

APPROXIMATE CONTENTS OF MATH 345

ALEX DEGTYAREV

ABSTRACT. As promised, here is a brief digest of what we covered in class. By no means is this text intended as a replacement for the textbook or lecture notes; rather, it indicates the parts of the textbook that we consider relevant. Though, as no one reads textbooks, including myself, I do not give precise chapter references.

Numerous digressions present material that I personally find interesting and important (and definitely worth further reading) but that is not officially included into the syllabus, hence neither to homeworks or exams.

CONTENTS

1. Vector calculus	2
1.1. The dot product	2
1.2. The cross product	3
1.3. Calculus	4
2. Curves	5
2.1. Plane curves	5
2.2. Space curves	6
2.3. Formulas	7
2.4. Global theorems	8
3. Surfaces: parametrizations <i>etc.</i>	11
3.1. Regular parametrizations	11
3.2. Thickening a parametrization	11
3.3. Differentiable (smooth) maps	13
3.4. Tangent planes/vectors	14
3.5. Differentials	16
4. Intrinsic geometry of surfaces: take I	18
4.1. The first fundamental form	18
4.2. Intrinsic distance and isometries	21
4.3. Surface area and 2-forms	22
4.4. Digression: the Poincaré model	23
4.5. Digression: the Lie bracket	24
4.6. Formulas	25
5. Curvatures	27
5.1. Normal vector and orientation	27
5.2. Normal and geodesic curvature; the second fundamental form	27
5.3. Principal curvatures and directions	28
5.4. The Gauss map and Weingarten operator	29
5.5. Digression: minimal surfaces	30
5.6. Digression: critical points of the distance function	31

5.7.	Compatibility equations	32
5.8.	Example: yet another characterization of spheres	34
5.9.	Digression: special parametrizations	35
6.	Intrinsic geometry continued	36
6.1.	The covariant derivative	36
6.2.	Parallel transport	39
6.3.	Geodesics	41
6.4.	The exponential map	43
6.5.	Complete surfaces	45
7.	Examples	47
7.1.	Graphs	47
7.2.	Surfaces of revolution	47
7.3.	Tubes	49
7.4.	Ruled surfaces	51
7.5.	Poincaré half-plane revisited	53
8.	Advanced topics	55
8.1.	The Gauss–Bonnet theorem	55
8.2.	Second variation of the arc length	59
8.3.	Jacobi fields	61
8.4.	Surfaces of constant Gaussian curvature	66

1. VECTOR CALCULUS

This is a brief summary of algebra and calculus of vectors in $\mathbb{R}^3$.

1.1. The dot product. Geometrically, the *dot* (or *scalar*) *product* of two vectors is defined as the *number*

$$\mathbf{a} \cdot \mathbf{b} := |\mathbf{a}||\mathbf{b}| \cos \theta \in \mathbb{R},$$

where θ is the angle between $\mathbf{a}$ and $\mathbf{b}$. Otherwise, on a higher dimensional space, it is a certain *positive definite* ($\mathbf{a}^2 \geq 0$, $\mathbf{a}^2 = 0$ iff $\mathbf{a} = 0$) *symmetric* ($\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$) bilinear form that is fixed in advance and considered part of the structure. It is used to “measure” lengths and angles:

$$|\mathbf{a}| := \sqrt{\mathbf{a}^2}, \quad \cos \theta = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}||\mathbf{b}|}.$$

In coordinates,

$$\mathbf{a} \cdot \mathbf{b} = a_x b_x + a_y b_y + a_z b_z$$

(the *standard inner product*).

Vectors $\mathbf{a}$, $\mathbf{b}$ are *orthogonal* (including the case $\mathbf{a} = 0$ or $\mathbf{b} = 0$) if and only if

$$\mathbf{a} \cdot \mathbf{b} = 0.$$

As a consequence, the equation of a line in $\mathbb{R}^2$ (plane in $\mathbb{R}^3$, *etc.*) passing through a point $\mathbf{a}$ and orthogonal to a vector $\mathbf{N} \neq 0$ is

$$(1.1) \quad (\mathbf{r} - \mathbf{a}) \cdot \mathbf{N} = 0.$$

1.2. The cross product. The *cross* (or *vector*) *product* is the *vector* $\mathbf{c} := \mathbf{a} \times \mathbf{b}$ defined *via*

$$|\mathbf{c}| = |\mathbf{a}||\mathbf{b}|\sin\theta, \quad \mathbf{c} \perp \mathbf{a}, \mathbf{b}, \quad (\mathbf{a}, \mathbf{b}, \mathbf{c}) \text{ is positively oriented.}$$

This operation makes sense *only in* $\mathbb{R}^3$. If $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$, then $\mathbf{a} \times \mathbf{b}$ is a multiple of $\mathbf{k}$. Observe the obvious $(\sin^2\theta + \cos^2\theta = 1)$; cf. (1.8) below for a generalization)

$$(1.2) \quad |\mathbf{a} \times \mathbf{b}|^2 + (\mathbf{a} \cdot \mathbf{b})^2 = |\mathbf{a}|^2|\mathbf{b}|^2, \quad |\mathbf{a} \times \mathbf{b}| \leq |\mathbf{a}||\mathbf{b}|, \quad |\mathbf{a} \cdot \mathbf{b}| \leq |\mathbf{a}||\mathbf{b}|;$$

the latter is known as the *Cauchy-Schwarz inequality*.

Vectors $\mathbf{a}, \mathbf{b}$ are linearly dependent if and only if

$$\mathbf{a} \times \mathbf{b} = \mathbf{0}.$$

In particular, the equation of a line in $\mathbb{R}^3$ passing through a point $\mathbf{a}$ and parallel to a vector $\mathbf{v} \neq \mathbf{0}$ is

$$(\mathbf{r} - \mathbf{a}) \times \mathbf{v} = \mathbf{0}.$$

The two products are subject to quite a number of useful (??) identities. Here are some of them.

$$(1.3) \quad \mathbf{a} \times \mathbf{b} := \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix},$$

$$(1.4) \quad \mathbf{b} \times \mathbf{a} = -\mathbf{a} \times \mathbf{b}, \quad \mathbf{a} \times \mathbf{a} = \mathbf{0},$$

$$(1.5) \quad (\mathbf{a}, \mathbf{b}, \mathbf{c}) := \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \begin{vmatrix} a_x & a_y & a_z \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{vmatrix},$$

$$(1.6) \quad \mathbf{a} \times (\mathbf{b} \times \mathbf{c}) + \mathbf{b} \times (\mathbf{c} \times \mathbf{a}) + \mathbf{c} \times (\mathbf{a} \times \mathbf{b}) = \mathbf{0},$$

$$(1.7) \quad \mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}(\mathbf{a} \cdot \mathbf{c}) - \mathbf{c}(\mathbf{a} \cdot \mathbf{b}),$$

$$(1.8) \quad (\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) = \begin{vmatrix} \mathbf{a} \cdot \mathbf{c} & \mathbf{a} \cdot \mathbf{d} \\ \mathbf{b} \cdot \mathbf{c} & \mathbf{b} \cdot \mathbf{d} \end{vmatrix},$$

$$(1.9) \quad (\mathbf{a} \times \mathbf{b}) \times (\mathbf{a} \times \mathbf{c}) = (\mathbf{a}, \mathbf{b}, \mathbf{c})\mathbf{a}.$$

The easiest/dumbest way to prove all that is to observe that both sides are polyliner and check the identities on all collections of basis vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$.

Geometrically, $|\mathbf{a} \times \mathbf{b}|$ is the area of the parallelogram spanned by $\mathbf{a}, \mathbf{b}$, and the *box* (or *mixed*) *product* $(\mathbf{a}, \mathbf{b}, \mathbf{c}) := \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$, see (1.5), is the *signed* volume of the parallelepiped (*aka* box) spanned by $\mathbf{a}, \mathbf{b}, \mathbf{c}$; the sign depends on the orientation of the triple. A consequence of (1.5) is that

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}) = -\mathbf{b} \cdot (\mathbf{a} \times \mathbf{c}) = \dots,$$

the sign depending on the parity of the permutation. Note that we do not reserve the parentheses exclusively for the box product, but we do use them whenever convenient; it is supposed to be clear from the context. Similar alternative notations that we do not use are

$$(\mathbf{a}, \mathbf{b}) = \langle \mathbf{a}, \mathbf{b} \rangle = \langle \mathbf{a} | \mathbf{b} \rangle \quad \text{for } \mathbf{a} \cdot \mathbf{b}, \quad [\mathbf{a}, \mathbf{b}] = \mathbf{a} \wedge \mathbf{b} \quad \text{for } \mathbf{a} \times \mathbf{b}.$$

(The last one is outrageous!)

1.3. Calculus. The limits, continuity, derivatives, *etc.* of vector valued functions, although defined formally within the framework of metric spaces, are merely done coordinatewise. For example, a function $\mathbf{r}(t) := x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}$ is continuous if and only if so are x, y, z , and

$$(x\mathbf{i} + y\mathbf{j} + z\mathbf{k})' = x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k}.$$

The derivative satisfies all “rules” of calculus, most notably the Leibnitz (product) rule for all thinkable products:

$$(f\mathbf{r})' = f'\mathbf{r} + f\mathbf{r}', \quad (\mathbf{r} \cdot \mathbf{s})' = \mathbf{r}' \cdot \mathbf{s} + \mathbf{r} \cdot \mathbf{s}', \quad (\mathbf{r} \times \mathbf{s})' = \mathbf{r}' \times \mathbf{s} + \mathbf{r} \times \mathbf{s}',$$

and

$$(\mathbf{a}, \mathbf{b}, \mathbf{c})' = (\mathbf{a}', \mathbf{b}, \mathbf{c}) + (\mathbf{a}, \mathbf{b}', \mathbf{c}) + (\mathbf{a}, \mathbf{b}, \mathbf{c}')$$

for the *box product* $(\mathbf{a}, \mathbf{b}, \mathbf{c}) := \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$, see (1.5).

A few examples of the Leibnitz (+ chain) rule:

$$|\mathbf{r}|' = (\sqrt{\mathbf{r}^2})' = \frac{\mathbf{r} \cdot \mathbf{r}'}{|\mathbf{r}|};$$

$$\text{if } |\mathbf{r}| = \text{const, then } \mathbf{r}' \perp \mathbf{r}, \text{ i.e., } \mathbf{r} \cdot \mathbf{r}' = 0.$$

Everything said applies to the partial derivative of multivariate vector-valued functions. We consistently use the standard abbreviation

$$\mathbf{r}_u := \frac{\partial \mathbf{r}}{\partial u}, \quad \text{etc.}$$

As in calculus, the partial commute:

$$\mathbf{r}_{uv} = \mathbf{r}_{vu}, \quad \text{or} \quad \frac{\partial^2}{\partial u \partial v} = \frac{\partial^2}{\partial v \partial u}.$$

The more meaningful notion of *differential* of a map is discusses in §3.5 below. In §6.1, in the context of tangent vector fields on surfaces, we discuss a version of partial derivatives that do *not* commute.

1.3.1. Vectors in calculus 102. Occasionally, we make use of the *pseudo-vector*

$$\nabla := \mathbf{i} \frac{\partial}{\partial x} + \mathbf{j} \frac{\partial}{\partial y} + \mathbf{k} \frac{\partial}{\partial z}$$

and operations

$$\text{grad } f := \nabla f, \quad \text{rot } \mathbf{F} = \text{curl } \mathbf{F} := \nabla \times \mathbf{F}, \quad \text{div } \mathbf{F} = \nabla \cdot \mathbf{F},$$

but only in a few examples. Remarkably, ∇ does have some of the algebraic properties of vectors mentioned above, most notably

$$\nabla \times (\nabla f) = 0, \quad \nabla \cdot (\nabla \times \mathbf{F}) = 0;$$

see also classical Stokes’ and Gauss–Ostrogradsky theorems in calculus textbooks. None of these makes sense unless “straight” Cartesian coordinates have been fixed. In general, one should speak about *differential forms* (§2.4.5, §3.4.2, §4.3), their *exterior derivative* d (which does have the property $d^2 = 0$), and generalized *Stokes’ theorem* (2.17). These topics are beyond the scope of this text.

2. CURVES

A *regular (parametrized) curve* is a map $\gamma: (a, b) \rightarrow \mathbb{R}^n$ (we stick to $n = 2, 3$) of class at least C^1 (continuously differentiable) such that

$$\gamma'(t) \neq 0 \quad \text{for any } t \in (a, b).$$

Often we speak about compact curves $[a, b] \rightarrow \mathbb{R}^n$; then, the derivatives $\gamma'(a)$ and $\gamma'(b)$ are understood as the one-sided ones.

Often one uses γ rather than $\mathbf{r}$ for the radius vector; this is particularly convenient when curves on surfaces are considered and $\mathbf{r}$ is reserved for the parametrization of the surface. We adopt this convention from now on.

Convention 2.1. In differential geometry we usually assume that the functions involved have as many continuous derivatives as needed for the formulas, still not insisting that they should be C^∞ or, God forbids, analytic. Note, though, that there is a popular activity (beyond the scope of this course) trying to understand the “minimal” assumptions for various global theorems (such as §2.4.4 or §2.4.6 below). A notorious example is Jordan’s curve theorem (§2.4.2 below), which is of purely topological nature, just a continuous curve without any regularity assumptions. (And that is one of the reasons why we do not attempt an artificially difficult proof by the means available in differential geometry.)

The derivative $\gamma'(t_0) \neq 0$ is a *tangent vector* to the curve at $\gamma(t_0)$. The *tangent line* is

$$(\mathbf{r} - \gamma(t_0)) \times \gamma'(t_0) = 0,$$

cf. (1.2). This should be regarded as a vector space $\{r\gamma'(t_0) \mid r \in \mathbb{R}\}$ rather than an affine line in $\mathbb{R}^2$. Tangent vectors are always “applied” to the tangency point.

The length of an arc between $\gamma(t_0)$ and $\gamma(t)$ is

$$s(t) = \int_{t_0}^t |\gamma'(t)| dt.$$

This function is invertible, $t = t(s)$; hence, when doing the *theory*, we always assume our curves parametrized by the arc length s (and the letter s in the context of curves always stands for the arc length). Typically, we use $'$ for d/dt with respect to an arbitrary parameter while spelling out d/ds explicitly.

The parametrization is by the arc length if and only if $|\gamma'| \equiv 1$; in this case, the *unit tangent vector* is denoted by $\mathbf{t} := d\gamma/ds$.

The natural parametrization is unique up to shifts $s \mapsto s - s_0$ and change of orientation $s \mapsto -s$. The latter changes the sign of $\mathbf{t}$ (the chain rule); therefore, the unit tangent vector $\mathbf{t}$ is orientation dependent.

Remark 2.2. No one knows exactly what an orientation of a curve is: it is something that changes when we change the “direction”, $t \mapsto -t$. For regular curves, one way to think about it is a coherent (*i.e.*, continuous) choice of one of the two unit tangent vectors at each points. This is always possible: just normalize γ' .

2.1. Plane curves. For a plane curve $\gamma: (a, b) \rightarrow \mathbb{R}^2$, in addition to the unit tangent vector $\mathbf{t} = d\gamma/ds$, we have the *signed normal vector* $\mathbf{n}_s$, which is the unit normal (orthogonal to $\mathbf{t}$) vector singled out by the condition that $(\mathbf{t}, \mathbf{n}_s)$ should be a positive frame:

$$\text{if } \mathbf{t} = x\mathbf{i} + y\mathbf{j}, \quad \text{then } \mathbf{n}_s = -y\mathbf{i} + x\mathbf{j}.$$

Referring to the ambient space $\mathbb{R}^3$, we can write $\mathbf{n}_s = \mathbf{k} \times \mathbf{t}$. Like $\mathbf{t}$, this vector depends on the orientation.

The *signed curvature* $\kappa_s \in \mathbb{R}$ is defined as the rate of change of the *direction* of $\mathbf{t}$ with respect to s :

$$(2.3) \quad \kappa_s := \frac{d\mathbf{t}}{ds} \cdot \mathbf{n}_s = \frac{d\theta}{ds}, \quad \text{where} \quad \mathbf{t}(s) = \mathbf{i} \cos \theta(s) + \mathbf{j} \sin \theta(s);$$

see (2.11), (2.12) below for the “formulas.” The sign depends on the orientation.

The Frenet equations for plane curves take the form

$$\frac{d\mathbf{t}}{ds} = \kappa_s \mathbf{n}_s, \quad \frac{d\mathbf{n}_s}{ds} = -\kappa_s \mathbf{t},$$

and the uniqueness theorem states that *any (say, continuous) function $\kappa_s(s)$ is the signed curvature of a plane curve unique up to rigid motion*. For example, this can be used to prove various statements that certain conditions on κ_s imply something about the curve: it would suffice to “guess” one curve with the right κ_s .

Warning 2.4. The uniqueness theorem has a limitation: κ_s is assumed to be a function of the *arc length* s . Unless they are constants, you cannot directly compare the curvatures $\kappa_s(u)$ and $\kappa_s(v)$ of two distinct curves written in terms of random parameters u, v .

One can regard a plane curve as lying in space and define the (unsigned) *curvature* and (principal) *normal*

$$\kappa := |\kappa_s| = \left| \frac{d\mathbf{t}}{ds} \right|, \quad \mathbf{n} := \frac{1}{\kappa} \frac{d\mathbf{t}}{ds}$$

similar to §2.2 below. The latter is only well defined if $\kappa \neq 0$; it does *not* depend on the orientation. It can either agree or disagree with $\mathbf{n}_s$. Geometrically, unlike $\mathbf{n}_s$, it always points “inside.”

Physicists would say that κ has dimension m^{-1} . Hence, the *radius of curvature* $R := 1/\kappa$ is a certain distance. The point

$$\mathbf{c}(s) := \gamma(s) - R(s)\mathbf{n}(s)$$

at distance R along the principal normal is the *center of curvature*. The significance of this point has been discussed in class. For example,

- $\mathbf{c}(s)$ is the center of the *osculating* (inflection tangent at $\gamma(s)$) circle; the radius of this circle is $R(s)$;
- the *distance function*

$$(2.5) \quad d(s) := |\gamma(s) - \mathbf{c}(s_0)|$$

has at s_0 a *degenerate* critical point, *i.e.*, both d' and d'' vanish.

2.2. Space curves. For a space curve $\gamma: (a, b) \rightarrow \mathbb{R}^3$, we do *not* have a signed curvature κ_s or vector $\mathbf{n}_s$. Instead, we have

<i>curvature</i>	$\kappa := \left \frac{d\mathbf{t}}{ds} \right ;$
(principal) <i>normal</i>	$\mathbf{n} := \frac{1}{\kappa} \frac{d\mathbf{t}}{ds};$
<i>osculating</i> (or <i>kissing</i>) <i>plane</i>	$:= \text{span}(\mathbf{t}, \mathbf{n});$
<i>binormal</i>	$\mathbf{b} := \mathbf{t} \times \mathbf{n};$

$$\text{(signed) torsion} \quad \tau := -\frac{d\mathbf{b}}{ds} \cdot \mathbf{n}.$$

The last four are only defined at points where $\kappa \neq 0$. The osculating plane is the “most tangent” of all tangent planes: it intersects the curve with order at least 3. Its equation is

$$(\mathbf{r} - \gamma(t_0)) \cdot \mathbf{b}(t_0) = 0,$$

cf. (1.1). Like κ , the torsion τ is the rate of change of the *direction* of the osculating plane with respect to the arc length s . Sometimes, we also speak about the

- *normal plane* $:= \text{span}(\mathbf{n}, \mathbf{b})$ and
- *rectifying plane* $:= \text{span}(\mathbf{t}, \mathbf{b})$.

Remark 2.6. Intuitively, we can give something a sign only when the underlying geometric object has “full dimension”, *i.e.*, that of the ambient space: then, we can compare the orientations. (A certain orientation of the ambient space is assumed fixed in advance.) Thus, we can speak about the sign of $(\mathbf{t}, \mathbf{n})$ in $\mathbb{R}^2$, but not in $\mathbb{R}^3$, where we need three vectors $(\mathbf{t}, \mathbf{n}, \mathbf{b})$. (A plane in $\mathbb{R}^3$ does not inherit any distinguished orientation from $\mathbb{R}^3$: you can always “flip it over” by a rigid motion.) Likewise, we can speak about signed area in $\mathbb{R}^2$, but not in $\mathbb{R}^3$, where we have signed volume instead.

The Frenet equations for $\mathbf{t}, \mathbf{n}, \mathbf{b}$ are (2.8) below, and the uniqueness theorem is slightly crippled (we have to assume that κ never vanishes): *any pair of continuous functions $\kappa(s) > 0$ and $\tau(s)$ can be realized as the curvature and torsion of a space curve, unique up to rigid motion.* Warning 2.4 fully applies.

2.3. Formulas. Here is a list of “important” formulas that pretty much cover the local theory of curves. They will be provided with the exam booklets.

The arc length:

$$(2.7) \quad ds = |\gamma'| dt; \quad \text{hence,} \quad s(t) = \int_{t_0}^t |\gamma'| dt.$$

Since $ds/dt = |\gamma'| > 0$, the function is invertible and we can always assume (though we can almost never find it) that $t = t(s)$, the *arc length parametrization*.

Frenet equations (and the definition of κ and τ):

$$(2.8) \quad \begin{aligned} \frac{d\mathbf{t}}{ds} &= \kappa \mathbf{n} \\ \frac{d\mathbf{n}}{ds} &= -\kappa \mathbf{t} + \tau \mathbf{b} \\ \frac{d\mathbf{b}}{ds} &= -\tau \mathbf{n} \end{aligned}$$

For plane curves, one can also change $\mathbf{n}$ to $\mathbf{n}_s$ and κ to κ_s , letting $\tau = 0$ and $\mathbf{b} = \mathbf{k} = \text{const}$. We do not need the assumption $\kappa_s \neq 0$ in this case.

Curvature and torsion (arbitrary parametrization):

$$(2.9) \quad \kappa = \frac{|\gamma' \times \gamma''|}{|\gamma'|^3},$$

$$(2.10) \quad \tau = \frac{\gamma' \cdot (\gamma'' \times \gamma''')}{|\gamma' \times \gamma''|^2}.$$

For the signed curvature of a plane curve $\gamma = (x, y)$, one can drop the $|\cdot|$ in (2.9):

$$(2.11) \quad \frac{\gamma' \times \gamma''}{|\gamma'|^3} = \kappa_s \mathbf{k}, \quad \text{or} \quad \kappa_s = \frac{x'y'' - x''y'}{(x'^2 + y'^2)^{3/2}} = \frac{\gamma'' \cdot \mathbf{n}_s}{|\gamma'|^2} = \frac{d\theta}{ds},$$

where θ is the angle between $\mathbf{t}$ and, say, the x -axis, see §2.4.1 below. If a plane curve is a graph $y = f(x)$, then, taking x for the parameter,

$$(2.12) \quad \kappa_s = \frac{f''}{(1 + f'^2)^{3/2}}.$$

Spelling out (2.10) is insane: we have a (3×3) -determinant upstairs and sum of squares (*aka* SOS) of three (2×2) -determinants downstairs.

Remark 2.13. As a practical hint, when differentiating (*e.g.*, deriving the formulas above), learn to look a few steps ahead and avoid unnecessary work. Since the derivatives are intended to appear in various dot/cross products, most terms will most likely vanish; hence, they do not need to be computed (as done in class).

2.4. Global theorems. Most theorems in this section have been taken for granted because they are intuitively obvious while not belonging to *differential* geometry. Nevertheless, you are supposed to be aware of the statements. Those that have been proved serve as examples of the kind of proofs expected from you.

Many of these theorems are about *closed curves* $\gamma: [a, b] \rightarrow \mathbb{R}^n$, *i.e.*, such that $\gamma(a) = \gamma(b)$ and sufficiently many (for the desired class of smoothness) derivatives of γ at a and b agree. A *simple closed curve* is a closed curve without self-intersections, *i.e.*, $t_1 < t_2$ implies $\gamma(t_1) \neq \gamma(t_2)$ unless $t_1 = a, t_2 = b$.

2.4.1. Hopf's Umlaufsatz. The unit vector $\mathbf{t}(s)$ can be written as

$$\mathbf{t}(s) = \mathbf{i} \cos \theta(s) + \mathbf{j} \sin \theta(s),$$

where $\theta \in S^1$ is an *angle*. Often, we need to disambiguate it to a *number* $\tilde{\theta}(s) \in \mathbb{R}^1$ in a continuous way. This is given by the so-called *path lifting lemma* (*homotopy lifting lemma*, *etc.*) in *homotopy theory*. Consider the map $\exp: \mathbb{R}^1 \rightarrow S^1, t \mapsto e^{it}$ (or $t \mapsto \mathbf{i} \cos t + \mathbf{j} \sin t$). Then, any continuous map (*aka path*) $\theta: [a, b] \rightarrow S^1$ *lifts* to $\mathbb{R}^1$, *i.e.*, there exists a continuous map $\tilde{\theta}: [a, b] \rightarrow \mathbb{R}^1$ such that $\exp \circ \tilde{\theta} = \theta$. This lift is unique as long as the value $\tilde{\theta}(a) \in \mathbb{R}^1$ over $\theta(a) \in S^1$ has been chosen.

The same applies to *homotopies* (families of paths), families of families, *etc.* For example, given a homotopy $h: [a, b] \times [0, 1] \rightarrow S^1$ (which can be regarded as a continuous family of paths $\theta_t(s) := h(s, t)$) and a lift (given by the path lifting lemma) $\tilde{\theta}_0$ of its restriction $\theta_0: [a, b] \times 0 \rightarrow S^1$, it extends to a unique lift $\tilde{h}$ of h .

For a regular curve, the lift of the angle $\theta(s)$ between $\mathbf{t}(s)$ and $\mathbf{i}$ can be given by a “formula”:

$$\tilde{\theta}(s) = \int_0^s \kappa_s(s) ds,$$

cf. (2.3). (But we still need to lift homotopies.) This leads us to the notion of *rotation index* of a regular curve $\gamma: [0, l] \rightarrow \mathbb{R}^2$ (the arc length parametrization):

$$(2.14) \quad \text{rot } \gamma := \frac{1}{2\pi} \int_0^l \kappa_s(s) ds.$$

The rotation index of a *closed* curve is obviously an integer.

Theorem 2.15 (Hopf's Umlaufsatz). *For a simple closed curve, $\text{rot } \gamma = \pm 1$.*

The proof is essentially topological: we homotope the *Gauss map* $s \mapsto \mathbf{t}(s)$ that we intend to analyze to a simpler map and observe that a continuously changing integer is constant. The homotopy in question is

$$h: \{(s_1, s_2) \mid a \leq s_1 \leq s_2 \leq b\} \rightarrow S^1, \quad (s_1, s_2) \mapsto \frac{\gamma(s_2) - \gamma(s_1)}{|\gamma(s_2) - \gamma(s_1)|},$$

extended to the boundary by continuity: $(s, s) \mapsto \mathbf{t}(s)$ and $(a, b) \mapsto -\mathbf{t}(a)$.

2.4.2. Jordan's curve theorem. *Let C be the trace of a simple closed curve. Then the complement $\mathbb{R}^2 \setminus C$ has exactly two connected components, with exactly one of them (called the interior of C) bounded.*

The statement holds for any merely *continuous* simple closed curve and is a one-line exercise in algebraic topology. Therefore, we skip the pain of proving it by inadequate means. Furthermore, with complex analysis (Riemann's theorem) thrown in, one shows that the interior is a topological disk. This latter refinement does not extend to higher dimensions.

2.4.3. Convex curves. *For the trace C of a simple closed curve γ , the following statements are equivalent:*

- (1) C is convex in the sense that it lies to one side of each of its tangents;
- (2) the curvature κ_s of γ never changes sign (although it is allowed to vanish);
- (3) the interior R of C (see §2.4.2) is convex in the usual sense: whenever a pair of points $\mathbf{a}, \mathbf{b}$ lies in R , so does the whole segment $[\mathbf{a}, \mathbf{b}]$.

Intuitively, this is more or less obvious. For a formal proof, read the textbook.

2.4.4. Fenchel's theorem. *For a closed curve $\gamma: [0, l] \rightarrow \mathbb{R}^3$ one has*

$$\int_0^l \kappa(s) ds \geq 2\pi,$$

the equality holding if and only if γ is a convex (§2.4.3) simple closed planar curve.

The ideas behind the proof are more important than the ultimate statement:

- (1) regard the *Gauss map* $s \mapsto \mathbf{t}(s)$ as a closed curve in the unit sphere S^2 ;
- (2) observe that the integral in question is merely the length of $\mathbf{t}$;
- (3) since γ is closed, $\mathbf{t}$ intersects each big circle in S^2 ;
- (4) a geometric input: (3) implies that $\text{length}(\mathbf{t}) \geq 2\pi$.

2.4.5. Green's theorem. *Let C be the trace of a simple closed curve γ and R , the closure of its interior. Let, further, P, Q be continuously differentiable functions on a neighborhood of R . Then*

$$\iint_R \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy = \int_C P dx + Q dy.$$

Every calculus 102 student is supposed to be able to do this at home: treat the P - and Q -parts separately and compute the double integral using Fubini (calculus 102) and Newton–Leibnitz (*aka* FTC, calculus 101).

This theorem is the key to establishing the path independence of line integrals (or potential energy of conservative fields in physics). Here is another consequence:

$$(2.16) \quad \text{Area}(R) := \iint_R dx dy = \int_C x dy = - \int_C y dx = \frac{1}{2} \int_C x dy - y dx.$$

We may use this theorem on surfaces, where it should be restated more carefully.

Green's theorem is a special case of *Stokes' theorem*, which incorporates Newton–Leibnitz, Green, classical Stokes, Gauss–Ostrogradsky, *etc.* and establishes a duality between geometric boundary and exterior derivative of differential forms:

$$(2.17) \quad \int_R d\omega = \int_{\partial R} \omega.$$

Unfortunately, even explaining the meaning of this formula is way beyond the scope of these notes. Though, like most great theorems, once observed, the proof is almost straightforward: Fubini + Newton–Leibnitz; it is in the “correct” definition of $d\omega$.

2.4.6. The isoperimetric inequality. *Let R be the interior of a simple closed curve C . Then*

$$4\pi \text{Area}(R) \leq \text{length}(C)^2,$$

the equality holding if and only if C is a circle. (The inequality can easily be remembered by eliminating the radius from the area of a disk *vs.* length of its boundary circle.)

The inequality holds for any curve, even not necessarily rectifiable (in the latter case the right hand side would be ∞ , making the inequality vacuous.) Below is a relatively simple proof for smooth (of class C^1) curves (not necessarily regular), *i.e.*, exactly the case where we can easily compute the arc length.

Let

$$t \mapsto \gamma(t) := x(t)\mathbf{i} + y(t)\mathbf{j}, \quad t \in [a, b],$$

be a parametrization of C . We can assume that C projects to $[-r, r]$ (the range of $x(t)$) in the x -axis, and let us compute the areas of R and of the disk of radius r , with the boundary circle S parametrized *via*

$$t \mapsto \alpha(t) := x(t)\mathbf{i} + \bar{y}(t)\mathbf{j}, \quad \bar{y}(t) := \pm\sqrt{r^2 - x(t)^2}.$$

By various parts of (2.16), using (1.2) and the fact that $|\alpha(t)| \equiv r$, we have

$$\begin{aligned} \text{Area}(R) + \pi r^2 &= \int_C x dy - \int_S y dx = \int_a^b (xy' - \bar{y}x') dt \\ &\leq \int_a^b |xy' - \bar{y}x'| dt = \int_a^b |\alpha \times \gamma'| dt \\ &\leq \int_a^b |\alpha| |\gamma'| dt = r \int_a^b |\gamma'| dt = r \text{length}(C). \end{aligned}$$

Since also

$$2\sqrt{\text{Area}(R) \cdot \pi r^2} \leq \text{Area}(R) + \pi r^2$$

(geometric *vs.* arithmetic mean), the inequality follows.

Note that the parametrization of S is very bad, $\bar{y}$ is not differentiable, but we never need to differentiate it! In fact, it is not even a parametrization, as it may keep running back and forth along the circle, but it is OK for computing the line integral: overlapping parts merely cancel.

3. SURFACES: PARAMETRIZATIONS ETC.

A *crucial difference* between surfaces and curves is that, for surfaces, we never have a distinguished canonical parametrization: most of the time, we do not have a global parametrization at all. Hence, whenever something is defined in terms of local coordinates, we need to check that the resulting object has an invariant geometric meaning. For most students this may be the first situation where one has to forget the calculus paradigm of variables $(x, y, \dots)$; for example, we have to carefully distinguish between functions (or pairs thereof), vector fields $P\mathbf{r}_u + Q\mathbf{r}_v$, and differential forms (covector fields) $P du + Q dv$. Though, cf. §4.1.3 below.

3.1. Regular parametrizations. Unlike curves, which are identified with their parametrization, a surface is defined as a geometric object, *viz.* a subset $X \subset \mathbb{R}^3$, which is “*locally* parametrizable.”

Formally, a (local) *regular parametrization* of X is a continuously differentiable (cf. Convention 2.1) injective function

$$\mathbf{r}: U \rightarrow \mathbb{R}^3, \quad \text{where } U \subset \mathbb{R}^2 \text{ is an open set in the } (u, v)\text{-plane}$$

such that

- $\mathbf{r}_u, \mathbf{r}_v$ are linearly independent ($\mathbf{r}_u \times \mathbf{r}_v \neq 0$) at each $(u, v) \in U$, and
- the image $\mathbf{r}(U)$ is $X \cap \tilde{U}$ for some open set $\tilde{U} \subset \mathbb{R}^3$.

We also say that $\mathbf{r}$ is a *local coordinate system* (*local coordinates*, a *local chart*, etc.) in the open (by definition) set $\tilde{U} \cap X \subset X$. Henceforth, $X \subset \mathbb{R}^3$ is called a *regular surface* if each point $x \in X$ is contained in (the image of) some local chart.

Formally, this definition does not cover surfaces with self-intersections and/or boundary, but we will not go there as we are mostly doing the local theory.

The assumption $\mathbf{r}_u \times \mathbf{r}_v \neq 0$ means that, at each point (u_0, v_0) , in the matrix

$$\begin{bmatrix} x_u & y_u & z_u \\ x_v & y_v & z_v \end{bmatrix}$$

at least one of the three (2×2) -minors does not vanish:

$$\frac{\partial(x, y)}{\partial(u, v)} := \begin{vmatrix} x_u & x_v \\ y_u & y_v \end{vmatrix} \neq 0, \quad \text{or} \quad \frac{\partial(x, z)}{\partial(u, v)} \neq 0, \quad \text{or} \quad \frac{\partial(y, z)}{\partial(u, v)} \neq 0.$$

Then, the implicit function theorem applies and, say, in the first case, *on a smaller neighborhood* of (u_0, v_0) , we can express (u, v) as functions of (x, y) and reparametrize the surface as

$$x = x, \quad y = y, \quad z = f(x, y), \quad \text{the } z\text{-component of } \mathbf{r}(u(x, y), v(x, y)).$$

Therefore, in a neighbourhood of each point, a surface looks like a graph

$$(3.1) \quad z = f(x, y), \quad \text{or} \quad y = g(x, z), \quad \text{or} \quad x = h(y, z).$$

Usually, all three would do but, at each point, *at least one* is guaranteed to work.

3.2. Thickening a parametrization. To prove formally some (intuitively clear) statements, instead of a “questionable” bivariate function $\mathbf{r}: U \rightarrow \mathbb{R}^3$ we prefer a *local diffeomorphism*

$$\bar{\mathbf{r}}: \bar{U} \rightarrow \mathbb{R}^3, \quad \text{where } \bar{U} \subset \mathbb{R}^3 \text{ is an open set in the } (u, v, w)\text{-space,}$$

so that X is the image of the coordinate plane $w = 0$.

One construction is straightforward, but not very natural: assuming one of (3.1), say, the first one, we merely define

$$\bar{\mathbf{r}}: (u, v, w) \mapsto \mathbf{r}(u, v) + w\mathbf{k},$$

so that the image of $\bar{\mathbf{r}}$ is fibered over X “vertically.”

Another construction is natural, we can patch the local pieces into one global *tubular neighbourhood*, but it needs $\mathbf{r}$ of class at least C^2 . (Why?) We let

$$\bar{\mathbf{r}}: (u, v, w) \mapsto \mathbf{r}(u, v) + w\mathbf{N}(u, v), \quad \mathbf{N} := \frac{\mathbf{r}_u \times \mathbf{r}_v}{|\mathbf{r}_u \times \mathbf{r}_v|},$$

so that the image is fibered over X in the normal directions. The partials are

$$\bar{\mathbf{r}}_u = \mathbf{r}_u + w\mathbf{N}_u, \quad \bar{\mathbf{r}}_v = \mathbf{r}_v + w\mathbf{N}_v, \quad \bar{\mathbf{r}}_w = \mathbf{N},$$

so that they are linear independent at $w = 0$ (as $(\mathbf{r}_u, \mathbf{r}_v, \mathbf{N}) = |\mathbf{r}_u \times \mathbf{r}_v| \neq 0$); by the multi-variate inverse function Theorem 3.5 below, this map is indeed a local diffeomorphism when restricted to (u, v, w) sufficiently close to $(u_0, v_0, 0)$.

Remark 3.2. Since we are doing local theory, we are always ready to restrict our attention to a smaller neighbourhood $V \subset U$ of the point in question. This smaller neighborhood may be given by the inverse/implicit function theorem, solutions to PDE's, or such. This is always assumed and not mentioned explicitly.

Thus, locally each regular surface is “standard”: it looks like the coordinate plane $w = 0$. But this description is too rough as it does not take into account the reach metric properties of $\mathbb{R}^3$.

Here is another application.

Theorem 3.3. *Any two regular parametrizations of a surface X are related by a diffeomorphism.*

More precisely, given $\mathbf{r}: U \rightarrow X$ and $\mathbf{s}: V \rightarrow X$ whose *supports* $\mathbf{r}(U)$, $\mathbf{s}(V)$ overlap, $W := \mathbf{r}(U) \cap \mathbf{s}(V) \neq \emptyset$, we have well-defined (set-theoretically) mutually inverse bijections

$$\mathbf{s}^{-1} \circ \mathbf{r}: \mathbf{r}^{-1}(W) \rightarrow \mathbf{s}^{-1}(W), \quad \mathbf{r}^{-1} \circ \mathbf{s}: \mathbf{s}^{-1}(W) \rightarrow \mathbf{r}^{-1}(W),$$

called *transition functions*. In plain English, we have two coordinate systems on W , viz. (u, v) and (s, t) , and we are discussing the change of coordinates. We assert that these functions are mutually inverse *diffeomorphisms*, i.e., differentiable of at least the same class as the minimum of $\mathbf{r}$, $\mathbf{s}$.

Convention 3.4. In view of Theorem 3.3, from now on we work with *one* function $\mathbf{r}$ (without even specifying its domain $U \subset \mathbb{R}^2$), often doing the computations in the *local coordinates* (u, v) . When necessary, e.g., when establishing the geometric invariance of our definitions, we consider a *change of coordinates*

$$u = u(s, t), \quad v = v(s, t) \quad \text{or} \quad s = s(u, v), \quad t = t(u, v)$$

(which may only make sense on a smaller domain $V \subset U$; since we are doing the local theory, we do not care). Theorem 3.3 guarantees that these are mutually inverse diffeomorphisms.

The problem with Theorem 3.3 is that it is not immediately clear how we can apply the inverse function theorem from calculus (Theorem 3.5 below) to establish the differentiability of $\mathbf{s}^{-1}$, or even what this differentiability would mean. However,

upon thickening everything (and restricting to smaller neighborhoods as dictated by [Theorem 3.5](#)), the differentiability of $\mathbf{s}^{-1}$ is given by [Theorem 3.5](#). Then we just compose and restrict to $\mathbf{r}^{-1}(W) \subset \bar{\mathbf{r}}^{-1}(\bar{W})$.

Theorem 3.5 (Digression: the [inverse function theorem](#)). *Let $U \subset \mathbb{R}^n$ be an open subset and $f: U \rightarrow \mathbb{R}^n$, $(x_1, \dots, x_n) \mapsto (u_1, \dots, u_n)$, a differentiable (at least C^1) map. Assume that, at some point $\bar{x} \in U$,*

$$\frac{\partial(u_1, \dots, u_n)}{\partial(x_1, \dots, x_n)} := \det \left[\frac{\partial u_i}{\partial x_j} \right] \neq 0.$$

Then, there is a neighborhood $V \subset \mathbb{R}^n$ of $\bar{u} := f(\bar{x})$ and a differentiable (of at least the same class as f) map $g: V \rightarrow \mathbb{R}^n$ such that $g(\bar{u}) = \bar{x}$ and $f \circ g = g \circ f = \text{id}$ (upon restriction to appropriate neighborhoods).

In plain English, in a sufficiently small neighborhood, we can solve “nicely” a non-linear system provided that we can solve its linearization. The proof should be done in calculus. Upon a *linear* change of variables we can assume that $[\partial u_i / \partial x_j] = \text{id}$, whereupon g is found by [Banach's contraction principle](#).

3.3. Differentiable (smooth) maps. From now on, I use “smooth” as a synonym for “continuously differentiable as many times as we need but *at least once*.”

Remark 3.6 (Digression). This apparent inaccuracy may be due to a remarkable fact from differential topology. We can consider the so-called *topological manifolds* (all maps involved, including charts, are only required to be continuous), and then we can try to introduce a [smooth structure](#) (of a class C^r , $1 \leq r \leq \infty$). Roughly, this means that, out of all charts, we only leave a subcollection with all transition functions C^r . Contrary to our low-dimensional intuition, not any manifold admits a C^1 -structure, and if it does, the structure does not need to be unique, even up to C^1 -diffeomorphism. (The so called *exotic* manifolds, like the 27 [exotic spheres](#) S^7 . A closed 4-manifold may have countably many non-diffeomorphic smooth structures, whereas $\mathbb{R}^4$ has a continuum of them! Accidentally, dimension 4 is also the first one where one has non-smoothable manifolds.) *However*, once a C^1 -structure has been fixed, it can be refined to any C^r , and this refinement is unique up to C^r -diffeomorphism. It is this reason why topologists do not care much about the class of smoothness. :)

None of these applies to *analytic structures*, which is way more than C^∞ .

Let $X \subset \mathbb{R}^m$, $Y \subset \mathbb{R}^n$ be regular surfaces, curves (to which all our theory obviously applies), or such. As usual, we stick to $m, n = 1, 2, 3$. We do not exclude the case of functions, *i.e.*, $Y = \mathbb{R}$, the “trivially parametrized” curve $x \mapsto x$ in $\mathbb{R}^1$.

Consider a map $f: X \rightarrow Y$. We want to understand what it means that f is *smooth*, or *differentiable*, although we are not trying to actually differentiate it yet (see [§3.5](#) below). There are at least two equivalent approaches to that.

3.3.1. The intrinsic approach. Pick a chart $\mathbf{r}$ about a point $x \in X$ and a chart $\mathbf{s}$ about $y := f(x) \in Y$ and declare f *smooth at x* if $\mathbf{s}^{-1} \circ f \circ \mathbf{r}$ is differentiable *in the usual sense*: note that this composition is a map from an open set of $\mathbb{R}^2$ to $\mathbb{R}^2$ (in the case of two surfaces), so the meaning of its differentiability is known from calculus. (We may need to restrict $\mathbf{r}$ to a smaller open set for the composition to make sense.) As usual, we say that f is smooth if so it is at each point $x \in X$.

In other words, we work in local coordinates (u, v) on X and (s, t) on Y , in which the composition $\mathbf{s}^{-1} \circ f \circ \mathbf{r}$ is merely a pair of functions

$$s = s(u, v), \quad t = t(u, v),$$

and we require that all partials s_u, s_v, t_u, t_v should exist and be continuous.

3.3.2. The external approach. Assuming $X, Y \subset \mathbb{R}^3$, we say that f is *smooth* at $x \in X$ if there is a neighbourhood $V \subset \mathbb{R}^3$ of x and a map $\tilde{f}: V \rightarrow \mathbb{R}^3$, smooth in the usual sense, such that $f = \tilde{f}|_X$ on $X \cap V$.

The equivalence of the two definitions is more or less obvious as soon as we use the thickening of the parametrizations, see §3.2. Both are equally useful: in practice we use both ways to describe maps, either local coordinates (u, v) or the ambient coordinates (x, y, z) . Some functions, like the curvature K , are defined on the surface only in the first place and should be treated in local coordinates. Others, like d in (2.5), are defined globally and it is convenient to use (x, y, z) . The essence of this section is that both languages can be used freely.

An extra advantage of §3.3.1 is the fact that it does not refer to the ambient space and, hence, is applicable to abstract surfaces.

3.4. Tangent planes/vectors. Roughly, intrinsic geometry of surfaces is all about tangent vectors. Hence, this notions should be thoroughly understood. There are many definitions, but for most students the following naïve one would suffice.

Definition 3.7. The *tangent plane* to a regular parametrized surface $\mathbf{r}: U \rightarrow \mathbb{R}^3$ at a point $x := \mathbf{r}(u_0, v_0)$ is the plane spanned by $\mathbf{r}_u(u_0, v_0)$, $\mathbf{r}_v(u_0, v_0)$. It is to be regarded as a vector space

$$T_x X := \{ \alpha \mathbf{r}_u + \beta \mathbf{r}_v \mid \alpha, \beta \in \mathbb{R} \}$$

whose elements are called *tangent vectors*.

Theorem 3.3 and the chain rule make sure that this plane is well defined: a reparametrization leads to a basis change in the same plane:

$$(3.8) \quad \mathbf{r}_s = u_s \mathbf{r}_u + v_s \mathbf{r}_v, \quad \mathbf{r}_t = u_t \mathbf{r}_u + v_t \mathbf{r}_v;$$

the transition matrix is the Jacobi matrix

$$J := \begin{bmatrix} u_s & u_t \\ v_s & v_t \end{bmatrix}.$$

For this (and not only this, see §3.4.3 below) reason one often uses the notation

$$\partial_u = \frac{\partial}{\partial u} := \mathbf{r}_u, \quad \partial_v = \frac{\partial}{\partial v} := \mathbf{r}_v.$$

Here, it is merely a *notation* for the two basis vectors. It agrees with (3.8).

Of course, we will take advantage of the fact that $T_x X$ is a subspace of $\mathbb{R}^3$: sometimes this is not necessary but convenient, sometimes, *e.g.*, in the discussion of curvatures, this is crucial. However, ideologically tangent vectors should always be regarded as applied to the tangency point; vectors tangent at distinct points should not be compared, or added, or such.

By no means should tangent vectors/vector fields be regarded as pairs of numbers/functions α, β : these are merely the *coordinates* in a chosen basis $\mathbf{r}_u, \mathbf{r}_v$. If the basis changes as in (3.8), they get multiplied by J^{-1} .

3.4.1. Curves on surfaces. A curve on a surface X is a curve in $\mathbb{R}^3$ whose trace happens to lie in X . Since we always want to work in the local coordinates on X , we use somewhat vague terminology and notation. Instead of a map $[a, b] \rightarrow \mathbb{R}^3$, we consider a map

$$\gamma: [a, b] \rightarrow U \subset \mathbb{R}^2, \quad t \mapsto (u(t), v(t)),$$

so that the curve *per se* is, in fact $\mathbf{r} \circ \gamma: [a, b] \rightarrow \mathbb{R}^3$. (This is guaranteed to lie in X .) Nevertheless, we speak of γ as a curve on X , even though technically it is a curve in $\mathbb{R}^2$. The tangent vector is

$$(3.9) \quad \frac{d}{dt}(\mathbf{r} \circ \gamma) = u' \mathbf{r}_u + v' \mathbf{r}_v \in TX$$

rather than $(u', v') \in \mathbb{R}^2$. The parameter t is the arc length if and only if

$$(3.10) \quad |u' \mathbf{r}_u + v' \mathbf{r}_v| = 1 \quad (\text{rather than } u'^2 + v'^2 = 1).$$

Of course, we cannot guarantee that the whole curve “fits” into our chosen chart. Hence, typically we break it into smaller pieces, each piece covered by its own chart. We will not consider “global” curves very often.

Now, we can define $T_p X$ as the set of vectors (3.9), evaluated at $t = 0$, over all curves $\mathbf{r} \circ \gamma$ such that $\mathbf{r}(\gamma(0)) = p$. More formally, an element of $T_p X$ is a class of curves $\gamma: (-\varepsilon, \varepsilon) \rightarrow U$, $\gamma(0) = (u_0, v_0) \mapsto p$, modulo the equivalence relation

$$(3.11) \quad \gamma \sim \tilde{\gamma} \quad \text{if and only if} \quad u'(0) = \tilde{u}'(0) \quad \text{and} \quad v'(0) = \tilde{v}'(0).$$

3.4.2. Differential forms. Equally important are *cotangent vectors*, i.e., elements of the dual space $T_x^* X$ (the *cotangent space*) or, in plain English, linear functionals $T_x X \rightarrow \mathbb{R}$. The dual basis $\mathbf{r}_u^*, \mathbf{r}_v^*$ is traditionally denoted by du, dv , and a mechanical application of the chain rule shows that a basis change results in

$$(3.12) \quad du = u_s ds + u_t dt, \quad dv = v_s ds + v_t dt.$$

Considering the opposite direction, as compared to (3.8), the transition matrix is precisely $(J^T)^{-1}$ (as dictated by linear algebra), justifying this notation.

A smooth covector field is called a *differential form* or, more precisely, *1-form*. In our notation, it looks like

$$P(u, v) du + Q(u, v) dv;$$

however, it is neither a pair of functions nor a vector field! (Though, see §4.1.3 below.) Contrary to the common belief, it is the 1-forms (not pairs of functions or vector fields) that can be integrated over curves on X . Naïvely, given a curve $\gamma: [a, b] \rightarrow \mathbb{R}^2$, see §3.4.1, we can formally write

$$\int_{\gamma} P du + Q dv := \int_a^b (P(\gamma(t))u' + Q(\gamma(t))v') dt$$

and observe that, should we change the parametrization of the surface, the change of variables formula agrees with (3.12) rather than (3.8). More geometrically, in order to create a Riemann sum, we need to “evaluate” the integrand on the tangent vectors of γ . Usually we achieve that by some sort of multiplication, which needs to be defined beforehand, but it is the covectors that, *by their very design*, are the objects created to be evaluated on tangent vectors!

3.4.3. Tangent vectors as derivations. Of course, the notation $\partial/\partial u, \partial/\partial v$ for the basis for TX is not accidental: functions can be differentiated with respect to tangent vectors (the directional derivative in calculus 102).

In coordinates, given a function $\varphi(u, v)$ and a point $p = \mathbf{r}(u_0, v_0)$, we define

$$\frac{\partial \varphi}{\partial \mathbf{v}} = \mathbf{v}(\varphi) := \alpha \varphi_u(u_0, v_0) + \beta \varphi_v(u_0, v_0) \in \mathbb{R} \quad \text{for } \mathbf{v} := \alpha \mathbf{r}_u + \beta \mathbf{r}_v \in T_p X$$

and use the chain rule to check that this expression agrees with (3.8): it does not change if we change the local coordinates.

A more geometric way would be to consider a curve $\gamma: (-\varepsilon, \varepsilon) \rightarrow \mathbb{R}^2$ (cf. §3.4.1) such that $\gamma(0) = (u_0, v_0)$ and $(\mathbf{r} \circ \gamma)'(0) = \mathbf{v}$ and define

$$\frac{\partial \varphi}{\partial \mathbf{v}} = \mathbf{v}(\varphi) := \left. \frac{d}{dt} \varphi(\gamma(t)) \right|_{t=0};$$

then, we check that this is independent of γ as long as $\gamma'(0) = (\alpha, \beta)$, cf. (3.11).

As yet another way, if φ is the restriction to X of a function $\bar{\varphi}$ on $\mathbb{R}^3$, see §3.3.2, we can regard $\mathbf{v}$ as a vector in $\mathbb{R}^3$ and take the usual directional derivative.

These observations lead us to an intrinsic definition of tangent vectors. Let $\mathcal{A}_p$ be the algebra of functions $\varphi: X \rightarrow \mathbb{R}$ smooth in a neighbourhood of a point $p \in X$. A *derivation* is a map $D: \mathcal{A}_p \rightarrow \mathbb{R}$ which is $\mathbb{R}$ -linear and satisfies the Leibnitz rule:

$$D(\varphi\psi) = D(\varphi)\psi(p) + \varphi(p)D(\psi).$$

Then, a tangent vector to X at p is nothing but a derivation.

In one direction, the correspondence is given by $\mathbf{v} \mapsto (\varphi \mapsto \mathbf{v}(\varphi))$ as described above. (All three definitions are equivalent due to the chain rule.) To see that every derivation comes from a tangent vector, we start from general nonsense:

$$D(1) = D(1 \cdot 1) = 2D(1), \quad \text{hence,} \quad D(1) = 0, \quad D(\text{const}) = 0$$

for any D . Then, *via* local coordinates, we identify $\mathcal{A}_p$ with the algebra of functions $\mathbb{R}^2 \rightarrow \mathbb{R}$ smooth about, say, the origin $(0, 0)$. A simple calculus lemma states that any such function can be written, not uniquely, as

$$\varphi(u, v) = \varphi(0, 0) + ug(u, v) + vh(u, v),$$

so that

$$D(\varphi) = D(u)g(0, 0) + D(v)h(0, 0).$$

Thus, apart from φ , the value $D(\varphi)$ depends on the constants $D(u), D(v) \in \mathbb{R}$ only. We conclude that derivations constitute a vector space generated by $\partial/\partial u, \partial/\partial v$.

I do not insist on your fluency with this definition. However, it has a number of advantages. First, it simplifies many constructions, cf. §3.5 or §4.5 below. Even more importantly, it applies to *abstract* surfaces, where we do not have the luxury of an ambient space containing all our vectors in the first place.

3.5. Differentials. As mentioned above, derivatives are numbers extremely useful in computation but lacking a geometric meaning. The correct geometric object to deal with is the *differential*.

Let $f: X \rightarrow Y$ be a smooth map between two, say, surfaces. The *differential of f at a point $x \in X$* is a *linear* map

$$df_x: T_x X \rightarrow T_y Y, \quad y := f(x).$$

This is, indeed, the “best linear approximation” except that we have to reconsider the meaning of “linear” in our non-linear settings of “curved” surfaces.

In local coordinates $\mathbf{r}(u, v)$ on X and $\mathbf{s}(s, t)$ on Y , the map f can be expressed as $s = s(u, v)$, $t = t(u, v)$, cf. §3.3.1, and df is given by the Jacobi matrix:

$$\mathbf{r}_u \mapsto s_u \mathbf{r}_s + t_u \mathbf{r}_t, \quad \mathbf{r}_v \mapsto s_v \mathbf{r}_s + t_v \mathbf{r}_t.$$

As usual, formally we just spell out the chain rule. This resembles (3.8), and that is not a surprise: in linear algebra, a basis change can easily be confused with a linear transformation. Of course, one needs to check that the map thus obtained is independent of the choice of local coordinates; otherwise, it would make no sense.

In the invariant terms of §3.4.3, the differential is *the very first thing that comes to one's mind*, next to general nonsense: for $\mathbf{v} \in T_x X$ and $\varphi: Y \rightarrow \mathbb{R}$, we let

$$(3.13) \quad df_x(\mathbf{v})(\varphi) := \mathbf{v}(\varphi \circ f).$$

Then, the coordinate form is yet another manifestation of the chain rule.

Yet another definition is in terms of curves, in the spirit of §3.4.1:

$$(3.14) \quad df(\gamma'(0)) = (f \circ \gamma)'(0) \quad \text{for a curve } \gamma: (-\varepsilon, \varepsilon) \rightarrow X.$$

The equivalence is given by the chain rule, see Remark 3.16 below.

Differentials appear everywhere, even if we do not notice them explicitly.

Example 3.15. Let $f: \mathbb{R}^m \rightarrow \mathbb{R}^n$. We can identify $T\mathbb{R}^m = \mathbb{R}^m$, $T\mathbb{R}^n = \mathbb{R}^n$; then, $df: \mathbb{R}^m \rightarrow \mathbb{R}^n$ is the linear map given by the Jacobi matrix $J_f := [\partial y_i / \partial x_j]$.

Remark 3.16. Accidentally, the concept of the differential explains the chain rule, too: it reads “the best linear approximation of a composition is the composition of the best linear approximations.” Or,

$$d(f \circ g) = df \circ dg,$$

which is immediately obvious from (3.13) or (3.14). In the settings of Example 3.15, this is

$$J_{f \circ g} = J_f J_g$$

which, upon spelling out the matrix product, is the awkward multi-variate chain rule found in calculus textbooks.

Example 3.17. If $f: X \rightarrow \mathbb{R}$ is a function, upon the identification $T\mathbb{R} = \mathbb{R}$ we get

$$df = f_u du + f_v dv: TX \rightarrow \mathbb{R},$$

i.e., a 1-form (a covector at each point $x \in X$). The 1-forms obtained in this way are called *exact*. They are easily integrated: for a curve $\gamma: [a, b] \rightarrow X$,

$$\int_{\gamma} df = f(b) - f(a).$$

This is a version of Newton–Leibnitz or yet another special form of Stokes (2.17).

Example 3.18. For a parametrization $\mathbf{r}: U \rightarrow X$, $U \subset \mathbb{R}^2$, identify $T\mathbb{R}^2 = \mathbb{R}^2$. Then, a tangent vector $\alpha \mathbf{r}_u + \beta \mathbf{r}_v \in TX$ is the image of $(\alpha, \beta) \in \mathbb{R}^2$ under $d\mathbf{r}$.

4. INTRINSIC GEOMETRY OF SURFACES: TAKE I

Intrinsic geometry is everything about the *metric* properties of *tangent* vectors.

4.1. The first fundamental form. As we know, see §1.1, metric properties of the plane, such as angle, length, area, are encoded in the dot product. Therefore, we define the *first fundamental form* l_x of a surface x as the restriction to $T_x X$ of the standard dot product on $\mathbb{R}^3$. In other words, we have

$$l_x(\mathbf{u}) := \mathbf{u}^2, \quad l_x(\mathbf{u}, \mathbf{v}) := \mathbf{u} \cdot \mathbf{v},$$

except that this applies to *tangent vectors* $\mathbf{u}, \mathbf{v} \in T_x X$ only.

As known from linear algebra (and is obvious), in coordinates any bilinear form is given by

$$(4.1) \quad l_x(\mathbf{u}, \mathbf{v}) = \mathbf{u}^T C \mathbf{v}$$

(in spite of §3.4, we cheat and identify vectors with their coordinates; otherwise, the notation would be intolerable), where C is the *Gram matrix* (in the given basis):

$$(4.2) \quad C := \begin{bmatrix} \mathbf{r}_u \cdot \mathbf{r}_u & \mathbf{r}_u \cdot \mathbf{r}_v \\ \mathbf{r}_v \cdot \mathbf{r}_u & \mathbf{r}_v \cdot \mathbf{r}_v \end{bmatrix} = \begin{bmatrix} E & F \\ F & G \end{bmatrix}.$$

Its coefficients are denoted by

$$E := \mathbf{r}_u^2, \quad F := \mathbf{r}_u \cdot \mathbf{r}_v = \mathbf{r}_v \cdot \mathbf{r}_u, \quad G := \mathbf{r}_v^2.$$

This form is symmetric (the two ex-diagonal coefficients are equal, denoted by F) and positive definite; in dimension 2 this boils down to

$$E, G > 0, \quad EG - F^2 > 0 \quad (\text{Cauchy-Schwarz}).$$

Convention 4.3. Just like $\kappa, \tau, \mathbf{t}, \mathbf{n}, \mathbf{b}$ for curves, throughout differential geometry of surfaces in $\mathbb{R}^3$ the letters E, F, G are reserved for the coefficients of the first fundamental form. Other fixed notations, to come up later, are

- $\mathbf{N}$ for the unit normal vector (§5.1);
- e, f, g for the coefficients of the second fundamental form II (§5.2);
- k_1, k_2, K, H for all kinds of curvatures (§5.3);
- Γ_{ij}^k for the Christoffel symbols (§5.7).

An alternative expression for l (used in the textbook) is

$$l = E du^2 + 2F du dv + G dv^2.$$

Formally, here du, dv are still the basis for T^*X , see §3.4.2, and the products are the *symmetrized* tensor products (as opposed to the *anti-symmetrized* product $\wedge$ used for differential forms, cf. §4.3.1 below), so that l is a section of the *symmetric power* $S^2 T^*X$ (which indeed it is). We will not follow this path: tensors are useful, but not in dimension 2 and not for the first reading.

4.1.1. Digression: tensor, symmetric, and exterior powers. I do *not* recommend reading this part; it is here for *very* curious students. In the realm of vector spaces of finite dimension, we can afford the luxury of defining **tensor powers** in terms of dual spaces, which is much easier. Recall that $(V^*)^* = V$: covectors can be evaluated on vectors, hence, vectors can also be evaluated on covectors; it is for this reason that the evaluation of a covector ω on a vector $\mathbf{v}$ is often written in the symmetric form of *Kronecker product*: $\langle \omega, \mathbf{v} \rangle := \omega(\mathbf{v})$. In the matrix language, this is the product of the row vector ω and column vector $\mathbf{v}$.

Let V be a vector space, $\dim V = n$, and let V^* be its dual. The m -th *tensor power* $\bigotimes^m V^*$ is the space of all m -linear (*i.e.*, linear with respect to each of its m arguments) functionals

$$\varphi: \underbrace{V \times \dots \times V}_{m \text{ times}} \rightarrow \mathbb{R}.$$

The *symmetric* $S^m V^* := \text{Sym}^m V^*$ and *exterior power* $\bigwedge^m V^*$ are the subspaces of, respectively, symmetric and skew-symmetric functionals. The *tensor product* of two functionals is obtained by listing their arguments, *declared independent*, next to each other: if $\varphi \in \bigotimes^p V^*$ and $\psi \in \bigotimes^q V^*$, then $\varphi \otimes \psi \in \bigotimes^{p+q} V^*$ is

$$(\varphi \otimes \psi)(\mathbf{u}_1, \dots, \mathbf{u}_p, \mathbf{v}_1, \dots, \mathbf{v}_q) := \varphi(\mathbf{u}_1, \dots, \mathbf{u}_p) \psi(\mathbf{v}_1, \dots, \mathbf{v}_q).$$

By the (weighted) summation over all permutations of the arguments, this operation can be symmetrized (denoted as ordinary product) or skew-symmetrized (denoted by $\wedge$, the *exterior product*), landing in S^{p+q} and $\bigwedge^{p+q}$, respectively.

Sticking to our notation, let $dx_1, dx_2, \dots, dx_n$ be a basis for V^* (dual to some basis for V). Stated in linear algebra and fairly obvious is the fact that an m -linear functional is determined by its values on all (appropriate) combinations of bases vectors. Hence, the basis for various powers are:

$$\begin{aligned} \bigotimes^m V^*: & \quad dx_{i_1} \otimes \dots \otimes dx_{i_m}, & i_1, \dots, i_m \in \{1, \dots, n\}, \\ S^m V^*: & \quad dx_{i_1} \dots dx_{i_m}, & 1 \leq i_1 \leq \dots \leq i_m \leq n, \\ \bigwedge^m V^*: & \quad dx_{i_1} \wedge \dots \wedge dx_{i_m}, & 1 \leq i_1 < \dots < i_m \leq n, \end{aligned}$$

and it is these expressions that appear in our constructions.

Since all this hardly makes much sense in dimension 2, we will not discuss contravariant (functionals on covectors) or mixed (arguments of both kind) tensors, indexing and summation conventions: they are only used by physicists and hard-core differential geometers. Besides, in the presence of the first fundamental form $\mathbf{l}$, there is a *canonical* isomorphism $TX = T^*X$, cf. §4.1.3 below.

4.1.2. Coordinate change; first applications. Once again, it should be kept in mind that E, F, G are numbers that we use in our computations. They depend on the parametrisation and are *not* invariants of the surface. (Some invariant quantities can be cooked up from these numbers; most notably, cf. Theorem 5.20 below.) The change of variables formula is given by (3.8) and linear algebra: the new coefficients $\tilde{E}, \tilde{F}, \tilde{G}$ are found from

$$(4.4) \quad \begin{bmatrix} \tilde{E} & \tilde{F} \\ \tilde{F} & \tilde{G} \end{bmatrix} = J^T \begin{bmatrix} E & F \\ F & G \end{bmatrix} J,$$

where J is the Jacobi matrix. I will not even try to spell this out!

Some of the formulas of §1 can be rewritten in terms of the first fundamental form and E, F, G . Given (4.1), I am not sure that this makes much sense, but just in case... Consider $\mathbf{u} := \alpha \mathbf{r}_u + \beta \mathbf{r}_v$ and $\mathbf{v} := \bar{\alpha} \mathbf{r}_u + \bar{\beta} \mathbf{r}_v$ in TX . Then:

$$\begin{aligned} |\mathbf{u}| &= \sqrt{\mathbf{l}(\mathbf{u})} = \sqrt{E\alpha^2 + 2F\alpha\beta + G\beta^2}, \\ \cos(\widehat{\mathbf{u}, \mathbf{v}}) &= \frac{\mathbf{l}(\mathbf{u}, \mathbf{v})}{|\mathbf{u}||\mathbf{v}|} = \frac{E\alpha\bar{\alpha} + F(\alpha\bar{\beta} + \bar{\alpha}\beta) + G\beta\bar{\beta}}{|\mathbf{u}||\mathbf{v}|}. \end{aligned}$$

The area A of the parallelogram spanned by $\mathbf{u}, \mathbf{v}$ is $|\mathbf{u} \times \mathbf{v}|$, but the cross product is not part of our current paradigm: the result is no longer a *tangent* vector. Hence,

we use (1.2):

$$(4.5) \quad A^2 = \mathbf{l}(\mathbf{u})\mathbf{l}(\mathbf{v}) - \mathbf{l}(\mathbf{u}, \mathbf{v})^2 = \underbrace{(EG - F^2)}_{\det(4.2)} \begin{vmatrix} \alpha & \beta \\ \bar{\alpha} & \bar{\beta} \end{vmatrix}^2$$

(try that at home).

4.1.3. *Gradient vector fields.* As known from linear algebra, in the presence of an inner product a space is *canonically* isomorphic to its dual:

$$\mathbf{u} \in T_x X \quad \text{giver rise to} \quad (\mathbf{v} \mapsto \mathbf{l}(\mathbf{u}, \mathbf{v})) \in T_x^* X,$$

and this assignment is a bijection. Hence, on *our* surfaces that do come with a metric, we can identify vector and covector fields. This identification is not quite straightforward, as $\mathbf{l}$ is typically not standard:

$$\alpha \mathbf{r}_u + \beta \mathbf{r}_v \not\leftrightarrow \alpha du + \beta dv;$$

one should use C^{-1} from (4.2) instead. For example, a function $\varphi: X \rightarrow \mathbb{R}$, apart from the differential $d\varphi$, has the *gradient vector field* $\nabla\varphi := \text{grad } \varphi$. But do not make a mistake and use a formula from calculus 102: if I am not mistaken,

$$\nabla\varphi \neq \varphi_u \mathbf{r}_u + \varphi_v \mathbf{r}_v; \quad \nabla\varphi = C^{-1} \begin{bmatrix} \varphi_u \\ \varphi_v \end{bmatrix} = \frac{(G\varphi_u - F\varphi_v)\mathbf{r}_u + (-F\varphi_u + E\varphi_v)\mathbf{r}_v}{EG - F^2}.$$

Just as in calculus, it is $\nabla\varphi$ that points in the direction of fastest accent of φ and is orthogonal (if course, with respect to $\mathbf{l}$) to its *level curves* $\{\varphi = \text{const}\}$. Given another vector $\mathbf{v}$, one has $\mathbf{l}(\nabla\varphi, \mathbf{v}) = d\varphi(\mathbf{v})$, essentially by the definition.

4.1.4. *Why not the dot product?* One may rightfully ask why we need to consider a weird “variable” form on weird variable planes is we already do have one nice dot product on the ambient space $\mathbb{R}^3$. There are several answers to that.

- (1) Similar to functions/maps, where we compute everything in terms of partial derivatives, we want to do the vector computation in local coordinates, too. For example, for the velocity vector of a curve, see §3.4.1, (3.10) becomes

$$|\mathbf{v}| = \sqrt{\mathbf{l}(\mathbf{v})} = \sqrt{Eu'^2 + 2Fu'v' + Gv'^2},$$

which is somewhat more explicit. The arc length becomes

$$(4.6) \quad \text{length}(\gamma) = \int_a^b \sqrt{Eu'^2 + 2Fu'v' + Gv'^2} dt.$$

- (2) As we will see, the form $\mathbf{l}$ and, especially, *the way it changes* has a dramatic effect on the geometry of the surface, cf. §4.5 below.
- (3) As always in mathematics, we want to concentrate on what we need and forget the rest of the universe. Of course, the dot product applied to vectors other than tangent is also important: it would tell us something about the curvature, *i.e.*, the way the surface sits in $\mathbb{R}^2$, see §5 below. But for now, it is only the tangent vectors that we are interested in.
- (4) We do not need to stick to the form inherited from $\mathbb{R}^3$. That is, we can start with a familiar surface and *artificially* change its first fundamental form to anything we like, provided that this form is symmetric and positive definite. Or, we can do that with abstract surfaces, where we have no ambient space at all and $\mathbf{l}$ is merely a new extra structure. Cf. §4.4 below.

4.2. Intrinsic distance and isometries. The *intrinsic distance* on a regular surface X is defined *via*

$$\text{dist}(x, y) := \inf \{ \text{length}(\gamma) \mid \gamma: [a, b] \rightarrow X \text{ a curve, } \gamma(a) = x, \gamma(b) = y \}.$$

Clearly, we can restrict this to smooth or even regular curves, with a slight complication when checking the axioms. This has all properties of distance,

- $\text{dist}(x, y) \geq 0$, $\text{dist}(x, y) = 0$ if and only if $x = y$ (the only curve of length 0 is constant);
- $\text{dist}(y, x) = \text{dist}(x, y)$ (obvious: re-orient γ);
- $\text{dist}(x, z) \leq \text{dist}(x, y) + \text{dist}(y, z)$ (the triangle inequality; extra work may be needed to smoothen the union of two curves),

making X a metric space. Clearly,

$$\text{dist}_X(x, y) \geq \text{dist}_{\mathbb{R}^3}(x, y),$$

the equality holding if and only if the line segment $[x, y]$ lies in X .

By (4.6), the distance on X depends on the first fundamental form only, making it part of intrinsic geometry.

Convention 4.7. By definition, we say that the inclusion $X \hookrightarrow \mathbb{R}^3$ is an *isometry*, or *isometric embedding*, even though it almost surely is *not* an isometry of metric spaces. By this we always mean that *the first fundamental form on X* (rather than the distance function) is inherited from that on $\mathbb{R}^3$, *i.e.*, the dot product.

More generally, a diffeomorphism $f: X \rightarrow Y$ is called an *isometry* if f preserves the intrinsic distance. For a *diffeomorphism* this requirement is equivalent to saying that df preserves the first fundamental form,

$$(4.8) \quad l_X(\mathbf{u}, \mathbf{v}) = l_Y(df(\mathbf{u}), df(\mathbf{v})), \quad \mathbf{u}, \mathbf{v} \in TX$$

or, in the matrix language of (4.2),

$$\begin{bmatrix} E_X & F_X \\ F_X & G_X \end{bmatrix} = J^T \begin{bmatrix} E_Y & F_Y \\ F_Y & G_Y \end{bmatrix} J,$$

where J is the Jacobi matrix of f .

Usually, this is too much to ask for and we settle for *local isometries*, requiring the smoothness and (4.8), which clearly implies that $\det J \neq 0$ and, hence, f is a local diffeomorphism by Theorem 3.5.

Example 4.9. As done in class, the parametrization

$$f: (u, v) \mapsto (\cos u, \sin u, v)$$

is a local isometry from the (u, v) -plane $\mathbb{R}^2$ to the cylinder Y : it does preserve the first fundamental form. It does not preserve distance: the points $(0, 0)$ and $(0, 2\pi)$ at distance 2π are mapped to the same point $(1, 0, 0) \in Y$. Even if we make f injective by restricting it to the strip $0 < u < 2\pi$, it will *not* preserve distances, mapping $(\varepsilon, 0)$ and $(2\pi - \varepsilon, 0)$ at distance $2\pi - 2\varepsilon$ to the points $(\cos \varepsilon, \pm \sin \varepsilon, 0)$ at distance 2ε . Note though that, if we restrict f to a strip of width $< \pi$, it *will* preserve the distance as well. This is a general phenomenon: any local isometry is distance preserving when restricted to a *sufficiently small* neighborhood U , *e.g.*, such that $f(U)$ lies in a *convex neighborhood* as in Corollary 6.37 below.

Analyzing [Example 4.9](#), we can easily describe the shortest curves connecting two points $f(u_1, v_1)$ and $f(u_2, v_2)$. In the plane, consider the straight line segments connecting (u_1, v_1) and any of $(u_2 + 2\pi n, v_2)$, $n \in \mathbb{Z}$. These segments are mapped to arcs of helices (circles, if $v_1 = v_2$) on Y , which are the “straightest” curves (to be called *geodesics*) connecting the two points. However, only one of them (two, if $u_2 = u_1 + \pi \bmod 2\pi$) is shortest, realizing the distance. We will study this later.

Now, we can officially define that intrinsic geometry of a surface is whatever is preserved by isometries, whereas its local intrinsic geometry is what is preserved by local isometries. Both are encoded in the first fundamental form, but for the former we also need some global topological data, so it is not quite differential geometry. With an abuse of language, intrinsic geometry is whatever can be expressed in terms of E, F, G ; all truly interesting things also involve their derivatives.

4.3. Surface area and 2-forms. Unfortunately, the rigorous treatment of the length of a curve as supremum of those of inscribed broken lines does not work for surfaces. You can check Wikipedia for the [Schwarz’s boot](#) (*aka Schwarz lantern*): to something as simple as the cylinder one can inscribe a sequence of polyhedral surfaces of unbounded area. Hence, we have no choice but formal usage of integrals. By (4.5), the area of a “very small” parallelogram spanned by $\mathbf{r}_u du, \mathbf{r}_v dv$ (here, I cannot think of any smart interpretation of the notorious du, dv : they are just some infinitesimally small coefficients) is

$$dA = \sqrt{EG - F^2} du dv,$$

see (4.5), and this expression is to be integrated over the whole surface (or, rather, the domain of the parameters u, v):

$$\text{Area}(X) = \iint_X \sqrt{EG - F^2} du dv.$$

If X is not covered by a single chart, we break it into pieces and sum up.

To make this an acceptable *definition*, we need to check its invariance under a change of coordinates. The key is (4.4), the multiplicativity of $\det$, and the fact that it is $\det J$ that appears in the change of variables formula in calculus 102.

Example 4.10. As done in class, for the standard parametrization of the unit sphere by colatitude u and longitude v , we have $E = 1$, $F = 0$, $G = \sin^2 u$. (Strictly speaking, this is not quite a regular parametrization, but something happening on a curve does not affect area/double integral.) Thus, restricting to the rectangle $\Omega := \{0 \leq u \leq \pi, 0 \leq v \leq 2\pi\}$, we have

$$\text{Area}(S^2) = \iint_{\Omega} \sin u du dv = \int_0^{2\pi} \int_0^{\pi} \sin u du dv = 4\pi.$$

4.3.1. Digression: 2-forms. At this point I cannot help mentioning the fact that it is not functions that we can integrate over surfaces. Instead, we should integrate *2-forms*, which are smooth sections of the *exterior power* $\bigwedge^2 T^*X$ or, in plain English (in dimension 2), expressions

$$f(u, v) du dv \quad (\text{should in fact be } f(u, v) du \wedge dv, \text{ see } \S 4.1.1)$$

that change accordingly under the change of coordinates. Declaring $du dv$ bilinear and *skew*-symmetric (orientation matters!), the change of variables is, *cf.* (3.12),

$$du dv = (u_s ds + u_t dt)(v_s ds + v_t dt) = (u_s v_t - u_t v_s) ds dt = (\det J) ds dt,$$

which agrees with the change of variables formula for double integrals.

Luckily, *in the presence of the first fundamental form* I (cf. §4.1.3), we have a distinguished *volume form*

$$d\text{vol} := \sqrt{EG - F^2} du dv,$$

sometimes (in this dimension) called the area form and denoted by dS or dA :

$$d\text{vol}(\mathbf{u}, \mathbf{v}) := \text{Area}_{\text{signed}}(\text{parallelogram spanned by } \mathbf{u}, \mathbf{v}).$$

Hence, we can pretend that we integrate a function f by multiplying it by $d\text{vol}$, and it is indeed the expression $f d\text{vol} = f dS = f dA = \dots$, not just f itself, that appears under the integral sign in calculus textbooks.

In this language, the eclectic formula for $\text{Area}(X)$ (cf. Example 5.25, where, instead of S^2 , we rightfully use some mysterious Ω for the domain of integration) becomes

$$\text{Area}(X) = \iint_X 1 \cdot d\text{vol},$$

which makes perfect sense. Computationwise, all this does not make an iota of difference, but it is always good to distinguish the way things are computed from what they actually mean.

4.4. Digression: the Poincaré model. To illustrate item (4) in §4.1.4, let us consider the upper half-plane

$$\mathbb{H} := \{(x, y) \in \mathbb{R}^2 \mid y > 0\}.$$

(What can be flatter than that?) The natural parametrization $(x, y) \mapsto (x, y)$ identifies all tangent planes with $\mathbb{R}^2$ itself, with the basis $\mathbf{r}_x = \mathbf{i}$, $\mathbf{r}_y = \mathbf{j}$. Now, in this basis, let us change the first fundamental form by declaring that

$$E(x, y) = G(x, y) := \frac{1}{y^2}, \quad F(x, y) = 0.$$

Immediately, everything changes. (We will do more computation later, see §7.5 below; for now, I just state some results.)

Unlike the original half-plane, the new surface is complete, *i.e.*, any Cauchy (with respect to the new distance) sequence converges (in the new distance).

The “straightest”, and also shortest curves realizing the distance between pairs of points are either vertical lines $x = \text{const}$ or semi-circles $y = \sqrt{r^2 - (x - a)^2}$, $r, a = \text{const}$, orthogonal to the *absolute* $y = 0$ (which is *not* part of our surface).

Declaring these shortest curves “lines”, we obtain a “geometry” very similar to the high school one, except for Euclid’s *parallel postulate*: through a given point, there are *at least two* lines disjoint from a given one.

This *hyperbolic geometry* can be developed to the same extent as that in high school, with its own trigonometry and such. We get funny formulas like

$$\sin A = \frac{\sinh a}{\sinh c}, \quad \cos A = \frac{\tanh b}{\tanh c}$$

for a right triangle (that is why those are *hyperbolic* functions) or (§8.1.5 below)

$$\text{Area}(\triangle ABC) = \pi - (A + B + C).$$

From the point of view of differential geometry what we got is a *complete surface of constant negative curvature* $K = -1$; as explained, it is a model for the hyperbolic geometry. To make things worth, *Hilbert’s theorem* states that such a surface does

not admit an isometric embedding (or even immersion) to $\mathbb{R}^3$, so that we cannot observe it by the “conventional” means. It is probably due to this fact that the discovery of hyperbolic geometry took so long (the very end of the 19th century). An example of an incomplete surface in $\mathbb{R}^3$ of constant negative curvature is one half of a *tractroid* (surface of revolution of a *tractrix*), given by

$$(u, v) \mapsto ((u - \tanh u), \operatorname{sech} u \cos v, \operatorname{sech} u \sin v).$$

Alas, the whole surface is singular (and not simply connected).

4.5. Digression: the Lie bracket. Here, we elaborate on item (2) in §4.1.4.

From linear algebra we know that *any inner product is standard*, i.e., there is an *orthonormal* basis $\mathbf{u}, \mathbf{v}$ in which that Gram matrix is the identity. Furthermore, this basis is obtained by the *Gram-Schmidt process*, which is differentiable: from the original basis $\mathbf{r}_u, \mathbf{r}_v$ we pass to

$$\tilde{\mathbf{u}} := \mathbf{r}_u, \quad \tilde{\mathbf{v}} := \mathbf{r}_v - \frac{l(\mathbf{r}_v, \tilde{\mathbf{u}})}{l(\tilde{\mathbf{u}})} \tilde{\mathbf{u}}$$

and normalize:

$$\mathbf{u} := \frac{\tilde{\mathbf{u}}}{|\tilde{\mathbf{u}}|}, \quad \mathbf{v} := \frac{\tilde{\mathbf{v}}}{|\tilde{\mathbf{v}}|}.$$

Thus, it appears that, no matter what we start with, we can always find a pair of linearly independent (at each point) vector fields in which l is standard:

$$E = G = 1, \quad F = 0.$$

That is, contrary to §4.4, there is no such thing as interesting intrinsic geometry.

The problem is that it is highly unlikely that $\mathbf{u}, \mathbf{v}$ are of the form $\mathbf{s}_u, \mathbf{s}_v$ for some parametrization $\mathbf{s}$; on the other hand, only this kind of bases is covered by our theory. To understand why $\mathbf{s}$ cannot be found, we need the notion of *Lie bracket* of two vector fields.

Let $\mathbf{u} := \alpha \mathbf{r}_u + \beta \mathbf{r}_v$ and $\mathbf{v} := \bar{\alpha} \mathbf{r}_u + \bar{\beta} \mathbf{r}_v$. According to §3.4.3, the vector $\mathbf{v}$ can be applied to a function φ at a point to yield a number. Since now $\mathbf{v}$ is a vector *field*, we can do that at each point, obtaining a new *function* $\mathbf{v}(\varphi)$, to which we can further apply $\mathbf{u}$. Or, we can do that in the opposite order and question if the result is the same. A simple computation in local coordinates shows that

$$(4.11) \quad \begin{aligned} & \mathbf{u}(\mathbf{v}(\varphi)) - \mathbf{v}(\mathbf{u}(\varphi)) \\ &= (\alpha \bar{\alpha}_u - \bar{\alpha} \alpha_u + \beta \bar{\alpha}_v - \bar{\beta} \alpha_v) \varphi_u + (\alpha \bar{\beta}_u - \bar{\alpha} \beta_u + \beta \bar{\beta}_v - \bar{\beta} \beta_v) \varphi_v; \end{aligned}$$

all second derivatives cancel because $\varphi_{uv} = \varphi_{vu}$. We conclude that

- in general, there is no reason to expect that $\mathbf{u}(\mathbf{v}(\varphi)) - \mathbf{v}(\mathbf{u}(\varphi)) = 0$, i.e., that the two vector fields *commute*, and
- $\mathbf{u}(\mathbf{v}(\varphi)) - \mathbf{v}(\mathbf{u}(\varphi))$ is a derivation, hence, a new vector field.

By definition,

$$[\mathbf{u}, \mathbf{v}](\cdot) := \mathbf{u}(\mathbf{v}(\cdot)) - \mathbf{v}(\mathbf{u}(\cdot))$$

is the *Lie bracket* of $\mathbf{u}, \mathbf{v}$. This operation has many properties of the cross product, most notably bilinearity (with respect to the multiplication by *constants*), (1.4), and (1.6). Thus, with respect to $[\cdot, \cdot]$, the (infinite dimensional) space of all vector fields on X constitutes a *Lie algebra*. The Leibnitz rule manifests itself *via*

$$(4.12) \quad [\mathbf{u}, \varphi \mathbf{v}] = \mathbf{u}(\varphi) \mathbf{v} + \varphi [\mathbf{u}, \mathbf{v}].$$

We do not discuss an analogue of §1.3, i.e., derivatives $[\mathbf{u}, \mathbf{v}]_u$ or such. Why?

Of course, (4.11) could have been taken for the definition, but then checking the independence of the coordinates would be quite involved.

Now we observe that, for any parametrization $\mathbf{r}$,

$$[\mathbf{r}_u, \mathbf{r}_v] = 0, \quad \text{literally,} \quad \frac{\partial^2}{\partial u \partial v} = \frac{\partial^2}{\partial v \partial u},$$

Clairaut's theorem from calculus or (4.11) with $\alpha, \beta, \bar{\alpha}, \bar{\beta} = \text{const}$. This fact is a serious obstruction showing that *most* pairs of vector fields on a surface, even linearly independent, do not come from a parametrization. In particular, most likely, neither do $\mathbf{u}, \mathbf{v}$ given by the Gram–Schmidt.

From this discussion, it should be clear that the next statement is the best that we can hope for: we can keep *directions*, but not vector fields.

Theorem 4.13. *Let $\mathbf{u}, \mathbf{v}$ be two vector fields linearly independent at a point $p \in X$. Then, there is a parametrization $\mathbf{r}$ of a neighborhood $U \ni p$ such that*

$$\mathbf{r}_u = \lambda \mathbf{u}, \quad \mathbf{r}_v = \mu \mathbf{v}$$

for some functions μ, ν , i.e., $\mathbf{r}_u, \mathbf{r}_v$ have the direction of $\mathbf{u}, \mathbf{v}$, respectively.

Proof. Locally, each vector field (aka first order ODE), say $\mathbf{u}$, has a *first integral*, i.e., a function φ such that

$$d\varphi \neq 0 \quad \text{but} \quad d\varphi(\mathbf{v}) \equiv 0;$$

in other words, $\mathbf{u}$ is tangent to the level curves of φ . We take first integrals φ, ψ of $\mathbf{u}, \mathbf{v}$, respectively and consider the map

$$f: U \rightarrow \mathbb{R}^2, \quad q \mapsto (\varphi(q), \psi(q)).$$

By the construction, and since $\mathbf{u}, \mathbf{v}$ are linearly independent at p ,

$$df_p(\mathbf{u}) = d\varphi_p(\mathbf{u})\mathbf{j} \neq 0 \quad \text{and} \quad df_p(\mathbf{v}) = d\psi_p(\mathbf{v})\mathbf{i} \neq 0,$$

i.e., df_p has maximal rank 2 and, hence, $\mathbf{r} := f^{-1}$ is given by Theorem 3.5. $\square$

Corollary 4.14. *Locally, a pair $(\mathbf{u}, \mathbf{v})$ of vector fields is of the form $(\mathbf{r}_u, \mathbf{r}_v)$ for a parametrization $\mathbf{r}$ if and only if $[\mathbf{u}, \mathbf{v}] \equiv 0$.*

Proof. Start from Theorem 4.13 and use (4.12) to see how $\mathbf{r}_u, \mathbf{r}_v$ can be rescaled to keep $[\varphi\mathbf{r}_u, \psi\mathbf{r}_v] \equiv 0$. We find $\varphi_v \equiv \psi_u \equiv 0$, which is realized by a re-parametrization of the form $\mathbf{r}(u(\tilde{u}), v(\tilde{v}))$. $\square$

4.6. Formulas. Here is the principal take-out from this section.

The *first fundamental form* is the dot product:

$$(4.15) \quad l(\mathbf{u}, \mathbf{v}) := \mathbf{u} \cdot \mathbf{v}, \quad l(\mathbf{u}) := \mathbf{u}^2, \quad \mathbf{u}, \mathbf{v} \in TX,$$

restricted to the tangent planes. Its coefficients are

$$(4.16) \quad E := \mathbf{r}_u^2, \quad F := \mathbf{r}_u \cdot \mathbf{r}_v = \mathbf{r}_v \cdot \mathbf{r}_u, \quad G := \mathbf{r}_v^2; \quad C := \begin{bmatrix} E & F \\ F & G \end{bmatrix},$$

so that

$$(4.17) \quad l(\mathbf{u}, \mathbf{v}) = \mathbf{u}^T C \mathbf{v} \quad (\text{matrix product}).$$

(The notation C for the Gram matrix is not commonly used.)

The idea is that, instead of (x, y, z) in $\mathbb{R}^3$, we want to use the local coordinates (u, v) on X , speaking about

- vectors $(\alpha, \beta) \in \mathbb{R}^2$ rather than $\alpha\mathbf{r}_u + \beta\mathbf{r}_v \in TX \subset \mathbb{R}^3$ or

- curves $\gamma: [a, b] \rightarrow \mathbb{R}^2$ rather than $\mathbf{r} \circ \gamma: [a, b] \rightarrow X \subset \mathbb{R}^3$.

(Strictly speaking, linear algebra dictates the column notation

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} \in \mathbb{R}^2,$$

but that would spread the lines too much; it is *assumed* in (4.17) and such.)

Thus, for vectors $\mathbf{u} = (\alpha, \beta)$, $\mathbf{v} = (\bar{\alpha}, \bar{\beta})$ and a curve $\gamma: t \mapsto (u(t), v(t))$, we have

$$(4.18) \quad l(\mathbf{u}) = E\alpha^2 + 2F\alpha\beta + G\beta^2,$$

$$(4.19) \quad l(\mathbf{u}, \mathbf{v}) = E\alpha\bar{\alpha} + F\alpha\bar{\beta} + F\beta\bar{\alpha} + G\beta\bar{\beta},$$

$$(4.20) \quad |\mathbf{u}| = \sqrt{l(\mathbf{u})} \quad (\text{see §1.1}),$$

$$(4.21) \quad \cos(\widehat{\mathbf{u}, \mathbf{v}}) = \frac{l(\mathbf{u}, \mathbf{v})}{|\mathbf{u}||\mathbf{v}|} \quad (\text{see §1.1}),$$

$$(4.22) \quad |\gamma'| = \sqrt{Eu'^2 + 2Fu'v' + Gv'^2},$$

$$(4.23) \quad \text{length}(\gamma) = \int_a^b \sqrt{Eu'^2 + 2Fu'v' + Gv'^2} dt.$$

For the moment, we do not discuss $\mathbf{n}_s$ or κ_s : later, they are to be replaced with their so-called *geodesic* counterparts $\mathbf{n}_g, \kappa_g$.

Once you understand the concept “work in $\mathbb{R}^2$ pretending that it is X ”, there is no need in memorizing all these formulas.

The surface area is

$$(4.24) \quad \text{Area}(X) = \iint_X 1 \cdot d\text{vol} = \iint_{\Omega} \sqrt{EG - F^2} du dv,$$

where Ω is the domain of *the parameters* (u, v) .

These formulas simplify in various special coordinate systems whose existence is to be proved later, most notably, in *isothermal coordinates*, where

$$(4.25) \quad E \equiv G \equiv e^{\rho(u,v)}, \quad F \equiv 0,$$

or *normal polar coordinates* (see §6.4.1 below), where

$$(4.26) \quad E \equiv 1, \quad F \equiv 0, \quad G = G(u, v).$$

See also other special parametrizations in §5.9 below.

Remark 4.27. In spite of apparent similarity, the existence of (4.26) is rather straightforward (we will do this in §6.4.1 below), whereas (4.25) is a big theorem involving PDE's; originally it was proved by Gauss in the analytic settings only. This statement does not extend to higher dimensions. Geometrically, it means that any surface is locally *conformal* to the plane, *i.e.*, is related by a diffeomorphism preserving angles (but not necessarily distances).

5. CURVATURES

We are ready to discuss the way that a surface is “curved” in $\mathbb{R}^3$. For this, we need to leave the comfortable tangent plane and consider normal vectors as well.

5.1. Normal vector and orientation. Let $\mathbf{r}$ be a regular parametrization. By definition, $\mathbf{r}_u, \mathbf{r}_v$ are linearly independent, *i.e.*, $\mathbf{r}_u \times \mathbf{r}_v \neq 0$. The latter vector is obviously orthogonal to $TX = \text{span}(\mathbf{r}_u, \mathbf{r}_v)$, and we can define the *unit normal vector*

$$\mathbf{N} := \frac{\mathbf{r}_u \times \mathbf{r}_v}{|\mathbf{r}_u \times \mathbf{r}_v|}.$$

Clearly, at each point $p \in X$, there are two unit vectors orthogonal to $T_p X$; a choice of one of them is called an *orientation* of X at p . Then, an *orientation* of X is a continuous choice of orientation at all points, *i.e.*, a unit normal field. As just explained, a regular parametrization defines an orientation. Therefore, locally each surface is orientable.

An orientation of X is the same as a “continuous” orientation of all tangent planes: the convention is that

$$((\text{positive basis for } TX), \mathbf{N}) = (\text{positive basis for } \mathbb{R}^3),$$

so that, by definition, $\mathbf{r}_u, \mathbf{r}_v$ is a positive basis for TX .

A change of variables, such as $u \leftrightarrow v$, may change the orientation. If X is oriented, we stick to *positive* parametrizations only.

Remark 5.1 (digression). An orientation of a *vector bundle* (a continuous family of vector spaces) of dimension n is a non-vanishing section of its *determinant line bundle*, *i.e.*, highest exterior power $\bigwedge^n$ (see §4.1.1). An orientation of a surface can be defined in many other ways, including algebraic topology (for which we do not need to assume that the surface is regular or even sits in $\mathbb{R}^3$).

Unlike curves, *not all surfaces are orientable!* The best known is the *Möbius strip* M (Example 7.17 below). We conclude that M has no global parametrization!

There also are *closed* (compact without boundary) non-orientable surfaces, the best known being the *Klein bottle*. A somewhat simpler example is the so-called *real projective plane* P , even though its images usually look worse: they have irregular points. An immersion of P to $\mathbb{R}^3$, the so-called *Boy’s surface*, can be observed in the *Mathematisches Forschungsinstitut Oberwolfach*, courtesy of Mercedes-Benz (though I always thought it was Volkswagen).

All these surfaces have bad points, at least self-intersections. A simple exercise in algebraic topology states that *a closed non-orientable surface cannot be embedded to $\mathbb{R}^3$* . Therefore, we will not encounter them in our course.

5.2. Normal and geodesic curvature; the second fundamental form. Let $\mathbf{r}$ be a *positive* parametrization of an *oriented* surface X .

Pick a point $p \in X$ and consider a curve $\tilde{\gamma} := \mathbf{r} \circ \gamma: [-\varepsilon, \varepsilon] \rightarrow X$, $\tilde{\gamma}(0) = p$. We assume the parametrization by the arc length (letting $' := d/ds$), so that $\mathbf{l}_p(\gamma') = 1$. In addition to the unit tangent vector $\mathbf{t}$ at p , we have the curvature vector $\kappa \mathbf{n}$; in general, it is neither tangent nor normal to X . Hence, we can consider its

- *normal projection* $\kappa_n := \kappa \mathbf{n} \cdot \mathbf{N}$ and
- *tangential projection* $\kappa_g \mathbf{n}_g \in T_p X$.

The latter belongs to the intrinsic geometry of X , and we leave it for now.

Definition 5.2. The *normal curvature* of a curve $\tilde{\gamma}$ on X is $\kappa_n := \kappa \mathbf{n} \cdot \mathbf{N} \in \mathbb{R}$.

Going back to §2.2, we find

$$\kappa_n = \mathbf{t}' \cdot \mathbf{N} = (u' \mathbf{r}_u + v' \mathbf{r}_v)' \cdot \mathbf{N} = (u'^2 \mathbf{r}_{uu} + 2u'v' \mathbf{r}_{uv} + v'^2 \mathbf{r}_{vv} + \dots) \cdot \mathbf{N},$$

where the multiples of $\mathbf{r}_u, \mathbf{r}_v$ have been omitted as orthogonal to $\mathbf{N}$. We conclude that *the normal curvature does not depend on $\tilde{\gamma}$* ; it only depends on the tangent vector $\mathbf{t}$ and is given by

$$\kappa_n(\mathbf{t}) = \mathbb{I}_p(\mathbf{t}),$$

where the coefficients of the *second fundamental form* $\mathbb{I}_p$ are given by

$$\begin{aligned} e &:= \mathbf{r}_{uu} \cdot \mathbf{N} = -\mathbf{r}_u \cdot \mathbf{N}_u, \\ f &:= \mathbf{r}_{uv} \cdot \mathbf{N} = -\mathbf{r}_u \cdot \mathbf{N}_v = -\mathbf{r}_v \cdot \mathbf{N}_u, \\ g &:= \mathbf{r}_{vv} \cdot \mathbf{N} = -\mathbf{r}_v \cdot \mathbf{N}_v. \end{aligned} \tag{5.3}$$

(The alternative expressions, to be used later, are obtained by differentiating the obvious $\mathbf{r}_u \cdot \mathbf{N} = \mathbf{r}_v \cdot \mathbf{N} = 0$.) In other words, we can speak about the *normal curvature of X in the direction of a unit vector $\mathbf{t} \in T_x X$* :

$$\kappa_n((\alpha, \beta)) = e\alpha^2 + 2f\alpha\beta + g\beta^2 \quad \text{provided that} \quad \mathbb{I}_p((\alpha, \beta)) = 1.$$

Extending this to all, not necessarily unit vectors, we get

$$\kappa_n(\mathbf{v}) = \frac{\mathbb{I}_p(\mathbf{v})}{\mathbb{I}_p(\mathbf{v})}, \quad \mathbf{v} \neq 0.$$

As usual, the coefficients e, f, g depend on the parametrization; however, the symmetric bilinear form $\mathbb{I}$ on TX (not necessarily positive definite) is a well-defined geometric object: the proof is in the purely geometric construction.

The prefix “normal” refers to the normal projection of the curvature vector $\kappa \mathbf{n}$. Alternatively, we can consider the *normal section* of X , *i.e.*, the curve cut off X by the plane $\Pi := \text{span}(\mathbf{t}, \mathbf{N})$. This is a planar curve and its curvature vector lies in Π ; being orthogonal to $\mathbf{t}$, it is a multiple of $\mathbf{N}$. Thus, orienting Π so that $\mathbf{N} = \mathbf{n}_s$, we see that $\kappa_n(\mathbf{t})$ is the signed curvature of the normal section through $\mathbf{t}$.

5.3. Principal curvatures and directions. According to linear algebra, in the presence of an inner product $\mathbb{I}_p$, any symmetric bilinear form $\mathbb{I}_p$ is *diagonalizable* (in an essentially unique way), *i.e.*, it has diagonal Gram matrix in an appropriate orthonormal basis. One way to explain this is to observe that there is a unique linear operator $W: TX \rightarrow TX$ such that $\mathbb{I}_p(\mathbf{u}, \mathbf{v}) = \mathbb{I}_p(W(\mathbf{u}), \mathbf{v})$. Since $\mathbb{I}_p$ is symmetric, W is self-adjoint and, as such, it has real eigenvalues and orthogonal eigenvectors. (In §5.4 below this operator is described geometrically.) To simplify matters, let us temporarily denote the matrices of $\mathbb{I}_p, \mathbb{I}_p, W$ by the same letters:

$$\mathbb{I}_p = \begin{bmatrix} E & F \\ F & G \end{bmatrix}, \quad \mathbb{I}_p = \begin{bmatrix} e & f \\ f & g \end{bmatrix}, \quad W = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}.$$

We want

$$\mathbf{u}^T \mathbb{I}_p \mathbf{v} = \mathbf{u}^T W^T \mathbb{I}_p \mathbf{v}, \quad \text{hence,} \quad W = (\mathbb{I}_p \mathbb{I}_p^{-1})^T = \mathbb{I}_p^{-1} \mathbb{I}_p. \tag{5.4}$$

This operator is self-adjoint *with respect to $\mathbb{I}_p$* :

$$W^T \mathbb{I}_p = \mathbb{I}_p W \quad (\text{both equal } \mathbb{I}_p).$$

Geometrically, the eigenvalues of W are the minimal and maximal values taken by $\mathbb{I}_p$ on the unit circle $\mathbb{I}_p(\mathbf{v}) = 1$. ([Lagrange multipliers](#), calculus 102, will help.)

Definition 5.5. The eigenvalues k_1, k_2 of the above operator W are called the *principal curvatures* of X at p , and the directions of its eigenvectors are the *principal directions*. The latter are well defined and orthogonal whenever $k_1 \neq k_2$; otherwise (an *umbilical point*, $k_1 = k_2$), any direction is principal.

Back to linear algebra, k_1, k_2 are the roots of the *characteristic equation* of W :

$$k^2 - 2Hk + K = 0,$$

where

$$(5.6) \quad H := \frac{1}{2} \text{trace } W = \frac{1}{2}(k_1 + k_2) = \frac{1}{2} \frac{Eg - 2Ff + Ge}{EG - F^2} \quad \text{is the mean curvature,}$$

$$(5.7) \quad K := \det W = k_1 k_2 = \frac{\det \text{II}}{\det \text{I}} = \frac{eg - f^2}{EG - F^2} \quad \text{is the Gaussian curvature.}$$

To find the principal directions $\mathbf{e}_{1,2}$, we need to honestly compute

$$(5.8) \quad W = \frac{1}{EG - F^2} \begin{bmatrix} -Ff + Ge & -Fg + Gf \\ Ef - Fe & Eg - Ff \end{bmatrix}$$

and solve the linear system

$$(W - k_{1,2} \text{id}_2) \mathbf{e}_{1,2} = 0.$$

This is differential geometry in all its beauty!

The curvatures H, k_1, k_2 are orientation dependent; K is not.

Since $\mathbf{e}_{1,2}$ are orthogonal, any unit vector can be written in the form

$$\mathbf{t} = \mathbf{e}_1 \cos \theta + \mathbf{e}_2 \sin \theta; \quad \text{then,} \quad \kappa_n(\mathbf{t}) = k_1 \cos^2 \theta + k_2 \sin^2 \theta.$$

We conclude once again that, assuming $k_1 \leq k_2$, we have

$$k_1 \leq \kappa_n(\mathbf{t}) \leq k_2.$$

In some cases, we can also define *asymptotic directions* as those vectors $\mathbf{a} \in TX$ for which $\kappa_n(\mathbf{a}) = 0$. They exist only if $K \leq 0$: if $K < 0$, there are two asymptotic directions, if $K = 0$, either there is one (the principal direction $\mathbf{e}_1$ with $k_1 = 0$) or each direction is asymptotic (if $k_1 = k_2 = 0$).

5.4. The Gauss map and Weingarten operator. Let X be an oriented surface and consider the *Gauss map* $\mathbf{N}: X \rightarrow S^2, p \mapsto \mathbf{N}(p)$, to the unit sphere.

The *differential of the Gauss map* is a linear map

$$d\mathbf{N}_p: T_p X \rightarrow T_{\mathbf{N}_p} S^2.$$

We observe that the tangent plane $T_{\mathbf{N}_p} S^2$ to the unit sphere S^2 is that orthogonal to $\mathbf{N}_p$, and so is the tangent plane $T_p X$. Hence, translating the point $\mathbf{N}_p \in \mathbb{R}^3$ back to p , we can identify the two planes (as vector spaces) and regard $d\mathbf{N}_p$ as a linear operator $T_p X \rightarrow T_p X$. Traditionally, they consider

$$W := -d\mathbf{N}_p: T_p X \rightarrow T_p X;$$

this is called the *Weingarten operator*.

Essentially by the definition, *e.g.*, using (3.14) and the coordinate curves,

$$d\mathbf{N}(\mathbf{r}_u) = \mathbf{N}_u, \quad d\mathbf{N}(\mathbf{r}_v) = \mathbf{N}_v.$$

Hence, $d\mathbf{N}$ is self-adjoint,

$$d\mathbf{N}(\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot d\mathbf{N}(\mathbf{v}).$$

It suffices to check this identity on pairs of basis vectors, and the only “nontrivial” pair is $\mathbf{r}_u, \mathbf{r}_v$, where it is given by (5.3):

$$d\mathbf{N}(\mathbf{r}_u) \cdot \mathbf{r}_v = \mathbf{N}_u \cdot \mathbf{r}_v = -\mathbf{N} \cdot \mathbf{r}_{uv} = \mathbf{r}_u \cdot \mathbf{N}_v = \mathbf{r}_u \cdot d\mathbf{N}(\mathbf{r}_v).$$

Hence, the *Weingarten operator* $W := -d\mathbf{N}$ is also self-adjoint and, as such, it gives rise to a symmetric bilinear form

$$(\mathbf{u}, \mathbf{v}) \mapsto W(\mathbf{u}) \cdot \mathbf{v},$$

$$(\mathbf{r}_u, \mathbf{r}_u) \mapsto -\mathbf{N}_u \cdot \mathbf{r}_u = e, \quad (\mathbf{r}_u, \mathbf{r}_v) \mapsto -\mathbf{N}_u \cdot \mathbf{r}_v = f, \quad (\mathbf{r}_v, \mathbf{r}_v) \mapsto -\mathbf{N}_v \cdot \mathbf{r}_v = g,$$

see (5.3). Thus, we obtain the second fundamental form $\mathbb{I}$, implying the $-d\mathbf{N}$ is nothing but the operator W considered in §5.3. This relation between the three fundamental objects is worth an explicit statement:

$$\mathbb{I}(d\mathbf{N}(\mathbf{u}), \mathbf{v}) = -\mathbb{I}(\mathbf{u}, \mathbf{v}) = \mathbb{I}(\mathbf{u}, d\mathbf{N}(\mathbf{v}));$$

$$H = -\frac{1}{2} \text{trace } d\mathbf{N}, \quad K = \det d\mathbf{N}.$$

(Recall that trace and determinant of a linear map make no sense; however, they are well defined for a linear *operator*, *i.e.*, self-map, *e.g.*, as the coefficients of its characteristic polynomial.)

In particular, we see that $K = \det d\mathbf{N}$ describes the distortion of the area under the Gauss map: for a “tiny” patch $R \subset X$ we have

$$(5.9) \quad \text{Area}(\mathbf{N}(R)) \approx K(p) \text{Area}(R), \quad p \in R.$$

Here, we take into account the orientation and consider the signed area. Thus, $\mathbf{N}$ is orientation preserving (resp. reversing) if $K > 0$ (resp. $K < 0$); if $K = 0$, then $d\mathbf{N}$ is degenerate and we cannot compare the orientation.

Assuming X closed (*i.e.*, compact without boundary), we can integrate (5.9) to a precise statement (which otherwise would be derived from the Gauss–Bonnet Theorem 8.1 below and Riemann–Hurwitz formula for $\mathbf{N}: X \rightarrow S^2$).

Theorem 5.10. *For a closed surface X one has*

$$\iint_X K \, d\text{vol} = 4\pi \deg(\mathbf{N}).$$

Here, the *degree* $\deg \mathbf{N}$ is the number of points in the pull-back $\mathbf{N}^{-1}(q) \subset X$ of a generic point (regular value) $q \in S^2$ counted with signs ± 1 according to the orientation. This is well defined, and an easy proof is given by algebraic topology. Roughly, under $\mathbf{N}$ the surface X covers S^2 (of area 4π) $\deg \mathbf{N}$ times; there may be bigger overlaps, but they would cancel due to the opposite orientations.

Theorem 5.10 is a literal analogue of (2.14), where $\text{rot } \gamma$ is nothing but the degree of the Gauss map $\mathbf{t}$. The two are simplest examples of a long series of statements connecting hardcore differential geometry and simple topology.

5.5. Digression: minimal surfaces. Here, we discuss the geometric meaning of the mean curvature H .

Definition 5.11. A surface X is called *minimal* if $H \equiv 0$.

This is an unfortunate classical terminology: in fact, such surfaces are *critical* for the area functional. For the reason of surface tension and such, soap films always take the shape of a minimal surface. (That is, when they can: soap bubbles do not count.) Indeed, common sense suggests that equilibria of a system are the critical

points of its potential energy; however, only the *minima* are stable. Below is a brief naïve explanation of the “minimality” of surfaces with $H \equiv 0$.

Fix a surface X parametrized by $\mathbf{r}$ and consider a *normal variation* thereof, *i.e.*, a family X_t parametrized by

$$\mathbf{r}_t(u, v) := \mathbf{r}(u, v) + th(u, v)\mathbf{N}(u, v),$$

where h is a certain fixed function. (Here and below, t is a parameter and the subscript is *not* the derivative.) We have

$$(\mathbf{r}_t)_u = \mathbf{r}_u + th\mathbf{N}_u + th_u\mathbf{N}, \quad (\mathbf{r}_t)_v = \mathbf{r}_v + th\mathbf{N}_v + th_v\mathbf{N},$$

whereupon, by (5.3),

$$\begin{aligned} E_t &= E - 2thE + O(t^2), & F_t &= F - 2thF + O(t^2), & G_t &= G - 2thG + O(t^2), \\ E_tG_t - F_t^2 &= (EG - F^2)(1 - 4thH + O(t^2)). \end{aligned}$$

Hence,

$$A(t) := \text{Area}(X_t) = \iint_{\Omega} \sqrt{EG - F^2}(1 - 2thH + O(t^2)) \, du \, dv$$

and

$$\frac{d}{dt}A(t)\Big|_{t=0} = - \iint_{\Omega} 2hH\sqrt{EG - F^2} \, du \, dv.$$

(This is called the *Gateaux derivative* of Area in the direction of h ; the latter plays the rôle of a “tangent vector” to the space of surfaces.) As in calculus, for a critical point we want this to be zero in *any* “direction” h ; since we assume everything continuous and $\sqrt{EG - F^2} > 0$, this is true if and only if $H \equiv 0$.

Clearly, if X maximizes/minimizes the area, it has to be critical: for any h , the function $A(t)$ will have a maximum/minimum at $t = 0$ and, thus, $A'(0) = 0$.

5.6. Digression: critical points of the distance function. Similar to curves, at each point $p \in X$ we can define two *radii of curvature* $R_{1,2} := 1/k_{1,2}$ and *centers of curvature* $\mathbf{c}_{1,2} := \mathbf{r} + R_{1,2}\mathbf{N}$ lying on the normal line. Here is an interpretation of these points in terms of the *second derivative test* from calculus 102.

Given a critical point ($f_x = f_y = 0$) of a function $f(x, y)$, we can define the *Hessian matrix*

$$\mathcal{H} := \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}.$$

This matrix has a geometric meaning at a critical point only, in which case it gives rise to a well-defined symmetric bilinear form

$$(\mathbf{u}, \mathbf{v}) \mapsto \mathcal{H}(\mathbf{u}, \mathbf{v}) := \mathbf{u}^T \mathcal{H} \mathbf{v}.$$

This form can be diagonalized (in many ways) and, by *Sylvester’s law of inertia*, it has a well defined *index* $\text{ind } \mathcal{H}$, which is the number of negative squares in (any) diagonal form. In other words, index is the dimension of any maximal negative definite subspace (and it is this observation that proves Sylvester’s theorem). The critical point is called *degenerate* if $\det \mathcal{H} = 0$ or, equivalently, the form $\mathcal{H}$ has a non-zero *radical*

$$(5.12) \quad \text{rad } \mathcal{H} := \{\mathbf{u} \in \mathbb{R}^2 \mid \mathcal{H}(\mathbf{u}, \mathbf{v}) = 0 \text{ for all } \mathbf{v} \in \mathbb{R}^2\}.$$

In calculus textbooks, a critical point of a bivariate function f is a

- minimum, if $\text{ind } \mathcal{H} = 0$ ($\mathcal{H}$ is positive definite), or

- saddle point, if $\text{ind } \mathcal{H} = 1$ ($\mathcal{H}$ is indefinite), or
- maximum, if $\text{ind } \mathcal{H} = 2$ ($\mathcal{H}$ is negative definite);
- the test is inconclusive if the point is *degenerate*, $\det \mathcal{H} = 0$.

The meaning of this is very simple: since $df = 0$, it is the second order term $\mathcal{H}$ that controls the behaviour of f close to the point. Other, higher order terms of the Taylor expansion are negligibly small *unless $\mathcal{H}$ has radical*: in this latter case higher order terms do matter and we need more data. (In the univariate calculus 101 this is known as the higher derivative test; in the multivariate case it would not be easy to state.)

Similar to (2.5), let us fix a point $\mathbf{a} \in \mathbb{R}^3$ and apply this machinery to the *distance function*

$$d(u, v) := |\mathbf{r}(u, v) - \mathbf{a}|^2.$$

A point $p = \mathbf{r}(u, v) \in X$ is critical if and only if, at this point,

$$d_u = \mathbf{r}_u \cdot (\mathbf{r} - \mathbf{a}) = 0, \quad d_v = \mathbf{r}_v \cdot (\mathbf{r} - \mathbf{a}) = 0.$$

In other words, $(\mathbf{r} - \mathbf{a}) \perp T_p X$, or $\mathbf{a}$ lies in the normal to X at p , $\mathbf{r} - \mathbf{a} = -\alpha \mathbf{N}$. Assuming this, the second derivatives are

$$d_{uu} = -\mathbf{r}_{uu} \cdot \alpha \mathbf{N} + \mathbf{r}_u^2 = -\alpha e + E, \quad d_{uv} = -\alpha f + F, \quad d_{vv} = -\alpha g + G.$$

Hence,

$$\mathcal{H} = \mathbf{l}_p - \alpha \mathbf{l}_p = \mathbf{l}_p(\text{id}_2 - \alpha W),$$

and the critical point is degenerate if and only if $\alpha = R_{1,2}$, i.e., $\mathbf{a}$ is one of the principal centers of curvature at p .

We can say more. The eigenvalues of (the self-adjoint operator $\mathbf{l}_p^{-1} \mathcal{H}$ associated with) $\mathcal{H}$ are $1 - \kappa_{1,2} \alpha$, and $\text{ind } \mathcal{H}$ is the number of negative eigenvalues. In other words, $\text{ind } \mathcal{H}$ is the number of principal centers of curvature between $\mathbf{a}$ and p .

5.7. Compatibility equations. Instead of the natural Frenet frame $\mathbf{t}, \mathbf{n}, \mathbf{b}$ for curves, for surfaces we have a not so natural frame $\mathbf{r}_u, \mathbf{r}_v, \mathbf{N}$. Still, we want an analogue of the Frenet equations (2.8) expressing the derivatives of the basis vectors in terms of the same basis. This analogue is

$$\begin{aligned} \mathbf{N}_u &= -w_{11} \mathbf{r}_u - w_{21} \mathbf{r}_v, \\ \mathbf{N}_v &= -w_{12} \mathbf{r}_u - w_{22} \mathbf{r}_v, \\ \mathbf{r}_{uu} &= \Gamma_{11}^1 \mathbf{r}_u + \Gamma_{11}^2 \mathbf{r}_v + e \mathbf{N}, \\ \mathbf{r}_{uv} &= \Gamma_{12}^1 \mathbf{r}_u + \Gamma_{12}^2 \mathbf{r}_v + f \mathbf{N}, \\ \mathbf{r}_{vv} &= \Gamma_{22}^1 \mathbf{r}_u + \Gamma_{22}^2 \mathbf{r}_v + g \mathbf{N}, \end{aligned} \tag{5.13}$$

The first two equations are the definition of the Weingarten operator as $W = -d\mathbf{N}$, and their coefficients are found from (pay attention to the indexing!)

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} = \begin{bmatrix} E & F \\ F & G \end{bmatrix}^{-1} \begin{bmatrix} e & f \\ f & g \end{bmatrix},$$

see (5.8). The $\mathbf{N}$ -components of the last three equations are immediate from (5.3). The other coefficients are called the *Christoffel symbols*; their indexing becomes more apparent if we rename $u \mapsto 1$, $v \mapsto 2$ and rewrite them in the form

$$\mathbf{r}_{ij} = \Gamma_{ij}^1 \mathbf{r}_1 + \Gamma_{ij}^2 \mathbf{r}_2 + \dots$$

In particular, almost by definition, they are symmetric: $\Gamma_{ij}^k = \Gamma_{ji}^k$.

To find the Christoffel symbols, we differentiate the entries of (4.2) to arrive at

$$(5.14) \quad \begin{aligned} \mathbf{r}_{uu} \cdot \mathbf{r}_u &= \frac{1}{2}E_u, & \mathbf{r}_{uv} \cdot \mathbf{r}_u &= \frac{1}{2}E_v, & \mathbf{r}_{vv} \cdot \mathbf{r}_u &= F_v - \frac{1}{2}G_u, \\ \mathbf{r}_{uu} \cdot \mathbf{r}_v &= F_u - \frac{1}{2}E_v, & \mathbf{r}_{uv} \cdot \mathbf{r}_v &= \frac{1}{2}G_u, & \mathbf{r}_{vv} \cdot \mathbf{r}_v &= \frac{1}{2}G_v. \end{aligned}$$

On the other hand, the same products are found by multiplying (5.13) by $\mathbf{r}_u, \mathbf{r}_v$. Equating, we obtain three linear systems in Γ_{**}^* (all three with the same coefficient matrix $\mathbb{I}$), which are easily solved:

$$(5.15) \quad \begin{bmatrix} \Gamma_{11}^1 & \Gamma_{12}^1 & \Gamma_{22}^1 \\ \Gamma_{11}^2 & \Gamma_{12}^2 & \Gamma_{22}^2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} E & F \\ F & G \end{bmatrix}^{-1} \begin{bmatrix} E_u & E_v & 2F_v - G_u \\ 2F_u - E_v & G_u & G_v \end{bmatrix}.$$

Needless to say that all these equations in their general form can only be used at times of total desperation, if you are absolutely out of better ideas!

The following statement is the most important take-out from (5.15); it will be explored in §6.

Corollary 5.16. *The Christoffel symbols Γ_{ij}^k belong to the intrinsic geometry, i.e., they are expressed in terms of E, F, G (and their derivatives) only.* $\triangleleft$

The formulas simplify in various special coordinate systems; e.g., in (4.26),

$$(5.17) \quad \Gamma_{11}^1 = \Gamma_{11}^2 = \Gamma_{12}^1 = 0, \quad \Gamma_{12}^2 = \frac{G_u}{2G}, \quad \Gamma_{22}^1 = -\frac{1}{2}G_u, \quad \Gamma_{22}^2 = \frac{G_v}{2G}$$

or in (4.25),

$$(5.18) \quad \Gamma_{11}^1 = \Gamma_{12}^2 = -\Gamma_{22}^1 = \frac{1}{2}\rho_u, \quad -\Gamma_{11}^2 = \Gamma_{12}^1 = \Gamma_{22}^2 = \frac{1}{2}\rho_v.$$

Equations (5.13) cannot be independent: we know that mixed partial derivatives are equal (e.g., $\mathbf{N}_{uv} = \mathbf{N}_{vu}$, $\mathbf{r}_{uv} = \mathbf{r}_{vu} = \mathbf{r}_{vuu}$, etc.) Differentiating (5.13) and substituting the same equations (5.13) to the new derivatives in the right hand side, we obtain a number of equations. Some of them are trivial, some have already been accounted for (e.g., when asserting that $\mathbb{I}$ is symmetric), some are new. In fact, we arrive at one *remarkable theorem* and two extra *compatibility equations*; all others are consequences thereof.

Theorem 5.19 (Mainardi–Codazzi equations). *One has*

$$\begin{aligned} e_v - f_u &= e\Gamma_{12}^1 + f(\Gamma_{12}^2 - \Gamma_{11}^1) - g\Gamma_{11}^2, \\ f_v - g_u &= e\Gamma_{22}^1 + f(\Gamma_{22}^2 - \Gamma_{12}^1) - g\Gamma_{12}^2. \end{aligned}$$

Proof. Compare the $\mathbf{N}$ -components in $(\mathbf{r}_{uu})_v = (\mathbf{r}_{uv})_u$ and $(\mathbf{r}_{uv})_v = (\mathbf{r}_{vv})_u$. $\square$

Theorem 5.20 (Gauss' Theorema Egregium). *One has*

$$\begin{aligned} EK &= \Gamma_{11}^1\Gamma_{12}^2 + \Gamma_{11}^2\Gamma_{22}^2 + (\Gamma_{11}^2)_v - \Gamma_{12}^1\Gamma_{11}^2 - \Gamma_{12}^2\Gamma_{12}^2 - (\Gamma_{12}^2)_u, \\ GK &= \Gamma_{22}^1\Gamma_{11}^1 + \Gamma_{22}^2\Gamma_{12}^1 + (\Gamma_{12}^1)_u - \Gamma_{12}^2\Gamma_{22}^1 - \Gamma_{12}^1\Gamma_{12}^1 - (\Gamma_{12}^1)_v, \\ FK &= \Gamma_{12}^2\Gamma_{12}^1 + (\Gamma_{12}^1)_u - \Gamma_{11}^2\Gamma_{22}^1 - (\Gamma_{11}^1)_v \\ &= \Gamma_{12}^1\Gamma_{12}^2 + (\Gamma_{12}^2)_v - \Gamma_{22}^1\Gamma_{11}^2 - (\Gamma_{22}^2)_u. \end{aligned}$$

Proof. For the first one, compare the $\mathbf{r}_v$ -components in $(\mathbf{r}_{uu})_v = (\mathbf{r}_{uv})_u$. Others are proved similarly. $\square$

It can be shown that, in the presence of [Theorem 5.19](#) and definition (5.7) of K , all relations in [Theorem 5.20](#) follow from any of the first two.

Remark 5.21. In view of (5.15) and (5.7), Theorems 5.19 and 5.20 can be regarded as certain relations among E, F, G, e, f, g . An analogue of the existence and uniqueness theorem for curves states that these are, indeed, all relations.

The value of [Theorem 5.20](#) can hardly be overestimated: the Gaussian curvature K , originally defined as a *curvature*, is part of intrinsic geometry! Some geometric properties may force a surface to be curved and, *vice versa*, a curved surface is bound to have non-trivial intrinsic geometry. *E.g.*, unlike cylinder or cone, a sphere with $K = 1/R^2$ cannot be flat: its intrinsic geometry differs from that of the plane.

[Theorem 5.20](#) simplifies in coordinates (4.26): in view of (5.17),

$$(5.22) \quad K = -\frac{(\sqrt{G})_{uu}}{\sqrt{G}} = \frac{G_u^2}{4G^2} - \frac{G_{uu}}{2G}.$$

In (4.25), it simplifies to

$$(5.23) \quad K = -\frac{1}{2}e^{-\rho}(\rho_{uu} + \rho_{vv}).$$

5.8. Example: yet another characterization of spheres. We illustrate the machinery of this section applied to simple problems of exam/homework level.

In most cases, instead of the ready “formulas”, it is easier to use the geometric definitions of the objects involved.

Example 5.24. *A surface X is (part of) a plane if and only if $k_1 \equiv k_2 \equiv 0$.*

In one direction, the statement is a straightforward computation for the plane. In the opposite, assume that $k_1 = k_2 = 0$. Since $k_{1,2}$ are the eigenvalues of W , this implies that $W = -d\mathbf{N} = 0$, *i.e.*, $\mathbf{N}_u = \mathbf{N}_v = 0$. The latter holds at each point, and we conclude that $\mathbf{N} = \text{const}$.

Then,

$$(\mathbf{r} \cdot \mathbf{N})_u = \mathbf{r}_u \cdot \mathbf{N} = 0, \quad (\mathbf{r} \cdot \mathbf{N})_v = \mathbf{r}_v \cdot \mathbf{N} = 0,$$

and X lies in a plane $\mathbf{r} \cdot \mathbf{N} = \text{const}$, see (1.1). $\square$

Example 5.25. *A surface X is (part of) a sphere (obviously, of radius R) if and only if $k_1 \equiv k_2 = R^{-1} = \text{const}$.*

Interpreting normal curvatures as those of normal sections, we conclude that a sphere of radius R does have $k_1 = k_2 = R^{-1}$. In the opposite direction, this time $W = R^{-1} \text{id}_2$ is a scalar operator. Our educated guess is that the center of the sphere in question is $\mathbf{c} := \mathbf{r} + R\mathbf{N}$, and we only need to show that $\mathbf{c} = \text{const}$. (Pay attention to the “+”: the curvatures are positive if the sphere is oriented by the *inward* normals.) As in the case of curves, using $R = \text{const}$, we have

$$\mathbf{c}_u = \mathbf{r}_u + R\mathbf{N}_u = \mathbf{r}_u + R d\mathbf{N}(\mathbf{r}_u) = \mathbf{r}_u - R \cdot W(\mathbf{r}_u) = 0,$$

and similarly $\mathbf{c}_v = 0$. $\square$

Example 5.26. *A surface X is (part of) a sphere or a plane if and only if each point of X is umbilical.*

This is the most interesting one. In view of Examples 5.24 and 5.25, it suffices to show that, if $k_1 = k_2$ at each point, this common value $k := k_1 = k_2$ is constant.

As in [Example 5.25](#), $W = k \text{id}_2$ is a scalar operator, *i.e.*, by its definition [\(5.4\)](#), we have $\mathbb{I} = k\mathbb{I}$:

$$e = kE, \quad f = kF, \quad g = kG.$$

Substituting this to Mainardy–Codazzi [Theorem 5.19](#) and using the linear equations that lead us to [\(5.15\)](#), *viz.*

$$E\Gamma_{12}^1 + F\Gamma_{12}^2 = \mathbf{r}_{uv} \cdot \mathbf{r}_u = \frac{1}{2}E_v, \quad F\Gamma_{11}^2 + G\Gamma_{11}^2 = \mathbf{r}_{uu} \cdot \mathbf{r}_v = F_u - \frac{1}{2}E_v,$$

etc., *cf.* [\(5.14\)](#), we arrive at

$$k_v E - k_u F = 0, \quad k_v F - k_u G = 0.$$

This system has determinant $F^2 - EG \neq 0$; hence, $k_u = k_v = 0$. $\square$

Remark 5.27. Of course, there is another way: $d\mathbf{N} = -W = -k \text{id}$ means that

$$\mathbf{N}_u = -k\mathbf{r}_u, \quad \mathbf{N}_v = -k\mathbf{r}_v.$$

Differentiate with respect to v and u , respectively,

$$k_v \mathbf{r}_u + \mathbf{r}_{uv} = -\mathbf{N}_{uv} = -\mathbf{N}_{vu} = k_u \mathbf{r}_v + \mathbf{r}_{vu},$$

and recall that $\mathbf{r}_{uv} = \mathbf{r}_{vu}$ and that $\mathbf{r}_u, \mathbf{r}_v$ are linearly independent.

5.9. Digression: special parametrizations. To conclude this section, we invoke [Theorem 4.13](#) with an appropriate pair $\mathbf{u}, \mathbf{v}$ of vector fields to prove the (local) existence of various special parametrizations of a regular surface X . Recall that the *coordinate lines* are mereli the curves $\mathbf{r}(u, \text{const})$ and $\mathbf{r}(\text{const}, v)$ on X .

Other special cases, such as isothermal coordinated [\(4.25\)](#) or *Tchebychev nets*

$$F \equiv G \equiv 1, \quad F = \cos \theta(u, v),$$

depend on PDE's and are beyond the scope of this text.

Theorem 5.28. *A neighborhood of each point $p \in X$ has a parametrization whose coordinate lines are orthogonal, $F \equiv 0$.*

Proof. Take for $\mathbf{u}, \mathbf{v}$ the two fields discussed at the beginning of [§4.5](#). $\square$

In [§6.4.1](#) below we will prove a stronger version of this theorem, ensuring that $F \equiv 0$ and $G \equiv 1$. A quick glance at the statements in this section suggests that, in general, we can control up to two of the six parameters E, F, G, e, f, g .

Theorem 5.29. *A neighborhood of each non-umbilical point $p \in X$ has a parametrization whose coordinate lines are lines of curvature, $F \equiv f \equiv 0$.*

Proof. Take for $\mathbf{u}, \mathbf{v}$ unit eigenvectors $\mathbf{e}_{1,2}$ of the Weingarten operator $W = -d\mathbf{N}$, see [\(5.4\)](#), ordered so that, *e.g.*, $k_1 < k_2$. $\square$

Remark 5.30. The assumption that p should be non-umbilical is not a mere deficiency of the proof, inability to separate the two fields at p itself. The behaviour of the lines of curvature about an [umbilical point](#) may be quite sophisticated.

Theorem 5.31. *A neighborhood of each hyperbolic point $p \in X$ (*i.e.*, $K(p) < 0$) has a parametrization whose coordinate lines are asymptotic, $e \equiv g \equiv 0$.*

Proof. Take for $\mathbf{u}, \mathbf{v}$ unit vectors in the two asymptotic directions. $\square$

6. INTRINSIC GEOMETRY CONTINUED

We are ready to explore the intrinsic geometry of a surface to full extent.

On the plane, lines are the “straightest” curves ($\kappa_s \equiv 0$) and, at the same time, the shortest ones (distance is realized by the unique line segment connecting two points). On a surface, we will define the so-called *geodesics* as straightest curves and discuss the extent to which they are shortest.

6.1. The covariant derivative. At this point, in spite of what has been said in §3.4, we do want to compare tangent vectors at distinct points. On the plane $\mathbb{R}^2$, where all tangent planes are naturally identified with $\mathbb{R}^2$ itself, a vector field is constant (we will say *parallel*) if and only if all its derivatives vanish. Hence, we need a sort of derivative that would take a tangent vector field and, unlike other derivatives known so far, produce a *tangent* vector.

Consider a curve γ on X and a tangent (to X , not necessarily γ) vector field $\mathbf{u}$ along γ ; in local coordinates, we can write it as

$$\mathbf{u}(t) = \alpha(t)\mathbf{r}_u + \beta(t)\mathbf{r}_v.$$

The *covariant derivative* of $\mathbf{u}$ at the point $p := \gamma(0)$ is the tangent vector

$$\left. \frac{D\mathbf{u}}{dt} \right|_{t=0} := \text{proj}_{T_p X} \left. \frac{d\mathbf{u}}{dt} \right|_{t=0},$$

where proj is the orthogonal projection. That is, we take the ordinary derivative and do the best that we can: project it back to the tangent plane. Assuming (with the usual abuse of language) that

$$\gamma: t \mapsto (u(t), v(t)),$$

we have

$$\begin{aligned} \mathbf{u}' &= \alpha' \mathbf{r}_u + \beta' \mathbf{r}_v + \alpha(\mathbf{r}_u)' + \beta(\mathbf{r}_v)' \\ &= \alpha' \mathbf{r}_u + \beta' \mathbf{r}_v + \alpha(u' \mathbf{r}_{uu} + v' \mathbf{r}_{uv}) + \beta(u' \mathbf{r}_{uv} + v' \mathbf{r}_{vv}) \end{aligned}$$

and, upon projecting to TX , *i.e.*, disregarding the $\mathbf{N}$ -component in (5.13),

$$\begin{aligned} (6.1) \quad \frac{D\mathbf{u}}{dt} &= (\alpha' + \alpha u' \Gamma_{11}^1 + \alpha v' \Gamma_{12}^1 + \beta u' \Gamma_{12}^1 + \beta v' \Gamma_{22}^1) \mathbf{r}_u \\ &\quad + (\beta' + \alpha u' \Gamma_{11}^2 + \alpha v' \Gamma_{12}^2 + \beta u' \Gamma_{12}^2 + \beta v' \Gamma_{22}^2) \mathbf{r}_v. \end{aligned}$$

We make two extremely important observations:

- (1) in spite of our external definition, the covariant derivative is part of the intrinsic geometry of X : it depends on E, F, G only, see (5.15);
- (2) assuming that $\mathbf{u}$ is defined on a whole neighborhood of γ , *i.e.*, α, β are functions of u, v , we can substitute

$$\alpha' = \alpha_u u' + \alpha_v v' = \partial \alpha / \partial \gamma', \quad \beta' = \beta_u u' + \beta_v v' = \partial \beta / \partial \gamma'$$

in (6.1) to see that $D\mathbf{u}/dt$ depends on the tangent vector $\mathbf{v} := \gamma'$ only.

In view of item (2), we can speak about the covariant derivative

$$\frac{D\mathbf{u}}{d\mathbf{v}} \in T_p X, \quad \mathbf{v} \in T_p X,$$

of a tangent vector field with respect to a tangent vector. The formula is obtained by substituting to (6.1) the above expressions for α', β' and replacing (u', v') with the coordinates $\bar{\alpha}, \bar{\beta}$ of $\mathbf{v}$:

$$(6.2) \quad \begin{aligned} \frac{D\mathbf{u}}{\partial \mathbf{v}} = & (\alpha_u \bar{\alpha} + \alpha_v \bar{\beta} + \alpha \bar{\alpha} \Gamma_{11}^1 + \alpha \bar{\beta} \Gamma_{12}^1 + \beta \bar{\alpha} \Gamma_{12}^1 + \beta \bar{\beta} \Gamma_{22}^1) \mathbf{r}_u \\ & + (\beta_u \bar{\alpha} + \beta_v \bar{\beta} + \alpha \bar{\alpha} \Gamma_{11}^2 + \alpha \bar{\beta} \Gamma_{12}^2 + \beta \bar{\alpha} \Gamma_{12}^2 + \beta \bar{\beta} \Gamma_{22}^2) \mathbf{r}_v. \end{aligned}$$

It is worth emphasizing that these lengthy formulas do define a vector, *i.e.*, they transform appropriately under the change of coordinates. Checking this directly would be quite tedious: one would also have to take into account the change of Γ_{ij}^k . But we merely refer to the fact that our definition was geometric.

Remark 6.3. Concerning item (1), of course there is a way to define the covariant derivative intrinsically: it is the so-called *Levi-Cevita connection*, which is the only connection that is reasonably symmetric, see (6.7) below, and compatible with the first fundamental form I, see (6.6) below. The term *connection* in this context means a sort of derivation (as defined by (6.4) and (6.6) below) for vector fields rather than functions; in a sense, it lets us connect nearby tangent planes.

It is an object of a totally new nature, not even a tensor, so we skip further details. Accidentally, in spite of the notation, Γ_{**}^* is not a tensor, either: by definition, these numbers are the coefficients of the Levi-Cevita connection, and the change of variables formulas are quite different. (They should follow from (4.4) and (5.15), but we never bothered to write them down, nor we ever will.)

The proof of existence and uniqueness is, essentially, our computation of Γ_{**}^* .

The following properties are immediate from the definition or from (6.2).

$$(6.4) \quad \frac{D\mathbf{u}}{\partial \mathbf{v}} \text{ is } \mathbb{R}\text{-linear with respect to } \mathbf{u} \text{ and } \mathbf{v};$$

it is subject to all thinkable forms of the Leibnitz rule:

$$(6.5) \quad \frac{D}{\partial \mathbf{v}}(f\mathbf{u}) = \mathbf{v}(f)\mathbf{u} + f \frac{D\mathbf{u}}{\partial \mathbf{v}};$$

$$(6.6) \quad \frac{\partial}{\partial \mathbf{w}}(\mathbf{u} \cdot \mathbf{v}) = \frac{D\mathbf{u}}{\partial \mathbf{w}} \cdot \mathbf{v} + \mathbf{u} \cdot \frac{D\mathbf{v}}{\partial \mathbf{w}}$$

(in local coordinates, $\cdot$ should be understood as I); it is as symmetric as possible:

$$(6.7) \quad \frac{D\mathbf{v}}{\partial \mathbf{u}} - \frac{D\mathbf{u}}{\partial \mathbf{v}} = [\mathbf{u}, \mathbf{v}]$$

(the Lie bracket): this follows directly from (6.2) and (4.11).

We will use all kinds of self-explanatory notation, such as

$$\frac{D}{\partial u} := \frac{D}{\partial \mathbf{r}_u}, \quad \frac{D}{\partial v} := \frac{D}{\partial \mathbf{r}_v}, \quad \frac{D}{dt} := \frac{D}{\partial \gamma'} \quad \text{for a curve } \gamma,$$

but we never use the awkward $\odot$ from the textbook. The modern notation found in the literature, especially in the general context of connections, is $\nabla_{\mathbf{v}} := D/\partial \mathbf{v}$.

The next statement shows that the covariant derivative is not quite symmetric.

Theorem 6.8. *For any vector field $\mathbf{w}$, one has*

$$\frac{D}{\partial v} \frac{D}{\partial u} \mathbf{w} - \frac{D}{\partial u} \frac{D}{\partial v} \mathbf{w} = K(\mathbf{r}_u \times \mathbf{r}_v) \times \mathbf{w}.$$

The proof of this theorem is a tedious straightforward computation eventually referring to [Theorem 5.20](#) (*cf.* also [§6.1.1](#)). Instead, we discuss a few peculiarities other than yet another reappearance of the omnipresent Gaussian curvature.

Remark 6.9. The right hand side of [Theorem 6.8](#) is a bit eclectic, as it refers to the cross product which is not part of our current framework. However, it *is* a tangent vector. One can either refer to [\(1.7\)](#) and rewrite it in the form

$$K(\mathbf{w} \cdot \mathbf{r}_u)\mathbf{r}_v - K(\mathbf{w} \cdot \mathbf{r}_v)\mathbf{r}_u$$

difficult to remember or observe that it obviously is

$$K|\mathbf{r}_u \times \mathbf{r}_v|\mathbf{w}^\perp = K\sqrt{EG - F^2}\mathbf{w}^\perp.$$

For the rest of this text, we reserve the notation

$$(6.10) \quad \mathbf{w}^\perp := \mathbf{N} \times \mathbf{w}$$

for the vector obtained from $\mathbf{w}$ by the rotation in the tangent plane through $\pi/2$ in the positive direction; recall that the direction (orientation of the tangent plane) is essentially determined by the order of (u, v) .

Remark 6.11. Another remarkable fact about [Theorem 6.8](#) is that its left hand side is a differential operator, it needs a vector *field*, whereas the right hand side is a tensor: its value at a point depends on the value of $\mathbf{w}$ at that point only! This tensor is very symmetric, see [\(6.13\)](#), [\(6.14\)](#) below.

6.1.1. *An “easy” proof of [Theorem 6.8](#).* The proof below is an easy computation, but it is a bit eclectic, appealing to normal vectors. Decompose a vector $\mathbf{w} \in \mathbb{R}^3$

$$\mathbf{w} = \mathbf{w}_\nu + \mathbf{w}_\tau, \quad \mathbf{w}_\nu := (\mathbf{w} \cdot \mathbf{N})\mathbf{N}, \quad \mathbf{w}_\tau := \mathbf{w} - \mathbf{w}_\nu \in TX,$$

into its *normal* and *tangential components*. Since the partials $\mathbf{w}_{uv} = \mathbf{w}_{vu}$ are equal,

$$0 = (\mathbf{w}_{vu})_\tau - (\mathbf{w}_{uv})_\tau = \left(\frac{\partial}{\partial v}(\mathbf{w}_u)_\tau - \frac{\partial}{\partial u}(\mathbf{w}_v)_\tau \right)_\tau + \left(\frac{\partial}{\partial v}(\mathbf{w}_u)_\nu - \frac{\partial}{\partial u}(\mathbf{w}_v)_\nu \right)_\tau.$$

The first term is the left hand side of [Theorem 6.8](#); hence, we need to compute the red term. It is immediate that the result is a tensor: if $\mathbf{w} = \alpha\mathbf{r}_u + \beta\mathbf{r}_v$, each time the terms involving the derivatives of α, β are on the “wrong” side, tangential after the first differentiation and normal after the second. Hence, it suffices to consider $\mathbf{w} = \mathbf{r}_u$ or $\mathbf{r}_v$, for which we only need the “easy” part of [\(5.13\)](#). Assume $\mathbf{w} = \mathbf{r}_u$ (for $\mathbf{r}_v$ the computation is literally the same). Then,

$$\begin{aligned} \frac{\partial}{\partial v}(\mathbf{r}_{uu})_\nu - \frac{\partial}{\partial u}(\mathbf{r}_{vu})_\nu &= \frac{\partial}{\partial v}(e\mathbf{N}) - \frac{\partial}{\partial u}(f\mathbf{N}) = e\mathbf{N}_v - f\mathbf{N}_u + \overbrace{(e_v - f_u)\mathbf{N}}^{0 \text{ in } TX} \\ &\approx (fw_{11} - ew_{12})\mathbf{r}_u + (fw_{21} - ew_{22})\mathbf{r}_v = \frac{eg - f^2}{EG - F^2}(F\mathbf{r}_u - E\mathbf{r}_v), \end{aligned}$$

as stated, see [\(5.8\)](#), [\(5.7\)](#), and the first expression in [Remark 6.9](#). $\square$

Remark 6.12. A big disadvantage of this proof is the fact that it does not apply to abstract surfaces like the one considered in [§4.4](#). However, for such surfaces we have no $\mathbb{I}$, hence neither W or k_1, k_2, H . Most importantly, we do not have an *independent* definition of the Gaussian curvature K . Therefore, we can *define* K using [Theorem 6.8](#) (or, better yet, [§6.1.2](#) below to make sure that the definition is independent of the parametrization); then, [Theorem 5.20](#) can be derived by the above mentioned tedious computation. [Theorem 5.19](#) no longer makes sense.

6.1.2. *Digression: the curvature of a connection.* In general, for a pair $\mathbf{u}, \mathbf{v}$ of not necessarily constant vector fields, equation (4.11) suggests that a correction term involving $[\mathbf{u}, \mathbf{v}]$ should be present. Indeed, we have

$$\frac{D}{\partial \mathbf{u}} \frac{D}{\partial \mathbf{v}} \mathbf{w} - \frac{D}{\partial \mathbf{v}} \frac{D}{\partial \mathbf{u}} \mathbf{w} - \frac{D \mathbf{w}}{\partial [\mathbf{u}, \mathbf{v}]} = K(\mathbf{w} \times (\mathbf{u} \times \mathbf{v})).$$

The left hand side makes sense for any connection, and the right hand side is always a tensor, called the *curvature tensor* of that connection. Usually, they add an extra parameter $\mathbf{z}$ and consider the 4-valent tensor

$$(6.13) \quad R(\mathbf{u}, \mathbf{v}; \mathbf{w}, \mathbf{z}) \stackrel{\text{for us}}{=} K \mathbf{z} \cdot (\mathbf{w} \times (\mathbf{u} \times \mathbf{v})) \stackrel{(1.7)}{=} -K \begin{vmatrix} \mathbf{w} \cdot \mathbf{u} & \mathbf{z} \cdot \mathbf{v} \\ \mathbf{w} \cdot \mathbf{v} & \mathbf{z} \cdot \mathbf{u} \end{vmatrix}$$

(I hope the sign is correct), which is very symmetric:

$$(6.14) \quad R(\mathbf{u}, \mathbf{v}; \mathbf{w}, \mathbf{z}) = R(\mathbf{w}, \mathbf{z}; \mathbf{u}, \mathbf{v}) = -R(\mathbf{v}, \mathbf{u}; \mathbf{w}, \mathbf{z}) = -R(\mathbf{u}, \mathbf{v}; \mathbf{z}, \mathbf{w}).$$

That is why, in dimension 2, it boils down to a single number K . In our settings, an easy proof follows §6.1.1, with §4.5 taken into account.

6.2. **Parallel transport.** It would be tempting to define a *parallel* (our new term for “constant”, *i.e.*, changing as little as possible) vector field $\mathbf{u}$ on X by requiring that

$$\frac{D\mathbf{u}}{\partial u} \equiv \frac{D\mathbf{u}}{\partial v} \equiv 0.$$

Unfortunately, Theorem 6.8 states that, on most interesting surfaces, such fields do not exist. Hence, we stick to parallel fields along curves.

A tangent vector field $\mathbf{u}(t)$ along a curve $\gamma: [a, b] \rightarrow X$ is called *parallel* if

$$\frac{D\mathbf{u}}{dt} = 0 \quad \text{for each } t \in [a, b].$$

In local coordinates, letting $\mathbf{u} := \alpha \mathbf{r}_u + \beta \mathbf{r}_v$ and $\gamma: t \mapsto (u(t), v(t))$ and using (6.1), we obtain a system of first order linear ordinary differential equations

$$(6.15) \quad \begin{aligned} \alpha' + \alpha u' \Gamma_{11}^1 + \alpha v' \Gamma_{12}^1 + \beta u' \Gamma_{12}^1 + \beta v' \Gamma_{22}^1 &= 0, \\ \beta' + \alpha u' \Gamma_{11}^2 + \alpha v' \Gamma_{12}^2 + \beta u' \Gamma_{12}^2 + \beta v' \Gamma_{22}^2 &= 0, \end{aligned}$$

and the general theory of linear ODE's tells us that such a field exists along the whole curve and is uniquely determined by the initial value $\mathbf{u}(a) = \mathbf{u}_p$. This field is called the *parallel transport* of the vector $\mathbf{u}_p$ at the point $p := \gamma(a)$ *along the curve* γ . Sometimes, it is the resulting linear map

$$P_\gamma: T_{\gamma(a)}X \rightarrow T_{\gamma(b)}X$$

that is called parallel transport. Simple examples show that P is *path dependent*. Though, P_γ does not depend on the parametrization of γ (the chain rule).

In particular, the parallel transport along a closed curve does *not* need to result in the original vector; typically, the *holonomy* $P_\gamma: T_pX \rightarrow T_pX$ is a non-trivial linear operator.

Remark 6.16. The regularity of γ is irrelevant for (6.15). Hence, we can define the parallel transport along any smooth curve, and then extend it to piecewise smooth curves. Then, we have

$$P_\gamma = P_{\gamma_2} \circ P_{\gamma_1}$$

whenever γ is broken into two pieces $\gamma_1: [a, c] \rightarrow X$ and $\gamma_2: [c, b] \rightarrow X$. Denoting by $-\gamma: [b, a] \rightarrow X$ the same curve with the opposite orientation, we have

$$P_{-\gamma} = P_{\gamma}^{-1}.$$

By (6.6), we have $\mathbf{u} \cdot \mathbf{v} = \text{const}$ for any two parallel vector fields $\mathbf{u}, \mathbf{v}$. Hence,

$$P_{\gamma}: T_{\gamma(a)}X \rightarrow T_{\gamma(b)}X$$

is an *isometry* of inner product spaces:

$$(6.17) \quad P_{\gamma}(\mathbf{u}) \cdot P_{\gamma}(\mathbf{v}) = \mathbf{u} \cdot \mathbf{v}.$$

If X is orientable, P_{γ} is orientation preserving and, considering that $SO(2)$ is as simple as S^1 , it suffices to know the image of a single non-zero vector.

More generally, by (6.6),

$$\frac{d}{dt}(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot \frac{D\mathbf{v}}{dt} \quad \text{if } \mathbf{u} \text{ is parallel, } \frac{D\mathbf{u}}{dt} = 0;$$

in particular, $\mathbf{u} \cdot \mathbf{v} = \text{const}$ if $\mathbf{v}$ is also parallel. Thus, if $\mathbf{u}$ is parallel, so is $\mathbf{v} := \mathbf{u}^{\perp}$, see (6.10), and

$$(6.18) \quad \frac{D}{dt}(\alpha(t)\mathbf{u}(t) + \beta(t)\mathbf{v}(t)) = \alpha'(t)\mathbf{u}(t) + \beta'(t)\mathbf{v}(t).$$

Example 6.19. The holonomy along the core of the [Möbius strip](#) is orientation reversing, the precise details depending on the model. Cf. [Example 7.17](#) below.

Example 6.20. We can easily solve (6.15) only in a few simple cases. Usually, we need to use tricks and appeal to our geometric intuition. For example, assume that two surfaces are tangent to each other at each point of a common curve γ , so that $T_{\gamma(t)}X = T_{\gamma(t)}Y$ for each t . Then, due to our extrinsic definition, $D_X/dt = D_Y/dt$ and, hence, the two parallel transports are equal. Since also our theory is invariant under (local) isometries, this observation is particularly useful if one of X, Y is locally isometric to the plane, where “parallel” means “constant.”

Problem 6.21. Pick a point on the sphere and compute the holonomy along (any) circle. *Hint:* use tangent cylinders/cones.

Example 6.22. Assume that γ is a normal section of X at each of its point. Then the unit tangent vector field $\mathbf{t}$ is parallel along γ . Indeed,

$$\frac{d\mathbf{t}}{ds} \text{ is normal to } X, \quad \text{hence,} \quad \frac{D\mathbf{t}}{ds} = 0.$$

For example, this is true for *big* circles on the sphere.

6.2.1. Curves on surfaces. For curves on a surface X we can develop a full analogue of §2.1. Assume X oriented and γ parametrized by the arc length. In addition to the unit tangent vector $\mathbf{t}$, we can define the *geodesic normal vector*

$$\mathbf{n}_g := \mathbf{t}^{\perp} = \mathbf{N} \times \mathbf{t},$$

see (6.10). Besides, we have the *geodesic curvature vector* and *geodesic curvature*:

$$\kappa_g \mathbf{n}_g := \frac{D\mathbf{t}}{ds} \in TX, \quad \kappa_g := \frac{D\mathbf{t}}{ds} \cdot \mathbf{n}_g \in \mathbb{R}.$$

These are the objects introduced at the beginning of §5.2.

Literally the same arguments as in §2.1 result in a geodesic version of the Frenet equations:

$$\frac{D\mathbf{t}}{ds} = \kappa_g \mathbf{n}_g, \quad \frac{D\mathbf{n}_g}{ds} = -\kappa_g \mathbf{t}.$$

We can also define the *angle function* and *rotation number*,

$$(6.23) \quad \theta(s) := \int_0^s \kappa_g ds, \quad \text{rot } \gamma := \frac{\Delta\theta}{2\pi} = \frac{1}{2\pi} \int_0^l \kappa_g ds,$$

except that the rotation number of a closed curve *no longer needs to be an integer!* If $\mathbf{u}$ and $\mathbf{v} := \mathbf{u}^\perp$ are parallel vector fields starting from $\mathbf{t}(0)$ and $\mathbf{n}_g(0)$, respectively, then

$$\mathbf{t}(s) = \mathbf{u}(s) \cos \theta(s) + \mathbf{v}(s) \sin \theta(s).$$

Hence, the linear isometry

$$\bar{P}_\gamma: T_{\gamma(0)}X \rightarrow T_{\gamma(l)}X, \quad \alpha\mathbf{t}(0) + \beta\mathbf{n}_g(0) \mapsto \alpha\mathbf{t}(l) + \beta\mathbf{n}_g(l),$$

differs from the parallel transport P_γ by $\Delta\theta$, regarded as an element of $SO(1)$. If γ is closed (of class C^1) and X is orientable, then $\bar{P}_\gamma = \text{id}_2$ and, hence,

$$(6.24) \quad P_\gamma = \text{rotation through } (-\Delta\theta = -2\pi \text{rot } \gamma).$$

6.3. Geodesics. On the plane, straight lines are characterized by the property that their unit tangent vector is constant. On a surface, we define geodesics as curves whose tangent vector changes “as little as possible.”

Definition 6.25. A curve $\gamma: \langle a, b \rangle \rightarrow X$ is called a (*parametrized*) *geodesic* if its velocity vector γ' is parallel:

$$\frac{D\gamma'}{dt} \equiv 0.$$

In other words, γ (with any parametrisation) is a geodesic if and only if $\kappa_g \equiv 0$.

In view of (6.17), geodesics are always parametrized at a constant speed, *i.e.*, by a *constant multiple* of the arc length. A boring example is the *constant geodesic* $\gamma = \text{const}$. More interesting are big circles on a sphere, see Example 6.22.

In coordinates, letting $\gamma: t \mapsto (u(t), v(t))$, equation (6.15) turns into a system of two *non-linear* second order ordinary differential equations

$$(6.26) \quad \begin{aligned} u'' + u'^2 \Gamma_{11}^1 + 2u'v' \Gamma_{12}^1 + v'^2 \Gamma_{22}^1 &= 0, \\ v'' + u'^2 \Gamma_{11}^2 + 2u'v' \Gamma_{12}^2 + v'^2 \Gamma_{22}^2 &= 0, \end{aligned}$$

and general theory (*aka* Picard–Lindelöf theorem) only asserts that, for any set of initial data, which, for a second order equation, are

$$\gamma(0) = p := (u_0, v_0) \in X, \quad \gamma'_0 = \mathbf{u} := (u'_0, v'_0) \in T_p X,$$

there is a unique solution *defined on a certain interval* $t \in (-\varepsilon, \varepsilon)$. In other words, just like in the case of lines on the plane, through each point in each direction there is a unique parametrized geodesic; however, unlike lines, *a priori* we cannot know for how long this geodesic will extend. This is to be discussed in §6.5 below.

Example 6.27. On the sphere S^2 , through each point in each direction there is a unique big circle. Hence, apart from constants, big circles (parametrized at constant speed) are all geodesics on S^2 , see Example 6.22.

Remark 6.28. As is well known, there are two reasons why a solution γ to an ordinary differential equation may not extend to the whole line $\mathbb{R}$:

- (1) γ leaves the domain of definition of the equation, or
- (2) γ blows up in finite time (e.g., $y' = -y^2$).

Geodesics are parametrized at constant speed, and they have no reason to blow up. However, leaving the domain is quite realistic: e.g., on a half-plane, some lines hit the border and do not extend, just because there is no surface beyond the border. In §6.4 below we will see that this is the only scenario possible for geodesics.

We need a more precise existence statement, known as *smooth dependence of solutions on the initial data*. For example, one can convert (6.26) to a first order system and refer to the *flow* of a vector field.

Theorem 6.29. *For each point $p \in X$, there is a neighborhood $U \ni p$ and a real number $\varepsilon > 0$ such that, for each pair $q \in U$ and $\mathbf{t} \in T_q X$, $|\mathbf{t}| = 1$, there exists a unique parametrized geodesic*

$$\gamma := \gamma(s; q, \mathbf{t}): (-\varepsilon, \varepsilon) \rightarrow X, \quad \gamma(0) = q, \quad \gamma'(0) = \mathbf{t}.$$

Furthermore, this geodesic γ depends smoothly on $(s, q, \mathbf{t})$. $\triangleleft$

Remark 6.30. Strictly speaking, we need a certain class of smoothness of X , at least C^3 , as Γ_{**}^* already use two derivatives. Though, see Remark 3.6. We did not discuss the tangent bundle TX as a smooth manifold, but it suffices to understand the last assertion of Theorem 6.29 coordinatewise: the geodesic γ is a pair (u, v) of smooth functions of quintuples $(s; u_0, v_0; u'_0, v'_0)$.

Remark 6.31. If $\gamma(s)$ is a solution to (6.26), so is a reparametrization $\gamma(r^{-1}s)$, $r = \text{const}$: it is defined on $(-r\varepsilon, r\varepsilon)$ and has velocity vector $r\mathbf{t}$ at 0. Hence, we can restate Theorem 6.29 in a more convenient form: there is a neighborhood $U \ni p$ and a number $\rho > 0$ such that the geodesic $\gamma(t; q, \mathbf{u})$ is defined and smooth on

$$[-1, 1] \times U \times \{\mathbf{u} \in T_q X \mid |\mathbf{u}| < \rho\}.$$

6.3.1. Geodesics as critical points. Contrary to common belief, geodesics are not always shortest curves. Sometimes they are almost longest, e.g., a big circle of central angle $> \pi$ has arbitrary close shorter curves connecting its endpoints.

The truth is that, similar to minimal surfaces (see §5.5), geodesics are the critical points of the arc length functional. To prove this, consider a *variation*, i.e., a 1-parameter family of curves

$$\gamma_t(s) := \gamma(s, t), \quad s \in [0, l], \quad t \in (-\varepsilon, \varepsilon).$$

We assume that

$$|\gamma'_0| \equiv 1, \quad \text{where} \quad \gamma' := \frac{\partial \gamma}{\partial s},$$

i.e., γ_0 is parametrized by the arc length (but not necessarily γ_t with $t \neq 0$) and that the *variation vector field*

$$h_0(s) := \frac{\partial \gamma}{\partial t} \Big|_{t=0} \quad \text{vanishes at the ends:} \quad h_0(0) = h_0(l) = 0.$$

Indeed, the variation vector field is the direction/speed in/at which the points of γ_0 start moving, and we want the two endpoints fixed: we consider curves *connecting a given pair of points*. Otherwise, nothing would prevent us from, e.g., shrinking the curve along itself. The precise condition that we actually need is the vanishing of the red term below; e.g., normal variations would also do. Somewhat stronger assumptions are to be imposed in §8.2 below.

We proceed as in §5.5. Consider the arc length function

$$\ell(t) := \text{length}(\gamma_t) = \int_0^t \left| \frac{\partial \gamma}{\partial s} \right| ds.$$

Then,

$$\ell'(0) = \int_0^t \frac{\partial}{\partial t} \left| \frac{\partial \gamma}{\partial s} \right| \Big|_{t=0} ds \stackrel{|\gamma'_0| \equiv 1}{=} \int_0^t \left(\gamma' \cdot \frac{D}{\partial t} \frac{\partial \gamma}{\partial s} \right) \Big|_{t=0} ds$$

By (6.6),

$$\left(\gamma' \cdot \frac{D}{\partial t} \frac{\partial \gamma}{\partial s} \right) \Big|_{t=0} = \gamma'_0 \cdot \frac{D h_0}{\partial s} = \frac{d}{ds} (\gamma'_0 \cdot h_0) - \left(\frac{D \gamma'_0}{ds} \cdot h_0 \right).$$

Hence,

$$\ell'(0) = \underbrace{(\gamma'_0 \cdot h_0) \Big|_{t=0}}_0 - \int_0^t \left(\frac{D \gamma'_0}{ds} \cdot h_0 \right) ds.$$

This vanishes for any h_0 if and only if $D \gamma'_0 / ds \equiv 0$, i.e., γ_0 is a geodesic.

Remark 6.32. One may ask how this computation would change should we use ordinary derivatives instead of covariant. The difference is at the very last step: we require that the integral should vanish for any *tangent* vector field h_0 ; hence, only the tangential component of γ'' matters in the dot product.

6.4. The exponential map. The term *exponential map* is due to the fact that this map is, indeed, $\exp$ in most classical cases. For example, $\exp: T_1 S^1 = \mathbb{R} \rightarrow S^1$ is the map $t \mapsto \exp(it)$, and it is this example that one should keep in mind when visualizing the exponential map. Or, the tangent spaces at 1 to most classical matrix groups, such as GL , SL , SO , U , SU , are appropriate subspaces of the respective matrix spaces, and the map is $m \mapsto \exp(m)$, sometimes with a coefficient.

Officially, the *exponential map* is the map

$$(6.33) \quad \exp_p: B_\rho(0) \subset T_p X \rightarrow X, \quad \mathbf{u} \mapsto \gamma(1, p, \mathbf{u})$$

defined on a small ball $B_\rho(0) \subset T_p X$ and given by Theorem 6.29 and Remark 6.31: we start a geodesic from p at speed $\mathbf{u}$ and travel for time 1.

The differential $d \exp_p$ at $0 \in T_p X$ is a linear map

$$T_0(T_p X) = T_p X \rightarrow T_p X,$$

and it is not a surprise that this differential is the identity; this assertion is almost a tautology using definition (3.14) via curves. By the inverse function Theorem 3.5, the restriction of $\exp_p$ to a smaller ball $B_\rho(0)$ is a diffeomorphism to a neighborhood of p in X . (We will need to shrink ρ and U given by Remark 6.31 several times; upon announcing each shrinkage, we will keep the old notation.) Referring to the *implicit function theorem* (essentially equivalent to Theorem 3.5), we see that

$$(6.34) \quad \begin{array}{l} \text{such a (smaller) radius } \rho \text{ works for all points } q \\ \text{in a (smaller) neighborhood } U \ni p. \end{array}$$

The image $\exp_p(B_\rho)$ of such a small ball is called a *normal neighborhood* of p or, sometimes, a (small) *geodesic ball* about p .

6.4.1. *Normal polar coordinates.* Undoubtedly, the “best” coordinates in the plane $T_p X$ are Cartesian; unfortunately, they are distorted too much by $\exp_p$. A better choice is the polar coordinates (r, θ) ; *via* $\exp_p$, we regard them as coordinates (r, θ) , henceforth called *normal polar*, on a normal neighborhood of p . Note that normal polar charts cover everything *except p itself*; however, due to (6.34), this point lies in normal polar charts about other, close points.

Lemma 6.35 (Gauss). *In normal polar coordinates (r, θ) about $p \in X$, one has*

$$E \equiv 1, \quad F \equiv 0.$$

Furthermore, one has

$$\sqrt{G}(p) := \lim_{r \rightarrow 0} \sqrt{G(r, \theta)} = 0, \quad (\sqrt{G})_r(p) := \lim_{r \rightarrow 0} \frac{1}{r} \sqrt{G(r, \theta)} = 1.$$

Proof. For $E \equiv 1$ observe that $\mathbf{r}_r$ are unit tangent vectors to the radial geodesics.

Denote by $\bar{\mathbf{r}}(r, \theta)$ the parametrization of $T_p X$, so that $\mathbf{r} := \exp_p \circ \bar{\mathbf{r}}$ and

$$\mathbf{r}_r = d\exp_p(\bar{\mathbf{r}}_r), \quad \mathbf{r}_\theta = d\exp_p(\bar{\mathbf{r}}_\theta).$$

The coefficients E, F, G do not make sense at p , but vectors and their dot products do. We can compute the limits

$$\lim_{r \rightarrow 0} \bar{\mathbf{r}}_\theta(r, \theta) = 0, \quad \lim_{r \rightarrow 0} \frac{1}{r} \bar{\mathbf{r}}_\theta(r, \theta) \text{ is a unit vector at } 0.$$

Since $d\exp_p$ at 0 is an isometry, this implies the last two statements and the fact that $\lim_{r \rightarrow 0} F(r, \theta) = 0$. To complete the proof, we show that F is constant along each radial ray: by (5.14),

$$F_r = \frac{1}{2}E_\theta + (\mathbf{r}_{rr} \cdot \mathbf{r}_\theta) = \frac{D}{dr} \mathbf{r}_r(r, \theta_0) \cdot \mathbf{r}_\theta = 0$$

since $E = \text{const}$ and $\mathbf{r}(r, \theta_0)$ is a parametrized geodesic, $\mathbf{r}_{rr} \perp TX$. $\square$

6.4.2. *Local minimal geodesics.* We are ready to prove that, *for each point q in a normal neighborhood $V_\rho \ni p$, there is a unique shortest curve connecting p and q , and this curve is the radial geodesic.*

Let $q = \mathbf{r}(r_0, \theta_0)$, $r_0 < \rho$. Then $\text{dist}(p, q) \leq r_0$ as p and q are connected by the radial geodesic of length r_0 . Consider another curve γ connecting p and q .

If γ lies in V_ρ , $\gamma: t \mapsto (r(t), \theta(t))$, then

$$(6.36) \quad \text{length}(\gamma) = \int_0^l \sqrt{r'^2 + G\theta'^2} dt \geq \int_0^l |r'| dt \geq |r(l) - r(0)| = r_0 - 0 = r_0,$$

the equality holding if and only if $\theta = \text{const}$ and $r(t)$ is monotonous, *i.e.*, γ is the radial geodesic, possibly re-parametrized.

If γ leaves V_ρ , by the previous argument its length is at least $\rho > r_0$ as, in this case, γ must cross the boundary circle at distance ρ from p . $\square$

Corollary 6.37 (convex neighborhood). *Each point $p \in X$ has a neighborhood $W \ni p$ which is convex in the sense that any two points $q', q'' \in W$ are connected by a unique shortest curve; this shortest curve is a geodesic.*

Proof. We use (6.34) and reduce ρ so that $V_\rho(p) \subset U$. Then, take $W := V_{\rho/2}(p)$. With this choice, we have

$$\text{dist}(q', q'') \leq \text{dist}(q', p) + \text{dist}(p, q'') < \rho$$

for any pair $q', q'' \in W$. Hence, q'' lies in the normal neighborhood $V_\rho(q')$ and, in this neighborhood, is connected to q' by a unique minimal geodesic. $\square$

Remark 6.38. With a bit more work, one can construct W truly convex, so that it contained the unique minimal geodesics connecting pairs of points $q', q'' \in W$.

In general, neither existence nor uniqueness of minimal geodesics hold globally. For the non-existence, consider a punctured plane and two points to the opposite sides of the puncture: they are not connected by a geodesic (a line) at all. We will deal with this issue in the next section. A classical example of the non-uniqueness is the two poles in a sphere. This cannot be repaired.

Corollary 6.39. *Any two surfaces X, Y of equal constant Gaussian curvature K are locally isometric. More precisely, given a pair of flags*

$$p \in X, \mathbf{u} \in T_p X \quad \text{and} \quad q \in Y, \mathbf{v} \in T_q Y, \quad |\mathbf{u}| = |\mathbf{v}| = 1,$$

there is a unique orientation preserving isometry f of a neighborhood $U \ni p$ to a neighborhood $V \ni q$ such that $f(p) = q$ and $df_p(\mathbf{u}) = \mathbf{v}$.

Proof. We take for U, V normal neighbourhoods of p, q , respectively of the same radius ρ and let

$$f := \exp_q \circ \bar{f} \circ \exp_p^{-1},$$

where $\bar{f} = df_p$ is the (only) orientation preserving linear isometry $T_p X \rightarrow T_q Y$ that takes $\mathbf{u}$ to $\mathbf{v}$. If $K = \text{const}$, equation (5.22) with the initial conditions given by Lemma 6.35 is easily solved:

$$\sqrt{G}(r, \theta) = \begin{cases} (\sqrt{K})^{-1} \sin(r\sqrt{K}), & \text{if } K > 0, \\ (\sqrt{-K})^{-1} \sinh(r\sqrt{-K}), & \text{if } K < 0, \\ r, & \text{if } K = 0, \end{cases}$$

and we conclude that f is indeed an isometry. It is unique since any isometry should take radial geodesics from p to those from q . $\square$

6.5. Complete surfaces. A regular surface X is called *complete* if it is complete (any Cauchy sequence converges) *with respect to its intrinsic distance*.

Example 6.40. A surface that is complete (equivalently, closed) as a subset of $\mathbb{R}^3$ is also complete in this new sense but. The converse, in general, is not true. For example, one can take the cylinder over the spiral

$$\rho = 1 + e^\theta$$

(polar coordinates) of infinite length; it is isometric to $\mathbb{R}^2$ but not closed in $\mathbb{R}^3$.

Occasionally, X is said to be *geodesically complete* if each parametrized geodesic $\gamma: (a, b) \rightarrow X$ extends to the whole line $\mathbb{R}^1$. This term is not widely used because we assert that the two properties are equivalent.

Geodesic completeness $\Rightarrow$ completeness. By the assumption, at each point $p \in X$ the exponential map $\exp_p$ is defined on the whole plane $T_p X$. Each Cauchy sequence x_n is bounded, $\text{dist}(x_1, x_n) \leq d$ for all $n \geq 1$. Hence, it lies in the *closed* geodesic ball $\exp_{x_1}(\bar{B}_d)$, which is compact. Compact spaces are complete.

Completeness $\Rightarrow$ geodesic completeness. Consider a geodesic

$$\gamma: [0, l) \rightarrow X$$

parametrized by the arc length and assume that it does *not* extend beyond l . The sequence $\gamma(l - 1/n)$ is Cauchy, hence it converges to a point $p \in X$. Then, clearly, also $\lim_{s \rightarrow l} \gamma(s) = p$. Thus, a certain restriction $\gamma: [a, l) \rightarrow X$ lies in a convex neighbourhood $W \ni p$. According to the construction that we used in the proof of [Corollary 6.37](#), such a geodesic cannot be “weird”: it is a radial geodesic from any of its points and, eventually, leaves W . This is a contradiction.

Theorem 6.41 (Hopf–Rinow). *On a complete surface X , any two points $p, q \in X$ are connected by a minimal geodesic.*

Note that we no longer state the uniqueness of “long” shortest geodesics: the two poles of a sphere would be a counterexample.

A comprehensible proof of [Theorem 6.41](#) needs pictures, and we refer to the textbook or the lectures. The brilliant idea is “shooting” a radial geodesic γ from p in the hope to hit q . To “aim” γ , we consider a closed normal neighborhood $\bar{V}_\rho(p)$. Since the boundary $\partial \bar{V}_\rho(p)$ is compact, the distance function

$$d(x) := \text{dist}(q, x)$$

take its absolute minimum at some point $r \in \partial \bar{V}_\rho(p)$, and we take for γ the radial geodesic from p to r , parametrized by the arc length.

To trace the progress, we consider the function

$$\delta(s) := \text{dist}(q, \gamma(s)).$$

We have

$$\delta(s) = \text{dist}(p, q) - s \quad \text{for } 0 \leq s \leq \rho.$$

If δ stays like this, we do hit q at time $s = \text{dist}(p, q)$, completing the proof. If not, consider the last time s_0 beyond which δ starts deviating and “re-aim” γ at the point $r := \gamma(s_0)$. If the new outgoing geodesic is at an angle other than π from the incoming γ , we can “cut the corner” (within a convex neighborhood of r) and find a curve from p to q of length less than $\text{dist}(p, q)$. If the angle is π , the new geodesic merely continues γ , contradicting our choice of s_0 . $\square$

7. EXAMPLES

We consider a few well-known classes of surfaces traditionally used as a source of examples. To each class, we apply the machinery developed in the previous sections and try to say as much as reasonably possible about the geometry of the surface. This section can be consulted frequently, whenever a new notion is introduced: I have chosen to work out each example in full rather than break it into several partial computations scattered throughout the text.

7.1. Graphs. Let X be the graph of a function φ :

$$\mathbf{r}(x, y) = (x, y, \varphi(x, y)), \quad (x, y) \in U.$$

Then,

$$\begin{aligned} \mathbf{r}_x &= (1, 0, \varphi_x), & \mathbf{r}_y &= (0, 1, \varphi_y), & \mathbf{N} &= \frac{(-\varphi_x, -\varphi_y, 1)}{\sigma}, & \sigma &:= \sqrt{1 + \varphi_x^2 + \varphi_y^2}. \\ \mathbf{r}_{xx} &= (0, 0, \varphi_{xx}), & \mathbf{r}_{xy} &= (0, 0, \varphi_{xy}), & \mathbf{r}_{yy} &= (0, 0, \varphi_{yy}). \end{aligned}$$

Therefore,

$$\begin{aligned} E &= 1 + \varphi_x^2, & F &= \varphi_x \varphi_y, & G &= 1 + \varphi_y^2, & \text{Area}(X) &= \iint_U \sigma \, dx \, dy. \\ e &= \frac{\varphi_{xx}}{\sigma}, & f &= \frac{\varphi_{xy}}{\sigma}, & g &= \frac{\varphi_{yy}}{\sigma}. \end{aligned}$$

Then,

$$H = \frac{(1 + \varphi_x^2)\varphi_{yy} - 2\varphi_x\varphi_y\varphi_{xy} + (1 + \varphi_y^2)\varphi_{xx}}{2\sigma^3}, \quad K = \frac{\varphi_{xx}\varphi_{yy} - \varphi_{xy}^2}{\sigma^3}.$$

If (x, y) is a critical point, $\varphi_x = \varphi_y = 0$, this simplifies:

$$\mathbf{I} = \text{id}_2, \quad \mathbf{W} = \mathbf{II} = \mathcal{H} \quad (\text{the Hessian matrix, see §5.6}).$$

We find that

- $K = 0$ if and only if (x, y) is degenerate;
- $K > 0$ if (x, y) is a minimum or a maximum;
- $K < 0$ if (x, y) is a saddle point.

This behaviour is similar to what has been discussed in §5.6: in a sense, the *height function* φ is the “distance” from $(0, 0, -\infty)$.

In general, there is not much that we can say about the geodesics on a graph.

7.2. Surfaces of revolution. Consider a curve

$$\alpha: s \mapsto (x(s), r(s)), \quad r > 0, \quad s \in [0, l],$$

in the (x, y) -plane, parametrized by the arc length ($x'^2 + r'^2 = 1$), and let X be the surface obtained by revolving (the trace of) α about the x -axis:

$$(7.1) \quad \mathbf{r}(s, \theta) = (x(s), r(s) \cos \theta, r(s) \sin \theta).$$

(To avoid confusion, we have renamed the y -coordinate of the generating curve to r : it represents the radius of the latitudinal circle rather than the y -coordinate of a point in X .) Then, letting $' := d/ds$,

$$\begin{aligned} \mathbf{r}_s &= (x', r' \cos \theta, r' \sin \theta), & \mathbf{r}_\theta &= (0, -r \sin \theta, r \cos \theta), \\ \mathbf{N} &= (r', -x' \cos \theta, -x' \sin \theta), \\ \mathbf{r}_{ss} &= (x'', r'' \cos \theta, r'' \sin \theta), & \mathbf{r}_{s\theta} &= (0, -r' \sin \theta, r' \cos \theta), \\ \mathbf{r}_{\theta\theta} &= (0, -r \cos \theta, -r \sin \theta). \end{aligned}$$

Denoting by R_θ the rotation through θ about the x -axis (which preserves both dot and cross products), it may be convenient to present this in more invariant terms, using the standard objects affiliated with the curve:

$$\begin{aligned}\mathbf{r}_s &= R_\theta(\mathbf{t}), & \mathbf{r}_\theta &= R_\theta(r\mathbf{k}), & \mathbf{N} &= -R_\theta(\mathbf{n}_s), \\ \mathbf{r}_{ss} &= R_\theta(\kappa_s \mathbf{n}_s), & \mathbf{r}_{s\theta} &= R_\theta(r'\mathbf{k}), & \mathbf{r}_{\theta\theta} &= -R_\theta(r\mathbf{j}).\end{aligned}$$

In any case, denoting by φ the angle between $\mathbf{t}$ and the x -axis.

$$\begin{aligned}E &= 1, & F &= 0, & G &= r^2, \\ e &= x''r' - x'r'' = -\kappa_s, & f &= 0, & g &= x'r = r \cos \varphi,\end{aligned}$$

whereupon

$$k_1 = \frac{e}{E} = -\kappa_s, \quad k_2 = \frac{g}{G} = \frac{x'}{r} = \frac{\cos \varphi}{r}, \quad K = -\frac{r''}{r};$$

for K , we chose to use (5.22) directly. Thus, k_1 is the curvature of the *meridian* $\theta = \text{const}$, whereas k_2 can be interpreted as the curvature r^{-1} of the *parallel* (circle of latitude) $s = \text{const}$ through the point scaled by $\cos \varphi$.

Consequently, X is *developable* (i.e., $K \equiv 0$, locally isometric to the plane, see Corollary 6.39) if and only if r is a linear function of s . Considering the arc length parametrization, $r = s \cos \varphi_0$, $x = s \sin \varphi_0$ for some $\varphi_0 = \text{const}$, and X is either

- a “round” cylinder, if $\varphi_0 = 0$, or
- a straight circular cone, if $0 < \varphi_0 < \pi/2$, or
- a “vertical” punctured plane, if $\varphi_0 = \pi/2$.

One can easily classify other surfaces of revolution of constant Gaussian curvature, but this is left to homeworks and/or exams.

As in calculus 101,

$$\text{Area}(X) = 2\pi \int_0^l r \, ds = \text{length}(\alpha) \cdot 2\pi R_c,$$

recovering the *Pappus–Guldinus theorem*: the second factor $2\pi R_c$ is the length of the circle described by the *centroid* (*barycenter*, *center of mass*) of the curve α .

We observe that (7.1) is an analogue of normal polar coordinates (4.26), with an extra benefit that the coordinate lines are lines of curvature (cf. Theorem 5.29). Using (5.17),

$$\Gamma_{12}^2 = \frac{r'}{r}, \quad \Gamma_{22}^1 = -rr', \quad \Gamma_{**}^* = 0 \quad \text{otherwise},$$

and the geodesic equation (6.26) takes the form

$$\ddot{s} - rr'\dot{\theta}^2 = 0, \quad r\ddot{\theta} + r'\dot{s}\dot{\theta} = 0,$$

the dots standing for the derivatives with respect to the parameter t of the geodesic $\gamma: t \mapsto (s(t), \theta(t))$ in question. We find that

- each meridian $t \mapsto (t, \theta_0)$ is a geodesic, whereas
- a parallel $t \mapsto (s_0, t)$ is geodesic if and only if it is *critical*, $r'(s_0) = 0$, i.e., the generating curve is parallel to the x -axis,

cf. Example 6.22. In general, the geodesic curvature of a parallel is

$$(7.2) \quad \kappa_g = r^{-1} \sin \varphi, \quad \text{where } \varphi \text{ is the angle from the } x\text{-axis to } \mathbf{t}, \text{ as above.}$$

Assuming t the arc length, we can write

$$\dot{s} = \cos \psi, \quad \dot{\theta} = \frac{\sin \psi}{r},$$

where $\psi(t)$ is the angle between γ and the meridian. Since $r' = \dot{r}/\dot{s} = \dot{r}/\cos\psi$, the first equation

$$-r\dot{\psi}\sin\psi\cos\psi - \dot{r}\sin^2\psi = -\sin\psi\frac{d}{dt}(r\sin\psi) = 0$$

has a first integral

$$(7.3) \quad r\sin\psi = \text{const.}$$

Formally, the meridians $\psi = 0$ also satisfy this equation, and so do *all* (not only critical) parallels $\psi = \pi/2$, $r = \text{const.}$ In the latter case, the parasitic phenomenon is due to the multiplication by $\cos\psi = 0$ in the course of solution.

We can also compute the holonomy along a parallel $\beta := \{s = \text{const}\}$: by (6.24), (6.23), and (7.2), it is

$$P_\beta = -2\pi \text{rot } \beta = -\int_0^{2\pi r} \kappa_g dt = -2\pi \sin\varphi \in SO(1).$$

(Pay attention to the arc length parametrization $t := r\theta \in [0, 2\pi r]$.) This angle is 0 only if β is a geodesic, $\varphi = 0$ or π . (Though, if $\varphi = \pi/2$, so that the generating curve is vertical, it is again $-2\pi = 0 \pmod{2\pi}$.) We could have achieved the same result by using the trick with tangent cones described in Example 6.20.

7.2.1. Spheres. As an exercise, represent a sphere as a surface of revolution and use high school geometry to check (7.3) for a big circle (see Example 6.27) that is neither a meridian nor equator.

7.2.2. Cylinders. On a cylinder of revolution ($r = \text{const}$), by (7.3), all geodesics are given by $\psi = \text{const}$: they are the meridians (generatrices, $\psi = 0$), latitudinal circles ($\psi = \pi/2$), or helices. Alternatively, we can see this by developing the cylinder: geodesics would map to horizontal, vertical, or slanted lines, respectively.

7.2.3. Cones. Consider a cone of revolution ($r = s \cos\varphi_0$), develop it to the plane, and use high school geometry to verify (7.3).

7.3. Tubes. Consider a curve $\alpha: [0, l] \rightarrow \mathbb{R}^3$, parametrized by the arc length, and assume that κ never vanishes, so that, at each point, we have well-defined Frenet frame $\mathbf{t}, \mathbf{n}, \mathbf{b}$. The map

$$(s, u, v) \mapsto \alpha(s) + u\mathbf{n} + v\mathbf{b}$$

is a local diffeomorphism at each point of the s -axis. Hence, due to the compactness of α , it is a diffeomorphism of a sufficiently small neighborhood $|u|, |v| < \varepsilon \ll 1$ onto a neighborhood of the trace of α . Restricting this map to a cylinder of radius $r < \varepsilon$ about the s -axis, we obtain the *tube* $X := X_r$ about α , parametrized *via*

$$\mathbf{r}(s, \theta) := \alpha(s) + r\mathbf{n}(s)\cos\theta + r\mathbf{b}(s)\sin\theta.$$

In view of Frenet equations (2.8), we have

$$\begin{aligned} \mathbf{r}_s &= \mathbf{t}(1 - r\kappa\cos\theta) - r\tau\mathbf{n}\sin\theta + r\tau\mathbf{b}\cos\theta, & \mathbf{r}_\theta &= -r\mathbf{n}\sin\theta + r\mathbf{b}\cos\theta, \\ \mathbf{N} &= -(\mathbf{n}\cos\theta + \mathbf{b}\sin\theta), & \sigma &:= |\mathbf{r}_s \times \mathbf{r}_\theta| = r(1 - r\kappa\cos\theta), \\ E &= (1 - r\kappa\cos\theta)^2 + r^2\tau^2, & F &= r^2\tau, \quad G = r^2. \end{aligned}$$

For the second fundamental form, we compute the derivatives of $\mathbf{N} = (\alpha - \mathbf{r})/r$:

$$\begin{aligned} \mathbf{N}_s &= \kappa\mathbf{t}\cos\theta + \tau\mathbf{n}\sin\theta - \tau\mathbf{b}\cos\theta, & \mathbf{N}_\theta &= \mathbf{n}\sin\theta - \mathbf{b}\cos\theta, \\ e &= -\kappa\cos\theta(1 - \kappa\cos\theta) + r\tau^2, & f &= r\tau, \quad g = r. \end{aligned}$$

Therefore,

$$(7.4) \quad H = \frac{1 - 2r\kappa \cos \theta}{2r(1 - r\kappa \cos \theta)}, \quad K = \frac{-\kappa \cos \theta}{r(1 - r\kappa \cos \theta)}.$$

An immediate remarkable consequence is that

$$\text{Area}(X) = \iint_X d\text{vol} = \int_0^l \int_0^{2\pi} r(1 - r\kappa \cos \theta) d\theta ds = \text{length}(\alpha) \cdot 2\pi r,$$

as if X were a straight cylinder. Likewise, the total Gaussian curvature is

$$\iint_X K d\text{vol} = - \int_0^l \int_0^{2\pi} \kappa \cos \theta d\theta ds = 0.$$

Though, that would also follow from the Gauss–Bonnet [Theorem 8.1](#) below.

Analyzing [\(7.4\)](#), we see that X can be divided into two “equal halves”

$$\begin{aligned} X_- &:= \{-\pi/2 \leq \theta \leq \pi/2\} = \{p \in X \mid K(p) \leq 0\}, \\ X_+ &:= \{\pi/2 \leq \theta \leq 3\pi/2\} = \{p \in X \mid K(p) \geq 0\}, \end{aligned}$$

and a similar computation shows that

$$(7.5) \quad \iint_{X_+} K d\text{vol} = - \iint_{X_-} K d\text{vol} = 2 \int_0^l \kappa ds$$

is related to the total curvature of α .

We do not discuss the Christoffel symbols or geodesic equation. Just notice that, unlike the case of surfaces of revolution (see [§7.2](#)), most meridians $\theta = \text{const}$ are not geodesics, whereas the parallels $s = \text{const}$ are, see [Example 6.22](#).

7.3.1. The Fary–Milnor Theorem. Fenchel’s theorem [§2.4.4](#) can be obtained as an immediate consequence of [\(7.5\)](#): it suffices to show that, assuming α closed, the Gauss map $\mathbf{N}: X_+ \rightarrow S^2$ is surjective, cf. [\(5.9\)](#) and [Theorem 5.10](#).

Remark 7.6. This time, X_+ is not closed and we cannot speak about the degree of $\mathbf{N}$ but, since $\mathbf{N}$ is orientation preserving, $K \geq 0$, each point makes a positive contribution to the area of S^2 . That is why we only get an inequality.

To this end, pick a generic point $\mathbf{a} \in S^2$, regarded as a unit vector, and consider the *height function*

$$h_{\mathbf{a}}: s \mapsto \alpha(s) \cdot \mathbf{a}.$$

By the compactness of $[0, l]$, this function attains its maximum at some point s_0 . Since

$$h' = \mathbf{t} \cdot \mathbf{a}, \quad h'' = \mathbf{t}' \cdot \mathbf{a} = \kappa \mathbf{n} \cdot \mathbf{a},$$

at a point s_0 of (local) maximum of h we must have

$$\mathbf{t}(s_0) \cdot \mathbf{a} = 0, \quad \mathbf{n}(s_0) \cdot \mathbf{a} \leq 0, \quad \text{hence} \quad \alpha(s_0) + r\mathbf{a} \in X_+;$$

this point is mapped to $\mathbf{a}$ by $\mathbf{N}$. (In calculus, another option would be that s_0 is one of the endpoints, not critical for h . However, since the curve is *closed*, we can “break” the parametrization at any other point.)

For a generalization of this theorem, let us say that a topological circle (the trace of a simple closed curve) in $\mathbb{R}^3$ is *unknotted* if it bounds an embedded topological disk; otherwise, the circle is *knotted*.

Theorem 7.7 (Fary–Milnor; cf. §2.4.4). *If a simple closed curve $\alpha: [0, l] \rightarrow \mathbb{R}^3$ is knotted, then*

$$\int_0^l \kappa ds \geq 4\pi.$$

Proof. By (7.5), it suffices to prove that the restriction $\mathbf{N}: X_+ \rightarrow S^2$ is at least two-to-one; then, the total Gaussian curvature of X_+ is at least $2 \text{Area}(S^2) = 8\pi$. By the previous discussion, that would follow from the fact that a generic height function $h_{\mathbf{a}}$ has at least two local maxima. If it had one (hence, also one local minimum), the trace of α would consist of a single “ascending” branch and a single “descending” one. Then, connecting points of equal height by segments, we would obtain an embedded disk bounded by the trace, so it would be unknotted. $\square$

Remark 7.8. More formally, extracting information about a smooth manifold from the critical points of a single generic function is the subject of [Morse theory](#).

7.4. Ruled surfaces. Fix a curve $\alpha: (a, b) \rightarrow \mathbb{R}^3$ (the *directrix*) and a unit vector field $\mathbf{w}$ on γ . The *ruled surface* (or *scroll*) defined by γ and $\mathbf{w}$ is the surface X parametrized *via*

$$\mathbf{r}(s, t) := \alpha(s) + t\mathbf{w}(s);$$

it is swept by the lines (*generatrices* or *generators*) $\ell(s)$ passing through the points $\alpha(s)$ and directed by the vectors $\mathbf{w}(s)$. Often, such surfaces have self-intersections and/or singular (irregular) points: we ignore the former and confine our analysis to the complement of the latter.

Assuming the arc length parametrization of γ and letting $' := d/ds$, we have

$$\begin{aligned} \mathbf{r}_s &= \mathbf{t} + t\mathbf{w}', & \mathbf{r}_t &= \mathbf{w}, & \mathbf{r}_{ss} &= \kappa\mathbf{n} + t\mathbf{w}'', & \mathbf{r}_{st} &= \mathbf{w}', & \mathbf{r}_{tt} &= 0 \\ (7.9) \quad E &= 1 + 2t(\mathbf{t} \cdot \mathbf{w}') + t^2|\mathbf{w}'|^2, & F &= \mathbf{t} \cdot \mathbf{w}, & G &= 1, \\ f &= \frac{\mathbf{w}' \cdot (\mathbf{t} \times \mathbf{w})}{|\mathbf{r}_s \times \mathbf{r}_t|}, & g &= 0; & K &= -\frac{(\mathbf{w}' \cdot (\mathbf{t} \times \mathbf{w}))^2}{|\mathbf{r}_s \times \mathbf{r}_t|^4} \leq 0. \end{aligned}$$

7.4.1. Developable ruled surfaces. A surface X is called *developable* if $K \equiv 0$; by [Corollary 6.39](#), such surfaces are locally isometric to the plane. For ruled surfaces, this condition translates into

$$\mathbf{w}' \cdot (\mathbf{t} \times \mathbf{w}) \equiv 0,$$

i.e., vectors $\mathbf{t}$, $\mathbf{w}$, $\mathbf{w}'$ are linearly dependent at each point of α .

Example 7.10 (cylinder). The simplest example of a developable ruled surface is a (generalized) cylinder, where

$$\mathbf{w} = \text{const}, \quad \text{hence} \quad \mathbf{w}' = 0;$$

all generatrices are parallel to each other.

Example 7.11 (cone). To make all generatrices pass through a single point, *e.g.*, the origin O , we take

$$\mathbf{w} = \frac{\alpha}{|\alpha|}, \quad \text{hence} \quad \mathbf{w}' = \frac{|\alpha|\mathbf{t} + |\alpha|'\alpha}{|\alpha|^2} \in \text{span}(\mathbf{t}, \mathbf{w});$$

thus, the resulting (generalized) cone is also developable. It is singular at O .

Example 7.12 (*tangent developable*). A more interesting example is the surface swept by all tangents to α ; we take $\mathbf{w} = \mathbf{t}$ and find that this surface is developable. Typically, it is “very” singular (cuspidal) at each point of α , see next section.

7.4.2. The line of striction. From now on, we assume that X is *non-cylindrical*, so that $\mathbf{w}' \neq 0$. (To simplify the analysis, we assume that $\mathbf{w}'$ has no zeroes.) Then, for each fixed s , the dependence of K on t is

$$K = -\frac{\text{const}}{|(\mathbf{t} \times \mathbf{w}) + t(\mathbf{w}' \times \mathbf{w})|^4}.$$

This function takes its minimum at

$$t_m := t_m(s) := -\frac{(\mathbf{t} \times \mathbf{w}) \cdot (\mathbf{w}' \times \mathbf{w})}{|\mathbf{w}' \times \mathbf{w}|^2} = -\frac{1}{|\mathbf{w}'|^2} \begin{vmatrix} \mathbf{t} \cdot \mathbf{w}' & \mathbf{t} \cdot \mathbf{w} \\ \mathbf{w} \cdot \mathbf{w}' & 1 \end{vmatrix} = -\frac{\mathbf{t} \cdot \mathbf{w}'}{|\mathbf{w}'|^2};$$

we used (1.8) and the fact that $|\mathbf{w}| = 1$, $\mathbf{w} \cdot \mathbf{w}' = 0$. The curve

$$(7.13) \quad \tilde{\alpha}: s \mapsto \alpha(s) + t_m(s)\mathbf{w}(s)$$

is called the *line of striction* of X . Since

$$(7.14) \quad \tilde{\alpha}' = \mathbf{t} + t_m \mathbf{w}' + t_m' \mathbf{w},$$

under the additional assumption that $\mathbf{t}, \mathbf{w}, \mathbf{w}'$ are linearly independent, *i.e.*, $K \neq 0$, this is a regular curve. Hence, we can re-parametrize the surface and assume that

$$\mathbf{t} \cdot \mathbf{w}' \equiv 0, \quad \text{i.e., } \alpha \text{ is the line of striction,}$$

parametrized by the arc length. The expression for K simplifies to

$$K = -\frac{\lambda^2}{(\lambda^2 + t^2)^2}, \quad \text{where } \lambda := \pm \frac{|\mathbf{t} \times \mathbf{w}|}{|\mathbf{w}'|}$$

is the *distribution parameter*. The latter can be given a sign (usually unnecessary) by observing that, since $\mathbf{w}'$ is orthogonal to both $\mathbf{t}$ and $\mathbf{w}$, we have a vector identity

$$\mathbf{t} \times \mathbf{w} = \lambda \mathbf{w}', \quad \text{or } \lambda := \frac{\mathbf{w}' \cdot (\mathbf{t} \times \mathbf{w})}{|\mathbf{w}'|^2}.$$

Next, from (7.9) we have

$$|\mathbf{r}_s \times \mathbf{r}_t|^2 = |\lambda \mathbf{w}' + t(\mathbf{w}' \times \mathbf{w})|^2 = |\mathbf{w}'|^2(\lambda^2 + t^2);$$

hence, X is singular (or, rather, the chosen parametrization is not regular) only on the line of striction and only at the points where $\lambda = 0$ or, equivalently, $K = 0$. For this reason, if X is developable, $K \equiv 0$, the line of striction is usually called the *cuspidal edge*, cf. Example 7.16 below.

Example 7.15 (cone). If X is a cone, see Example 7.11, the “line” of striction shrinks to the single vertex O ; hence, our re-parametrization does not work in this case. One can check that this is the only case where $\tilde{\alpha}' \equiv 0$ in (7.14).

Example 7.16 (tangent developable). If the line of striction is a regular curve, we have

$$\lambda = 0 \quad \text{if and only if} \quad \mathbf{w} = \pm \mathbf{t},$$

i.e., the generatrix is tangent to the line of striction. Thus, roughly, a developable ruled surface is either a cylinder, or a cone, or the tangent developable of its line of striction (*aka* cuspidal edge), see Example 7.12. Simple local analysis shows that, generically, X does look like a family of semi-cubic parabolas with vertices on α .

Example 7.17 (Möbius strip). We take

$$\alpha(s) := (\cos s, \sin s, 0) \quad \mathbf{w}(s) := \mathbf{n} \cos \frac{s}{2} + \mathbf{b} \sin \frac{s}{2}.$$

Despite the “globally looking” parametrization, the surface is non-orientable.

7.5. Poincaré half-plane revisited (see §4.4). Consider the upper half-plane $\mathbb{H}$ with the first fundamental form defined in §4.4. In the notation of (4.25), it is

$$(7.18) \quad E = G = e^\rho, \quad F = 0, \quad \rho(x, y) := -2 \ln y.$$

By (5.18) and (5.23),

$$\Gamma_{11}^1 = \Gamma_{12}^2 = \Gamma_{22}^1 = 0, \quad \Gamma_{11}^2 = -\Gamma_{12}^1 = -\Gamma_{22}^2 = \frac{1}{y}, \quad K \equiv -1;$$

that is, $\mathbb{H}$ is a surface of constant negative Gaussian curvature. Since it is an abstract surface, we do not have e, f, g or $k_{1,2}, H$.

Geodesic equation (6.26) becomes

$$(7.19) \quad yx'' = 2x'y', \quad yy'' = y'^2 - x'^2,$$

not easy to solve. It is not so difficult, either. *If we assume that $x' \neq 0$* , the first equation is integrated to

$$x' = cy^2,$$

whereupon the second can probably be solved by a standard substitution.

However, we choose an alternative route. If we assume $x = \text{const}$, then the first equation becomes a tautology and the second one turns into

$$yy' \frac{dy'}{dy} = y'^2, \quad y' = cy, \quad y = Ce^{cs} = e^{c(s-s_0)}.$$

Therefore, vertical lines

$$(7.20) \quad x = x_0 = \text{const}, \quad y = e^{c(s-s_0)}$$

are parametrized geodesics; the parametrization is by the arc length if $c = \pm 1$.

To find the other geodesics, we cheat and describe the group of auto-isometries of $\mathbb{H}$ instead. For this, we cheat again and regard $\mathbb{H}$ as the upper half-plane

$$\mathbb{H} = \{z = x + iy \in \mathbb{C} \mid \text{Im } z > 0\} \subset \mathbb{C}.$$

We know that the group of biholomorphic automorphisms of $\mathbb{H}$ is

$$(7.21) \quad PSL(2, \mathbb{R}) = \left\{ z \mapsto \frac{az + b}{cz + d} \mid a, b, c, d \in \mathbb{R}, \begin{vmatrix} a & b \\ c & d \end{vmatrix} = 1 \right\} / \pm 1.$$

(It suffices to know that this group acts on $\mathbb{H}$, which is obvious; we never use the fact that there are no other automorphisms.) Any element of this group decomposes into a sequence of

- (1) horizontal shifts $z \mapsto z + a$, $a \in \mathbb{R}$,
- (2) dilations $z \mapsto rz$, $r \in \mathbb{R}_+$, or
- (3) inversion $z \mapsto -1/z$.

Since (7.18) is conformal to the Euclidean metric (see Remark 4.27), with respect to which holomorphic functions are known to be conformal, instead of (4.4) it suffices to check that, for a function $\varphi: z := x + iy \mapsto \tilde{z} := \tilde{x} + i\tilde{y}$,

$$\det \mathbf{l}_z = (\det d\varphi_z)^2 \det \mathbf{l}_{\tilde{z}}, \quad \text{or, equivalently,} \quad \tilde{y} = |\varphi'(z)|y,$$

i.e., in addition to angles, $d\varphi$ also preserves volume; in dimension 2, this is enough for isometry. This is straightforward for (1)–(3) above; hence, $PSL(2, \mathbb{R})$ does act on $\mathbb{H}$ by isometries. Since this action is flag-transitive, Corollary 6.39 (or, rather, its uniqueness part) implies that this is the full group of orientation preserving auto-isometries of $\mathbb{H}$. For orientation reversing, take, in addition, $z \mapsto -\bar{z}$.

Now, apply inversion (3) to the geodesic $s \mapsto (1, e^s)$ to get

$$(7.22) \quad x(s) = -\frac{1}{1+e^{2s}}, \quad y(s) = \frac{e^s}{1+e^{2s}}.$$

This is the semi-circle $(2x+1)^2 + 4y^2 = 1$ of radius $1/2$ centered at $z = -1/2$. It is parametrized by the arc length *with respect to the hyperbolic metric* (7.18).

Using (1) and (2) above, we can obtain from (7.22) all other semi-circles centered on the absolute $y = 0$; *a posteriori* one can check that they satisfy (7.19). (Note that the latter assumes the arc length parametrization; hence, we could not have used the “usual” parametrization of the semi-circles.) From the uniqueness part of Theorem 6.29 we conclude that these are all geodesics on $\mathbb{H}$.

Since both (7.20) and (7.22) are defined on the whole line $s \in \mathbb{R}$, the surface $\mathbb{H}$ is complete, see §6.5. Of course, it is not complete in the usual Euclidean metric, as it is not closed in $\mathbb{R}^2$.

7.5.1. Ideal triangles. One of the interesting phenomena of hyperbolic geometry is the existence of the so-called *ideal triangles* $\triangle ABC$ with all three vertices in the absolute $\{y = 0\} \cup \infty$. (Recall that formally the absolute is not part of $\mathbb{H}$.) All three angles of such a triangle are 0, and all three edges are parallel to each other! (The latter means that two lines meet at a point of the absolute; this is stronger than just being disjoint.) Since the action of (7.21) on the absolute is 3-transitive, any two ideal triangles are isometric. An example is the triangle bounded by

$$\text{two lines } x = \pm 1 \text{ and semi-circle } x^2 + y^2 = 1,$$

or that bounded by three semi-circles

$$(2x \pm 1)^2 + 4y^2 = 1, \quad x^2 + y^2 = 1.$$

In §8.1.5 below we will see that each ideal triangle has area π , which is maximal possible for a hyperbolic triangle (assuming $K = -1$).

Accidentally, in hyperbolic geometry there is no similarity: a triple A, B, C of angles satisfying $A + B + C < \pi$ (see §8.1.5 below) determines a hyperbolic triangle up to isometry. The hyperbolic plane has a natural unit of length, which is $|K|^{-1/2}$. This is quite similar to the sphere, where the radius $R = K^{-1/2}$ is a natural unit and, hence, the sides of spherical triangles are also regarded as angles.

Hyperbolic and spherical trigonometry is a fascinating subject; alas, nowadays it is considered more appropriate for recreational mathematics.

8. ADVANCED TOPICS

Most likely, not all of these topics will be covered in class.

8.1. The Gauss–Bonnet theorem. Undoubtedly, the principal statement of this section is the most beautiful result of all classical differential geometry: it relates hard analysis to simple topological invariants.

Theorem 8.1 (Gauss–Bonnet). *Let R be an oriented compact region in a regular surface X , bounded by a piecewise regular curve ∂R , possibly empty. Then,*

$$\iint_R K \, d\text{vol} + \int_{\partial R} \kappa_g ds + \sum \theta_i = 2\pi\chi(R).$$

Certainly, a lot of explanation is due before we can prove this theorem.

Since κ_g is orientation dependent, the orientation of ∂R matters. We *always* assume that ∂R has the *boundary orientation*, so that $\mathbf{n}_g$ points inside R .

Since ∂R is assumed *piecewise* regular, it may have “corners”, *i.e.*, points $\gamma(s_i)$ where two regular pieces meet. The tangents do not need to agree, and we define the *outer angle* θ_i as

$$\theta_i = \text{angle from } \mathbf{t}(s_i - 0) \text{ to } \mathbf{t}(s_i + 0); \quad -\pi \leq \theta_i \leq \pi.$$

In applications one often considers the conventional *inner angles* $\bar{\theta}_i := \pi - \theta_i$ of the “polygon” R , but outer angles are more convenient for the statement.

Recall that one way to define the curvature $\kappa_g = d\theta/ds$, see (6.23), is as the derivative of the angle function. At a corner s_i , this function θ has a jump by θ_i , and a physicist would say that ∂R has δ -curvature $\theta_i\delta(s - s_i)$, in the sense of *Dirac δ -function*

$$\delta(x) = \begin{cases} 0, & \text{if } x \neq 0, \\ \infty, & \text{if } x = 0, \end{cases} \quad \int_{-\infty}^{\infty} \delta(x) \, dx = 1.$$

Such a function does not exist but, in the realm of *generalized functions*, it is indeed the generalized derivative of the *Heaviside step function*. This way, we could have written $\int_{\partial R} + \sum$ as a single integral, and this sum is indeed the total rotation angle $2\pi \text{rot}(\partial R)$ as in (6.23). However, since

- we are trying to do rigorous mathematics,
- in many applications, it is the angles that are important, *cf.* §8.1.5 below,
- eventually, it is the angles that give rise to the mysterious right hand side of the equation, see the proof in §8.1.3 below,

we prefer to stick to the classical form of the theorem.

Finally, the *Euler characteristic* $\chi(R)$ in the right hand side is a certain purely topological invariant of the region R ; it is briefly discussed in §8.1.2 below.

8.1.1. The local version. Let us prove [Theorem 8.1](#) for a “small” region R . More precisely, assume that R is bounded by a simple closed piecewise regular curve and fits into a single normal polar chart (r, θ) , see §6.4.1. In particular,

$$E \equiv 1, \quad F \equiv 0, \quad d\text{vol} = \sqrt{G} \, dr \, d\theta, \quad K = -\frac{(\sqrt{G})_{rr}}{\sqrt{G}}.$$

Since $\chi(R) = 1$, we can rewrite the identity in the form

$$(8.2) \quad \int_{\partial R} \kappa_g ds + \sum \theta_i = 2\pi + \iint_R (\sqrt{G})_{rr} dr \, d\theta,$$

and the key to success is the correct interpretation of each term.

The left hand side of (8.2) is the rotation angle

$$2\pi \operatorname{rot}(\partial R).$$

(In §6.2.1 it was denoted by $\Delta\theta$, but now θ is reserved for a coordinate. Obvious generalization of this concept to piecewise regular curves is left to the reader.)

The term 2π in the right hand side is what would have been given by Hopf's Umlaufsatz Theorem 2.15 should we have worked in the true Euclidean (r, θ) -plane. A way to get this number (to which our proof of Theorem 2.15 fully applies) is as follows. Consider a *global* (defined on a neighborhood of the whole region R) unit vector field $\mathbf{w}$ and write the unit tangent vector to ∂R as

$$\mathbf{t}(s) = \mathbf{w} \cos \varphi(s) + \mathbf{w}^\perp \sin \varphi(s),$$

where $\varphi(s)$ is a *real valued* angle function, cf. (2.3). Then $\Delta\varphi = 2\pi$.

Remark 8.3. In §2.1 and §2.4.1 we used the “constant” vector field $\mathbf{w} = \mathbf{i}$ and standard Euclidean metric (angle) on the plane. However, this does not matter *as long as $\mathbf{w}$ is global*. This can easily be proved using the idea of homotopy. We can *continuously* change $\mathbf{w}$ to $\mathbf{i}$ and the first fundamental form $\mathbf{l}$, to the standard Euclidean dot product. Then, $\Delta\varphi$ will change continuously. On the other hand, at each moment $\Delta\varphi/2\pi$ is an integer; hence, it will have to stay constant.

A curious student may ask why this argument does not apply to a field $\mathbf{w}$ defined on ∂R only. Since ∂R is *not simply connected*, this field $\mathbf{w}$ may be twisted (*i.e.*, not homotopic to $\mathbf{i}$) in the first place, like $\mathbf{t}$ is. A formal justification is in algebraic topology; hence, we just refer to common sense. In fact, $\operatorname{rot}(\partial R)$ is precisely the *obstruction* to a homotopy between $\mathbf{t}$ and $\mathbf{i}$ on ∂R .

Now, the problem is that $\mathbf{w}$ does not need to be parallel along ∂R , and our wrong computation resulting in 2π needs a correction, which apparently is the last term of (8.2). Let $\mathbf{u}$ be a *parallel* vector field on ∂R and $\mathbf{v} := \mathbf{u}^\perp$. Then, on ∂R ,

$$\mathbf{w}(s) = \mathbf{u}(s) \cos \psi(s) + \mathbf{v}(s) \sin \psi(s)$$

for some function ψ , and the true rotation angle is $\Delta\varphi + \Delta\psi$, cf. §6.2.1. We have

$$\frac{D\mathbf{w}}{ds} = \frac{d\psi}{ds}(-\mathbf{u} \sin \psi + \mathbf{v} \cos \psi) = \frac{d\psi}{ds} \mathbf{w}^\perp;$$

hence,

$$\frac{d\psi}{ds} = \frac{D\mathbf{w}}{ds} \cdot \mathbf{w}^\perp \quad \text{and} \quad \Delta\psi = \int_{\partial R} \left(\frac{D\mathbf{w}}{ds} \cdot \mathbf{w}^\perp \right) ds.$$

To complete the proof, we take $\mathbf{w} := \mathbf{r}_r$, hence $\mathbf{w}^\perp = \mathbf{r}_\theta / \sqrt{G}$, and parametrize ∂R via $r(s), \theta(s)$. By (6.1) and (4.26), (5.17),

$$\frac{D\mathbf{r}_r}{ds} \cdot \frac{\mathbf{r}_\theta}{\sqrt{G}} = \left(\Gamma_{11}^2 \frac{dr}{ds} + \Gamma_{12}^2 \frac{d\theta}{ds} \right) \left(\mathbf{r}_\theta \cdot \frac{\mathbf{r}_\theta}{\sqrt{G}} \right) = \frac{G_r}{2\sqrt{G}} \frac{d\theta}{ds} = (\sqrt{G})_r \frac{d\theta}{ds}$$

(we do not need the $\mathbf{r}_r$ -component since $\mathbf{r}_r \cdot \mathbf{r}_\theta = 0$). Hence, the missing correction term is

$$\Delta\psi = \int_{\partial R} (\sqrt{G})_r \frac{d\theta}{ds} ds = \int_{\partial R} (\sqrt{G})_r d\theta \stackrel{\text{Green}}{=} \iint_R (\sqrt{G})_{rr} dr d\theta.$$

At the last step, we refer to Green's theorem §2.4.5. □

8.1.2. *Euler characteristic.* The term is due to Euler's celebrated theorem about planar graphs or convex polyhedra; in modern terms, it states that $\chi(S^2) = 2$. The generalization is mostly due to Poincaré.

Consider a compact surface X . Here, we do allow surfaces with boundary, even though we never defined them. One can think about a compact region R as in the statement of Theorem 8.1. Subdivide X into several “convex” polyhedra, bounded by curves, called *edges*, which meet at points called *vertices*. (Here, by “convex” we mean something that looks like an ordinary triangle, quadrilateral, pentagon, *etc.*, without “holes”; it is difficult to give a rigorous definition at this stage.) We must do this so that any two polyhedra either are disjoint or intersect at one common edge or vertex. Consider the numbers

$$\#_0 := \#(\text{vertices}), \quad \#_1 := \#(\text{edges}), \quad \#_2 := \#(\text{polyhedra})$$

and define the *Euler characteristic* of X via

$$\chi(X) := \#_0 - \#_1 + \#_2 \in \mathbb{Z}.$$

Remarkably, this integer is independent of the chosen triangulation; furthermore, it is a *topological* invariant of X . A rigorous proof is done in algebraic topology, using the so-called *Betti numbers*, where χ is also extended to all decent topological spaces and is shown to be a *homotopy* invariant.

A naïve commonly known proof, going back to Euler, is based on the assumption that any two triangulations have a common subdivision: this is fairly obvious but not true. Nevertheless, this proof gives us a good explanation of the invariance: it suffices to consider one *elementary* subdivision, *viz.* adding a vertex in the middle of an edge or adding a “diagonal” of one of the polygons, and check how the quantities involved change. This is left to the reader.

We can extend χ to the 1-dimensional objects (just let $\#_2 = 0$), but we do not get anything interesting:

$$\chi(\text{segment}) = 1, \quad \chi(S^1) = 0.$$

Euler characteristic is obviously additive:

$$\chi(X \cup Y) = \chi(X) + \chi(Y) - \chi(X \cap Y).$$

In particular, if X and Y are two surfaces glued together along common boundary *circles*, we have $\chi(X \cup Y) = \chi(X) + \chi(Y)$.

A simple computation shows that

$$\chi(\text{disk}) = 1, \quad \chi(S^2) = 2, \quad \chi(\text{cylinder}) = 0.$$

Then,

$$\chi(X \setminus D^2) = \chi(X) - 1, \quad \chi(X \# T^2) = \chi(X) - 2,$$

where

- $X \setminus D^2$ is X with a *hole* (an open disk removed), and
- $X \# T^2$ is X with a *handle*: two holes connected by an extra cylinder.

For example, $S^2 \# T^2$ is the torus $T^2 = S^1 \times S^1$.

A *classification theorem* states that any compact orientable surface is the sphere with a certain number $g \geq 0$ (called *genus*) of handles and a number $n \geq 0$ of holes. One possible notation is $S_{g,n}$ and $S_g := S_{g,0}$. From the above,

$$(8.4) \quad \chi(S_g) = 2 - 2g, \quad \chi(S_{g,n}) = 2 - 2g - n.$$

In particular, the sphere S^2 is the only closed surface with $\chi > 0$.

8.1.3. *Proof of Theorem 8.1.* We subdivide R into a number of smaller “polygonal” regions R_i , each covered by §8.1.1, apply the theorem to each region R_i ,

$$\iint_{R_i} K d\text{vol} + \int_{\partial R_i} \kappa_g ds + \sum \theta_{ij} = 2\pi,$$

and sum up the results.

The double integrals sum up to $\iint_R$.

The line integrals sum up to $\int_{\partial R}$: each inner edge appears in the sum twice, with two opposite orientations, and, since κ_g is orientation dependent, these summands cancel each other, leaving only the integrals over the outer edges, *i.e.*, ∂R .

The right hand side sums up to $2\pi \cdot \#_2$.

To analyze the remaining sum, consider a vertex v and denote by $\deg v$ its *degree* (or *valency*), *i.e.*, the number of edges incident to this vertex. Note that the number of regions R_i incident to v is $\deg v$ if v is inner and $(\deg v - 1)$ otherwise (as the “outer” region does not count). Let θ_k be the outer angles of these regions and denote by $\bar{\theta}_k := \pi - \theta_k$ the corresponding inner angles. We consider three cases.

If v is an inner vertex of R , then

$$\sum \theta_k = \pi \deg v - \sum \bar{\theta}_k = \pi(\deg v - 2) \quad \text{as} \quad \sum \bar{\theta}_k = 2\pi.$$

If v is a smooth point on the boundary ∂R , then

$$\sum \theta_k = \pi(\deg v - 1) - \sum \bar{\theta}_k = \pi(\deg v - 2) \quad \text{as} \quad \sum \bar{\theta}_k = \pi.$$

Finally, if v is a corner of R , with the outer angle $\theta(v)$, then

$$\sum \theta_k = \pi(\deg v - 1) - \sum \bar{\theta}_k = \pi(\deg v - 1) - \bar{\theta}(v) = \pi(\deg v - 2) + \theta(v).$$

Hence,

$$\sum \theta_{ij} = \sum \theta(v_i) + \pi \sum (\deg v_i - 2),$$

the summation running over all vertices v_i . The first term is precisely the sum in the left hand side of Theorem 8.1. Taking the second term to the right hand side, we obtain

$$2\pi \left(\#_2 - \frac{1}{2} \sum \deg v_i + \#_0 \right) = 2\pi \chi(R),$$

where we recall that $\sum \deg v_i = 2\#_1$ in any graph. $\square$

8.1.4. *Holonomy revisited.* Recall, see (6.24), that the holonomy P_γ along a closed curve γ is essentially the angle

$$P_\gamma = (-2\pi \text{rot } \gamma \bmod 2\pi) \in SO(1).$$

Hence, if $\gamma = \partial R$ is a (piecewise regular) simple closed curve bounding a region R , Theorem 8.1 yields

$$P_\gamma = \iint_R K d\text{vol} \bmod 2\pi \in SO(1).$$

We can use this observation to re-interpret the Gaussian curvature K as the “infinitesimal holonomy per unit of area.” Physicists would call it *volume derivative*, *cf.* the “physical” definitions of the rotor $\nabla \times \mathbf{F}$ and divergence $\nabla \cdot \mathbf{F}$ of a vector field $\mathbf{F}$ as volume derivatives of its circulation and flux, respectively.

More precisely, K measures the deviation, per unit of area, of the rotation angle $2\pi \text{rot } \gamma$ from the 2π predicted by Hopf’s Umlaufsatz Theorem 2.15.

Corollary 8.5 (of [Theorem 8.1](#)). *For a point $p \in X$, one has*

$$K(p) = \lim_{R \rightarrow p} \frac{2\pi - 2\pi \operatorname{rot}(\partial R)}{\operatorname{Area}(R)},$$

where $R \ni p$ is a compact region bounded by a piecewise regular simple closed curve ∂R and the limit is in the sense, e.g., $\operatorname{diameter}(R) \rightarrow 0$. $\square$

8.1.5. Spherical and hyperbolic triangles. Let $R := \triangle A_1 A_2 A_3$ be a *geodesic triangle* (i.e., region bounded by three geodesics intersecting at three points) on a surface X of constant Gaussian curvature K . We assume that R is an “honest” triangle, homeomorphic to a disk, so that $\chi(R) = 1$. Traditionally, we let $0 \leq A_i \leq 2\pi$ also stand for the *inner* angles of the triangle, so that the corresponding outer angles are $\theta_i := \pi - A_i$. Since the sides are geodesics, $\kappa_g \equiv 0$ and [Theorem 8.1](#) reads

$$K \operatorname{Area}(\triangle A_1 A_2 A_3) = A_1 + A_2 + A_3 - \pi.$$

If $K = 0$ (the plane, assuming that X is complete and simply connected, see [§8.4](#) below), we get the classical $A_1 + A_2 + A_3 = \pi$ from high school geometry.

If $K > 0$ (e.g., a sphere), we always have $A_1 + A_2 + A_3 > \pi$, and the *spherical defect* $A_1 + A_2 + A_3 - \pi$ is proportional to the area:

$$\operatorname{Area}(\triangle A_1 A_2 A_3) = R^2(A_1 + A_2 + A_3 - \pi), \quad R := \sqrt{K^{-1}} \text{ (the radius)}.$$

The sphere itself can be regarded as a triangle with angles, e.g., $3\pi/2, 3\pi/2, 2\pi$.

If $K < 0$, then $A_1 + A_2 + A_3 < \pi$ and the *hyperbolic defect* $\pi - (A_1 + A_2 + A_3)$ is proportional to the area. If X is complete and simply connected, then it is the so-called *hyperbolic plane* (see [§8.4](#) below); declaring geodesics lines, one obtains a model of non-Euclidean geometry. The area of a triangle does not exceed $|K|^{-1}\pi$. The largest triangles are the ideal ones (see [§7.5.1](#)); their area is $|K|^{-1}\pi$.

Accidentally, let us also discuss the existence of a *geodesic bigon*, i.e., a region R homeomorphic to a disk and bounded by two geodesics that meet at two points. We no longer assume K constant. Denoting by $0 \leq \bar{\theta}_i \leq 2\pi$ the two *inner* angles, we have

$$\iint_R K \, d\operatorname{vol} = \bar{\theta}_1 + \bar{\theta}_2.$$

Recalling that two *distinct* geodesics cannot meet at angle 0, see [Theorem 6.29](#), we conclude that

$$(8.6) \quad \text{if } K \leq 0 \text{ at all points, geodesic bigons do not exist.}$$

Bigons do exist on the sphere: a pair of meridians.

Likewise, if $K \leq 0$, there are no *monogons* or “0-gons” (closed geodesics), still assuming that the curves bound a topological disk.

8.2. Second variation of the arc length. In the settings and notation of [§6.3.1](#), let us repeat the computation of the *first variation* $\ell'(t)$ more carefully, without substituting $t = 0$ right away:

$$\ell'(t) = \int_0^l \frac{\partial}{\partial t} \left| \frac{\partial \gamma}{\partial s} \right| ds = \int_0^l \frac{1}{|\gamma'|} \left(\gamma' \cdot \frac{D}{\partial t} \frac{\partial \gamma}{\partial s} \right) ds = \int_0^l \frac{1}{|\gamma'|} \left(\gamma' \cdot \frac{D h_t}{\partial s} \right) ds,$$

where $h_t := \partial \gamma / \partial t$. (In [§6.3.1](#), we only used h_0 .) Now, we make extra assumptions:

(1) γ_0 is a geodesic parametrized by the arc length,

$$|\gamma'_0| \equiv 1, \quad \frac{D\gamma'_0}{ds} \equiv 0,$$

so that $\ell'(0) = 0$ (see §6.3.1);

(2) γ is a *proper normal variation*,

$$\gamma(0, t) = \text{const}, \quad \gamma(l, t) = \text{const} \quad (\text{proper}), \quad h_0 \cdot \gamma'_0 \equiv 0 \quad (\text{normal}).$$

Under these assumptions, compute the *second variation*

$$\ell''(0) = \int_0^l \frac{\partial}{\partial t} \left(\frac{1}{|\gamma'|} \right) \left(\gamma' \cdot \frac{Dh_t}{\partial s} \right) ds + \int_0^l \frac{1}{|\gamma'_0|} \frac{\partial}{\partial t} \left(\gamma' \cdot \frac{Dh_t}{\partial s} \right) ds,$$

all at $t = 0$.

In the first integral,

$$\left(\gamma' \cdot \frac{Dh_t}{\partial s} \right) \Big|_{t=0} = \frac{d}{ds} (\gamma'_0 \cdot h_0) - \left(\frac{D\gamma'_0}{ds} \cdot h_0 \right) = 0,$$

so that we do not need to compute the derivative in front of this expression. For the second,

$$\begin{aligned} \frac{\partial}{\partial t} \left(\gamma' \cdot \frac{Dh_t}{\partial s} \right) \Big|_{t=0} &= \left(\frac{D}{\partial t} \frac{\partial \gamma}{\partial s} \cdot \frac{Dh_t}{\partial s} \right) \Big|_{t=0} + \left(\gamma'_0 \cdot \frac{D}{\partial t} \frac{Dh_t}{\partial s} \right) \Big|_{t=0} \\ &= \left| \frac{Dh_0}{ds} \right|^2 + \underbrace{\left(\gamma'_0 \cdot \frac{D}{\partial s} \frac{Dh_t}{\partial t} \right) \Big|_{t=0}}_{\#1} + K \underbrace{\gamma'_0 \cdot ((\gamma'_0 \times h_0) \times h_0)}_{\#2}. \end{aligned}$$

Here, we refer to [Theorem 6.8](#) and, with a slight abuse, treat $\mathbf{r} := \gamma(s, t)$ as a parametrization, so that $\mathbf{r}_s = \gamma'$ and $\mathbf{r}_t = h_t$. A more rigorous reference would be §6.1.2. The last term is easy, *e.g.*, by (1.7):

$$\#2 = \gamma'_0 \cdot (h_0(\gamma'_0 \cdot h_0) - \gamma'_0(h_0 \cdot h_0)) = -(\gamma'_0 \cdot \gamma'_0)(h_0 \cdot h_0) = -|h_0|^2.$$

For the middle one, we argue backwards:

$$\#1 = \frac{d}{ds} \left(\gamma'_0 \cdot \frac{Dh_t}{\partial t} \right) \Big|_{t=0} - \left(\frac{D\gamma'_0}{ds} \cdot \frac{Dh_t}{\partial t} \right) \Big|_{t=0};$$

upon integration, it gives us

$$\left(\gamma'_0 \cdot \frac{Dh_t}{\partial t} \Big|_{t=0} \right) \Big|_{s=0}^l = 0 \quad \text{by assumption (2)}.$$

Summarizing, we find that, under assumptions (1), (2) above, the second variation of the arc length is

$$(8.7) \quad \ell''(0) = \int_0^l \left(\left| \frac{Dh_0}{ds} \right|^2 - K|h_0|^2 \right) ds,$$

where h_0 is the variation vector field along the original geodesic γ_0 . Since h_0 is normal, it has the form $h_0(s) = \eta(s)\mathbf{n}_g(s)$ and (8.7) becomes, *cf.* (6.18),

$$(8.8) \quad \ell''(0) = \int_0^l (\eta'^2 - K\eta^2) ds.$$

Note that all these arguments are quite standard in calculus of variations: we differentiate under $\int$ and integrate by parts (that is what $\int$ (6.6) is). The ultimate

goal is expressing the result in terms of the original “point” γ_0 and “direction” h_0 ; as in calculus, we expect the result linear/quadratic for the first/second variation.

Apart from a nice exercise in calculus of variations, we obtain a curious result about long geodesics on surfaces of positive curvature. (To recover the constants, think about big semi-circles on a sphere of radius $R = \sqrt{K^{-1}}$.)

Theorem 8.9 (Bonnet). *Assume that $K(s) \geq \delta > 0$ at each point of a geodesic γ . Then γ cannot be minimal if*

$$\text{length}(\gamma) > \frac{\pi}{\sqrt{\delta}}.$$

Proof. Let $l := \text{length}(\gamma)$ and consider the (obviously proper) normal variation with the variation field

$$h_0(s) := \mathbf{n}_g \sin \frac{\pi s}{l}.$$

By (8.8),

$$\ell''(0) = \int_0^l \left(\frac{\pi^2}{l^2} \cos^2 \frac{\pi s}{l} - K \sin^2 \frac{\pi s}{l} \right) ds \leq \frac{\pi^2 - l^2 \delta}{2l} < 0.$$

By the second derivative test, arbitrary close to γ there are shorter curves. $\square$

Remark 8.10. In the proof, we silently assumed that any vector field h_0 on γ_0 gives rise to a variation. This is indeed true, *e.g.*, we can take

$$\gamma(s, t) := \exp_{\gamma_0(s)}(th_0(s)).$$

A more conceptual explanation of [Theorem 8.9](#) is given in [§8.3.3](#) below. The following statement is a consequence of [Theorems 8.9](#) and [6.41](#).

Corollary 8.11 (Bonnet). *Assume that $K \geq \delta > 0$ at each point of a complete surface X . Then X is compact (equivalently, closed, as we never consider surfaces with boundary) and*

$$\text{diameter}(X) \leq \frac{\pi}{\sqrt{\delta}}. \quad \triangleleft$$

By the Gauss–Bonnet [Theorem 8.1](#) and (8.4), assuming X orientable, it must be diffeomorphic (not necessarily isometric) to the sphere.

8.3. Jacobi fields. Let us consider a *geodesic variation*, *i.e.*, a family

$$\gamma_t(s) := \gamma(s, t) \quad \text{s. t.} \quad \gamma_t(s) \text{ is a parametrized geodesic for each fixed } t.$$

In general, we do *not* assume that s is the arc length; it can be a multiple thereof (depending on t). A *Jacobi field* on a geodesic $\gamma_0(s)$ is the variation vector field

$$J := \frac{\partial \gamma}{\partial t} \Big|_{t=0}$$

of a geodesic variation. Using [Theorem 6.8](#), we have

$$\frac{D}{\partial s} \frac{D}{\partial s} \frac{\partial \gamma}{\partial t} = \frac{D}{\partial s} \frac{D}{\partial t} \frac{\partial \gamma}{\partial s} = \frac{D}{\partial t} \frac{D}{\partial s} \frac{\partial \gamma}{\partial s} - K \left(\gamma' \times \frac{\partial \gamma}{\partial t} \right) \times \gamma'.$$

Substituting $t = 0$ and disregarding the red term (since all $\gamma_t(s)$ are geodesics),

$$(8.12) \quad \frac{D^2 J}{ds^2} + K(\gamma'_0 \times J) \times \gamma'_0 = 0.$$

Decomposing $J = \tau\gamma'_0 + \nu(\gamma'_0)^\perp$ into its tangential and normal components and recalling that both γ'_0 and $(\gamma'_0)^\perp$ are parallel along γ_0 , we convert vector equation (8.12) into two independent scalar ones:

$$(8.13) \quad \tau'' = 0, \quad \nu'' + \lambda^2 K \nu = 0, \quad \lambda := |\gamma'_0| = \text{const.}$$

We arrive at the following statement.

Theorem 8.14. *Let $\gamma_0: [0, l] \rightarrow X$ be a compact parametrized geodesic. Then, a vector field J on γ_0 is a Jacobi field if and only if it satisfies (8.12).*

In particular, there is a unique Jacobi field J satisfying any pair of initial conditions

$$J(0) = J_0, \quad \left. \frac{DJ}{ds} \right|_{s=0} = J'_0.$$

Hence, Jacobi fields constitute a vector space of dimension 4.

Proof. The necessity has just been proved, and the existence on the whole domain $[0, l]$ follows from the fact that (8.12) is linear. For the sufficiency, we need to find a geodesic variation whose variation vector field satisfies a given pair of initial conditions. This variation can be constructed from radial geodesics:

$$\gamma(s, t) := \exp_{\alpha(t)}(s\mathbf{v}(t)),$$

where $\alpha: (-\varepsilon, \varepsilon) \rightarrow X$ is a wisely chosen curve and $\mathbf{v}(t)$ is a wisely chosen field on α :

$$\alpha(0) = \gamma_0(0), \quad \alpha'(0) = J_0; \quad \mathbf{v}(0) = \gamma'_0(0), \quad \left. \frac{D\mathbf{v}}{dt} \right|_{t=0} = J'_0.$$

By the compactness, γ is defined on $[0, l] \times (-\varepsilon, \varepsilon)$ for $\varepsilon \ll 1$. $\square$

Note that the construction of Remark 8.10 no longer applies: there, geodesics are the transverse sections $\gamma(\text{const}, t)$, whereas here we want the curves $\gamma(s, \text{const})$.

Equation (8.12) has an interesting corollary:

$$\left(J_1 \cdot \frac{DJ_2}{ds} \right) - \left(\frac{DJ_1}{ds} \cdot J_2 \right) = \text{const}$$

for any two Jacobi fields J_1, J_2 . For proof, differentiate and use the symmetry (6.14) of the curvature tensor (6.13).

According to (8.13), the tangential component τ of a Jacobi field is not very interesting: it is always a linear function, and the corresponding geodesic variations are merely re-parametrizations $\gamma(s, t) := \gamma_0(ats + bt)$. Therefore, most of the time we deal with *normal* (or *orthogonal*) *Jacobi fields*.

8.3.1. Conjugate points. The initial value problem in Theorem 8.14 always has a unique solution: this is a fundamental fact about ODE's. Thus, let us consider the more subtle *boundary value problem*. A motivating example is the equation

$$\nu'' + \nu = 0, \quad \nu(s_1) = \nu_1, \quad \nu(s_2) = \nu_2.$$

Since the general solution to the equation *per se* is

$$\nu(s) = A \sin(s - s_0), \quad A, s_0 = \text{const},$$

the boundary value problem has a unique solution *provided that* $s_2 \neq s_1 \bmod \pi$. Otherwise, if $s_2 = s_1 + \pi$, a solution exists if and only if $\nu_2 = -\nu_1$; in fact, any solution vanishing at s_1 also vanishes at all $s_1 + \pi k$, $k \in \mathbb{Z}$, and all such solutions are scalar multiples of any one of them.

Definition 8.15. Two points $\gamma(s_1), \gamma(s_2)$, $s_1 \neq s_2$, on a parametrized geodesic $\gamma: [a, b] \rightarrow X$ are called *conjugate* if there exists a Jacobi field $J \not\equiv 0$ such that $J(s_1) = J(s_2) = 0$. We use flexible terminology:

- two points $\gamma(s_1), \gamma(s_2) \in X$ are conjugate with respect to (or along) γ , or
- two points $\gamma(s_1), \gamma(s_2)$ on γ are conjugate, or even
- two values s_1, s_2 of the parameter (usually arc length) are conjugate.

(The presence of λ^2 in (8.13) implies that conjugate points do not depend on the parametrization.) Often, points $\gamma(s)$ conjugate to $\gamma(0)$ are called just conjugate.

Intuitively, the first conjugate point $s_1 > 0$ on a geodesic $\gamma: [0, l] \rightarrow X$ is the point where another geodesic $\bar{\gamma}$ through $\gamma(0)$, *sufficiently close* to γ , meets γ again. (For example, meridians on a surface of revolution, *e.g.*, sphere.) However, this should not be understood literally: formally, the assertion is about “infinitesimally close” geodesics $\bar{\gamma}$, which are non-existent beasts.

On surfaces, the structure of conjugate points is quite simple. As mentioned above, (8.13) implies that the tangential component of any Jacobi field is linear; hence, unless identically zero, it cannot vanish at more than one point. Therefore, we confine ourselves to *normal* Jacobi fields. Any such field J has a discrete zero locus $Z(J)$, possibly empty or consisting of a single point, and any other normal Jacobi field $\bar{J}$ vanishing at a point $s_0 \in Z(J)$ is a constant scalar multiple of J ; in particular, $Z(\bar{J}) = Z(J)$. All points of $Z(J)$ are conjugate to each other.

In order to analyze the existence of conjugate points, we need a special case of the [Sturm–Picone comparison theorem](#).

Lemma 8.16. For $i = 1, 2$, let $\nu_i(s)$ be a solution to

$$\nu_i''(s) + K_i(s)\nu_i(s) = 0, \quad \nu_i(0) = 0,$$

and let $c_i \in (0, \infty]$ be the first zero of ν_i . If $K_1(s) \leq K_2(s)$ for all s , then $c_2 \leq c_1$.

Proof. Without loss of generality, we can assume that

$$\nu_1(s) > 0 \quad \text{for } 0 < s < c_1; \quad \text{hence,} \quad \nu_1'(0) > 0, \quad \nu_1'(c_1) < 0.$$

(Neither derivative can vanish, as otherwise $\nu_1 \equiv 0$.) Assume also that $\nu_2(s) > 0$ for all $s \in (0, c_1]$. Integrating by parts, we have

$$\begin{aligned} 0 &= \int_0^{c_1} (\nu_1(\nu_2'' + K_2\nu_2) - \nu_2(\nu_1'' + K_1\nu_1)) ds \\ &= (\nu_1\nu_2' - \nu_1'\nu_2) \Big|_0^{c_1} + \int_0^{c_1} (K_2 - K_1)\nu_1\nu_2 ds > 0. \quad \square \end{aligned}$$

Corollary 8.17. Let γ be a geodesic parametrized by the arc length.

- (1) If $K \leq 0$ on γ , then γ has no conjugate points.
- (2) If $K \geq \delta > 0$ on γ , then γ has a conjugate point $s \leq \pi/\sqrt{\delta}$.

The corollary is inconclusive if $K \geq 0$ but is not bounded by a constant $\delta > 0$. $\triangleleft$

8.3.2. Differential of the exponential map. Recall (cf. §5.4), that, given a point $p \in X$, the tangent plane $T_{\mathbf{v}}(T_p X)$ at a point $\mathbf{v} \in T_p X$ can be naturally identified with $T_p X$ itself. (Formally, this identification is the differential of the translation $\mathbf{u} \mapsto \mathbf{u} + \mathbf{v}$, but that is an overkill.) Hence, the differential of the exponential map can be regarded as a linear transformation

$$(8.18) \quad (d\exp_p)_{\mathbf{v}}: T_p(X) \rightarrow T_{\exp_p(\mathbf{v})}X.$$

Recall also that the exponential map itself is defined *via*

$$\exp_p(\mathbf{v}) := \gamma_{\mathbf{v}}(1), \quad \text{where } \gamma_{\mathbf{v}} \text{ is a geodesic, } \gamma_{\mathbf{v}}(0) = p, \gamma'_{\mathbf{v}}(0) = \mathbf{v}.$$

Theorem 8.19. *In the notation above, the differential $d\exp_p$ is given by*

$$(d\exp_p)_{\mathbf{v}}(\mathbf{u}) = J(1),$$

where J is a Jacobi field on $\gamma_{\mathbf{v}}$ such that $J(0) = 0$ and $\frac{DJ}{ds}(0) = \mathbf{u}$. $\triangleleft$

This can be described somewhat more explicitly: under identification (8.18),

$$(d\exp_p)_{\mathbf{v}}: \quad \mathbf{v} \mapsto \gamma'_{\mathbf{v}}(1), \quad \mathbf{v}^{\perp} \mapsto \sqrt{G}(1)(\gamma'_{\mathbf{v}}(1))^{\perp},$$

where the function $\sqrt{G}: [0, 1] \rightarrow \mathbb{R}$ is formally found from

$$(\sqrt{G})'' + |\mathbf{v}|^2 K(\gamma_{\mathbf{v}}(s))\sqrt{G} = 0, \quad \sqrt{G}(0) = 0, \quad (\sqrt{G})'(0) = |\mathbf{v}|,$$

cf. (5.22) or the proof of Corollary 6.39. This is, indeed, $\sqrt{G(\gamma_{\mathbf{v}}(s))}$ provided that the latter makes sense, i.e., that G never vanishes along $\gamma_{\mathbf{v}}$ (or, equivalently, $\gamma_{\mathbf{v}}$ is covered by a normal polar chart).

To prove Theorem 8.19, consider the variation

$$(s, t) \mapsto s(\mathbf{v} + t\mathbf{u}), \quad s \in [0, 1], \quad t \in (-\varepsilon, \varepsilon),$$

compose it with $\exp_p$ to get a variation of $\gamma_{\mathbf{v}}$ consisting of radial geodesics (hence, Jacobi fields), and compute $\partial/\partial t$ at $(1, 0)$. Since, by the very assumption, $\exp_p$ is defined at $\mathbf{v}$, it is also defined in a neighborhood of the segment $s\mathbf{v}$, $s \in [0, 1]$.

Corollary 8.20. *The exponential map $\exp_p$ is a local diffeomorphism at a point $\mathbf{v} \in T_p X$ if and only if $\gamma_{\mathbf{v}}(1)$ is not conjugate to $\gamma_{\mathbf{v}}(0)$ along $\gamma_{\mathbf{v}}$.* $\triangleleft$

8.3.3. Critical points of the arc length. The second variation (8.8) of the arc length is a quadratic form in ν . Mimicking the finite dimensional case (cf. §5.6), let us convert it to a symmetric bilinear form (on the infinite dimensional space of proper normal variation vector fields on a geodesic γ)

$$(8.21) \quad \mathcal{H}(\mu, \nu) := \int_0^l (\mu' \nu' - K \mu \nu) ds \stackrel{\text{by parts}}{=} - \int_0^l \mu(\nu'' + K \nu) ds,$$

where we used integration by parts and the assumption that both normal variations are proper, $\mu(0) = \nu(0) = \mu(l) = \nu(l) = 0$. Comparing this to (8.13), we see that

$$\nu \in \text{rad } \mathcal{H} \quad \text{iff} \quad \nu \text{ is a Jacobi field.}$$

Since ν is assumed proper,

a geodesic γ is a degenerate critical point iff its endpoints are conjugate.

In fact, the index of γ (see §5.6) as a critical point of the arc length functional is

$$(8.22) \quad \text{ind } \gamma = \#\{\text{conjugate points on } \gamma\}$$

The proof would take us too far; e.g., we need to specify more carefully the space of variations allowed, cf. the proof of Theorem 8.23(2) below. An excellent treatment is found in Milnor's book.¹

Thus, we confine ourselves to a classical partial result that decides whether a given geodesic is *locally* minimal, i.e., shortest among all *sufficiently close* curves. (Do not mix this with local minimal geodesics considered in §6.4.2.)

¹John Milnor, "Morse theory." *Annals of Mathematics Studies*, vol. 51, Princeton University Press, Princeton, N. J., 1963

Theorem 8.23 (Jacobi). *Consider a geodesic $\gamma_0: [0, l] \rightarrow X$.*

- (1) *If γ_0 has no conjugate points, then, for any proper variation γ_t of γ_0 ,*

$$\text{length}(\gamma_t) \geq \text{length}(\gamma_0) \quad \text{for } 0 < |t| < \varepsilon \ll 1;$$

strict inequality whenever the trace of γ_t differs from that of γ_0 .

- (2) *If γ_0 has a conjugate point $0 < s_0 < l$, then there is a proper variation γ_t of γ_0 such that*

$$\text{length}(\gamma_t) < \text{length}(\gamma_0) \quad \text{for } 0 < |t| < \varepsilon \ll 1.$$

Hence, in this case, γ_0 is not minimal, even locally.

Remark 8.24. In full agreement with calculus 102 (cf. §5.6), we do not make any conclusion if γ_0 has a single conjugate point which is $s_0 = l$, the other endpoint. In this case the Hessian form $\mathcal{H}$ is negative *semidefinite*, and we do not have enough data. For example, let γ_0 be a meridian connecting two poles of the unit sphere S^2 . Clearly,

- on S^2 itself, γ_0 is minimal; however, there are arbitrary close to γ_0 curves (other meridians) of the same length π .

Now, we can deform S^2 making sure that sufficiently many derivatives do not change on γ_0 , so that neither do K or conjugate points (again, only on γ_0). For example, locally S^2 is given by

$$y = \sqrt{1 - x^2 - z^2},$$

γ_0 corresponding to $x = 0$, and we can add εx^4 , $|\varepsilon| \ll 1$, to this equation. Then,

- if $\varepsilon > 0$, the new surface is “fatter” about γ_0 and close curves are longer;
- if $\varepsilon < 0$, the new surface is “slimmer” and there are shorter close curves.

Proof of Theorem 8.23(1). Let $p := \gamma_0(0)$ and $\mathbf{v} := \gamma_0'(0)$. By the assumption and Theorem 8.19, the exponential map $\exp_p$ is a local diffeomorphism at each point of the closed ray $t\mathbf{v} \subset T_p X$, $t \in [0, l]$. By the compactness, it restricts to a diffeomorphism from a neighborhood of this ray to a neighborhood of γ_0 , i.e., γ_0 is covered by a normal polar chart. Hence, we can refer to (6.36). $\square$

We ignored the technicalities related to the possible self-intersections of γ_0 : in this case, the words “neighborhood of γ_0 ” should be understood correctly.

Proof of Theorem 8.23(2). In the definition of intrinsic distance in §4.2, we allow *piecewise* smooth curves; hence, it is natural to consider piecewise smooth, or *broken* variations as well. An extra complication is that, since the derivative is no longer a continuous function, all integrals should be understood “piece-by-piece”,

$$\int_0^l := \int_0^{s_0} + \sum_{i=1}^k \int_{s_{i-1}}^{s_i} + \int_{s_k}^l$$

for a curve/variation broken at $0 < s_0 < s_1 < \dots < s_k < l$. This applies to (8.21) as well and, since integration by parts is involved and the fields μ, ν do not need to vanish at the intermediate points, we obtain an extra term:

$$\mathcal{H}(\mu, \nu) = - \sum_{i=1}^k \mu(s_i) \Delta \nu'(s_i) - \int_0^l \mu(\nu'' + K\nu) ds,$$

where

$$\Delta \nu'(s_i) := \nu'(s_i + 0) - \nu'(s_i - 0)$$

is the jump of the derivative at s_i . Recall that, in spite of its appearance, $\mathcal{H}$ is a *symmetric* bilinear form, as it originates from (8.8).

To prove the theorem, we consider the unique Jacobi field

$$J := \tilde{\nu} \mathbf{t}^\perp \quad \text{such that} \quad \tilde{\nu}(0) = \tilde{\nu}(s_0) = 0, \quad \tilde{\nu}'(s_0) = -1$$

and make use of the normal variation vector field $\eta \mathbf{t}^\perp$,

$$\eta := r\mu + \nu$$

broken at one point s_0 , where

- $r \in \mathbb{R}$ is a small parameter to be determined later,
- μ is a smooth function such that $\mu(0) = \mu(l) = 0$, $\mu(s_0) = 1$, and
- ν is the piecewise differentiable function (broken Jacobi field) defined *via*

$$\nu(s) := \begin{cases} \tilde{\nu}(s) & \text{if } 0 \leq s \leq s_0, \\ 0 & \text{if } s_0 \leq s \leq l. \end{cases}$$

We have $\Delta \nu'(s_0) = 1$ and, since $\mathcal{H}$ is bilinear,

$$\mathcal{H}(\eta) = r^2 \mathcal{H}(\mu) + 2r \mathcal{H}(\mu, \nu) + \mathcal{H}(\nu) = r^2 \mathcal{H}(\mu) - 2r.$$

Thus, no matter what $\mathcal{H}(\mu)$ is, $\mathcal{H}(\eta) < 0$ for $0 < r \ll 1$. $\square$

A posteriori, one can refer to the continuity of $\mathcal{H}$ and deform η to a smooth variation. However, using broken Jacobi fields makes the proof cleaner, as it is the Jacobi fields that are sensitive to conjugate points. In fact, in the proof of Morse's theorem (8.22), a maximal negative definite subspace of the space of variations is generated by broken Jacobi fields.

Example 8.25. The proof of Theorem 8.23(2) can be illustrated by a meridian γ_0 of length $l > \pi$ on the unit sphere. We can proceed in two steps.

The first variation ν shifts the arc $[0, \pi]$ of γ_0 to a close meridian, leaving the arc $[\pi, l]$ intact. The new curve has the same length l , but it has a corner at π .

Then, the subsequent smaller variation $r\mu$ merely cuts this corner, producing a strictly shorter curve. This cut could have been chosen to smoothen the curve.

Remark 8.26. In the proof of Bonnet Theorem 8.9, we managed to construct a smooth shorter curve directly. To do so, we used more subtle information about the “nature” of the conjugate point, *viz.* the hypothesis of Corollary 8.17(2).

8.4. Surfaces of constant Gaussian curvature. under construction

Statements to be discussed here:

- closed, $K > 0 \Rightarrow d\mathbb{N}: X \rightarrow S^2$ is a diffeomorphism;
- complete, $K = \text{const} > 0$ (must be closed by Theorem 8.9) $\Rightarrow$ sphere (not true for abstract surfaces: there also is real projective plane, neither true for incomplete: there are pseudo-spheres of revolution);
- complete, *simply connected*, $K \leq 0 \Rightarrow \exp_p: T_p X \rightarrow X$ is a diffeomorphism (obviously not true for non-simply connected, *e.g.*, cylinders);
- complete, *simply connected*, $K \equiv 0 \Rightarrow$ plane;
- complete, *simply connected*, $K = \text{const} < 0 \Rightarrow$ Poincaré half-plane.

Further observations:

- in $\mathbb{R}^3$, any closed X has a point with $K(p) > 0$; if $X \not\cong S^2$, there also is a point with $K(q) < 0$.

Existence of surfaces of constant Gaussian curvature:

- an *abstract* torus S_1 admits a flat metric $K \equiv 0$; clearly, $S_1 = \mathbb{R}^2/\mathbb{Z}^2$;
- an *abstract* X with $\chi(X) < 0$ admits a complete metric with $K \equiv -1$; *e.g.*, this applies to S_g , $g \geq 2$.

For the hyperbolic metrics, we can obtain X as $\mathbb{H}/G$ for an appropriate subgroup $G \subset PSL(2, \mathbb{R})$. This construction explains also the fact that complete metrics of constant curvature $K \equiv -1$ are essentially the same as complex structures.

None of these can be observed in $\mathbb{R}^3$ due to Hilbert's theorem.