DIFFERENTIAL GEOMETRY NOTES[1]

I. Curves

II. Surfaces

[1]Connor Jackman, connor.jackman@cimat.mx . For Math and Science Semester in Guanajuato (MSSG) 2021 at CIMAT.

I. Curves

§1: local normal forms

A *plane curve* in the Euclidean plane, $\mathbb{E}^2$, may be described in the following ways:

- as a *geometric locus*: a 'certain' subset, $\mathcal{C} \subset \mathbb{E}^2$, of the plane
- *implicitly*: as a level set (pre-image), $f^{-1}(a)$, of a map $f : \mathbb{E}^2 \to \mathbb{R}$
- via a *parametrization*: the image, $c(\mathbb{R})$, of a map $c : \mathbb{R} \to \mathbb{E}^2$.

One passes from *synthetic* geometry to *analytic* geometry by taking coordinates on the plane, $\mathbb{E}^2 \cong \mathbb{R}^2$, so that plane curves are described –implicitly or parametrically– by equations. Note that only 'certain' subsets of the plane earn the name plane curves, a more rigorous definition of a plane curve will require additional conditions. For example we will –unless otherwise mentioned– always assume sufficient smoothness or non-degeneracy conditions (eg non-zero differentials) on the function c (resp. f) involved in the curves' parametric (resp. implicit) description. Curves defined by such functions are called *regular curves* and we say c (resp. f) is a regular parametrization (resp. the curve is a regular level set of f).

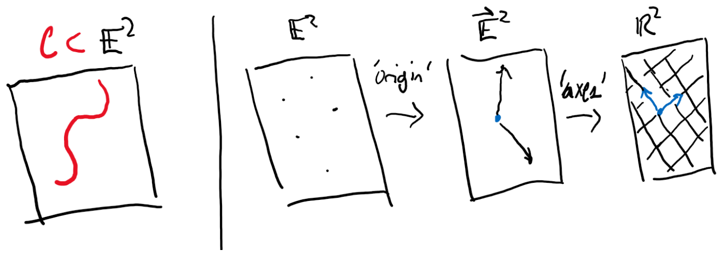

Figure 1. A curve $\mathcal{C}$ in the Euclidean plane. The Euclidean plane may be identified with $\mathbb{R}^2$ by taking coordinates: choosing an origin and (orthonormal) axes.

A fundamental question in the geometry of plane curves is their *equivalence problem*: given two plane curves, when does there exist an isometry of the plane taking one to the other?

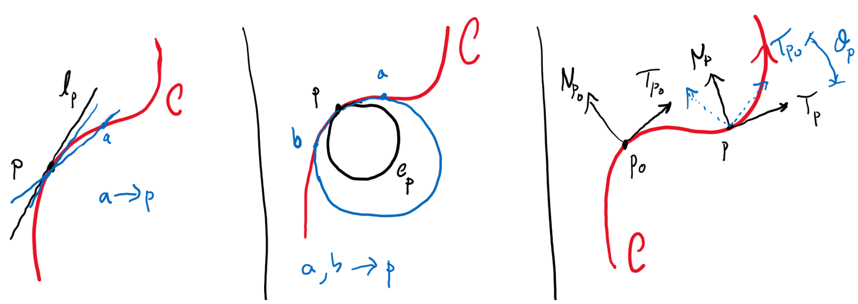

Figure 2. The curvature of a plane curve may be defined 'geometrically' in terms of its osculating circle or 'dynamically' in terms of the changes of a moving frame attached to the curve.

We first consider the 'curvature' of a plane curve. This is a *local* property of the curve: for each point $p \in \mathcal{C}$ on the curve, the curvature at p is a number representing 'how curved' the curve is at p, which only depends on the points of $\mathcal{C}$ 'infinitesimally close' to p. Straight lines have (constant) curvature values of zero, while 'tighter turning curves' values further from zero. Moreover, the curvature is an *invariant* under isometries: if two curves may be taken to eachother by an isometry, their corresponding points will have the same curvature values.

Geometrically, this curvature may be defined as follows. Recall the 'direction of the curve' at $p \in \mathcal{C}$ is given via its best approximation by a line or *tangent line*, ℓ_p, at p: the limit of the lines connecting $a, p \in \mathcal{C}$ as $a \to p$. Likewise, the 'turning of the curve' at p is given via its best approximation by a circle or *osculating circle*, c_p, at p: the limit of the circles[1] passing through $a, b, p \in \mathcal{C}$ as $a, b \to p$. We call the radius, $r(p)$, of c_p the *radius of curvature* at p and its inverse, $k(p) := 1/r(p)$, the (unsigned) *curvature* at p.

Dynamically, one may introduce the curvature of a curve by studying how an 'adapted frame' to the curve changes, i.e. a measure of change in the tangent direction to the curve. Orienting the curve, we have $T, N : \mathcal{C} \to S^1$ (unit circle), sending a point p to a unit vector along the tangent (T_p), normal (N_p) line.

Upon fixing a reference frame, say $\mathcal{F}_o = \{T_{p_o}, N_{p_o}\}$, we have $\mathcal{F}_p = R_p \mathcal{F}_o$ for $\mathcal{C} \to \mathrm{SO}_2, p \mapsto R_p$ a 'curve of rotations'. Then:[2]

$$d\mathcal{F} = dR\mathcal{F}_o = dRR^{-1}\mathcal{F},$$

or writing $R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$, for $\theta : \mathcal{C} \to \mathbb{R}$ we have in components:

$$dT = Nd\theta, \quad dN = -Td\theta.$$

Evaluating these differentials at T, i.e. measuring the changes when moving with unit speed, we arrive at the *planar Frenet-Serret equations*:

$$T' := dT(T) = \kappa N, \quad N' = -\kappa T$$

where $\kappa : \mathcal{C} \to \mathbb{R}, p \mapsto d\theta(T)$ is called the *signed curvature* of $\mathcal{C}$. Note that the sign of κ depends on the choice of direction for the normal and orientation of the curve. Customarily N is chosen as a counterclockwise rotation of T.

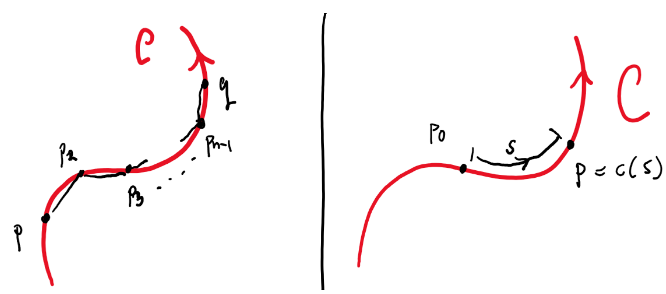

Figure 3. The length of a (oriented) curve between two points is a limit of lengths of its polygonal approximations. A curve is *rectifiable* when the length between any two points on it is finite. Any C^1 curve (admitting parametrizations by continuously differentiable functions) is rectifiable. Rectifiable curves may be parametrized by arc-length.

The *arc-length* of a curve between two points $p, q \in \mathcal{C}$ is another fundamental geometric property. It is defined as a limit of the lengths of inscribed segments: orient the curve segment from p to q and set

$$\ell(\mathcal{C}_{p,q}) := \sup_{p=p_1<p_2<\ldots<p_{n-1}<p_n=q\in\mathcal{C}} \sum_1^{n-1} |p_i p_{i+1}|.$$

Especially convenient for deriving formulas are parametrizations by arc-length: from an orientation of $\mathcal{C}$ and basepoint $p_o \in C$, we let $s \mapsto c(s) \in \mathcal{C}$ be the point of the curve with *signed* arc-length s to p_o.

[1]It is possible these circles limit to or contain lines which we consider as 'circles of infinite radius'.
[2]See the remarks §5, for some explanation of this 'd' (differentials) notation.

Let us present some equations for the computation of the above quantities. Consider a parametric, $c : \mathbb{R} \to \mathbb{E}^2$, or implicit, $f : \mathbb{E}^2 \to \mathbb{R}$, description of a curve. Then:

- the arc-length between $p = c(a)$ and $q = c(b)$, with $a < b$, is given by:

$$\ell(C_{p,q}) = \int_a^b |\dot{c}(t)| \, dt.$$

  In particular to pass from a general parametrization $t \mapsto c(t)$ to an arc-length parametrization, one solves $ds = |\dot{c}(t)| \, dt$. Parametrizations by arc-length, $s \mapsto c(s)$ are characterized by $|c'(s)| = 1$.

- the tangent line at $p = c(t) \in f^{-1}(a)$ is given by:

$$\ell_p = \{c(t) + \lambda \dot{c}(t) : \lambda \in \mathbb{R}\} = \{x : \vec{xp} \cdot \nabla_p f = 0\},$$

- the signed curvature at $p = c(t) \in f^{-1}(a)$ is given by:

$$\kappa(p) = \frac{\det(\dot{c}(t), \ddot{c}(t))}{|\dot{c}(t)|^3} = \pm \frac{d_p^2 f(T,T)}{|\nabla_p f|},$$

- the osculating circle at $p = c(t)$ has center

$$p + \frac{1}{\kappa(t)|\dot{c}(t)|}(-\dot{y}(t), \dot{x}(t)).$$

Now we consider *spatial curves*, certain subsets of Euclidean three space given say by intersecting level sets of functions or as the image of some regular parametrization. As before, we may define an osculating circle at each point of the curve yielding a radius of curvature and unsigned curvature. When this radius is finite, the osculating circle is contained in a plane, called the *osculating plane* at the point.

If the osculating plane is constant, then the 'spatial' curve is contained in this plane – we are just dealing with a plane curve. Change of the osculating plane may be considered as a measure of how much the curve is 'twisting' around in space, i.e. its failure to be contained in a fixed plane.

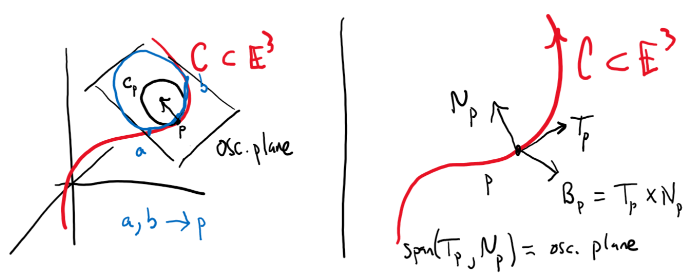

Figure 4. A spatial curve may be studied by the changes in a moving frame attached to the curve, T_p is tangent to the curve, N_p directed towards the center of its osculating circle at p and $B_p = T_p \times N_p$ normal to the osculating plane at p.

We may derive equations to capture the twisting of the curve in space by the same dynamic approach taken with plane curves. Orienting the curve, we have $T, N, B : \mathcal{C} \to S^2$ (unit sphere) where T is along the tangent line, N is in the osculating plane directed towards the center of the osculating circle and $B = T \times N$ is normal to the osculating plane. Upon fixing a reference frame, say $\mathcal{F}_o = \{T_{p_o}, N_{p_o}, B_{p_o}\}$, we have $\mathcal{F}_p = R_p \mathcal{F}_o$ for $\mathcal{C} \to \mathrm{SO}_3, p \mapsto R_p$ a 'curve of rotations'. Then:

$$d\mathcal{F} = dR\mathcal{F}_o = dRR^{-1}\mathcal{F},$$

or letting $dRR^{-1} : T\mathcal{C} \to \mathfrak{so}_3$ be represented by an infinitesimal axis of rotation, $d\omega : T\mathcal{C} \to \mathbb{R}^3$, we have in components: $dT = d\omega \times T,\ \ dN = d\omega \times N,\ \ dB = d\omega \times B$. Evaluating these differentials at T, i.e. measuring the changes when moving with unit speed, we arrive at:

$$T' = \vec{\omega} \times T, \quad N' = \vec{\omega} \times N, \quad B' = \vec{\omega} \times B$$

where $\vec{\omega} : \mathcal{C} \to \mathbb{R}^3, p \mapsto d\omega(T)$ is called the *Darboux vector* along $\mathcal{C}$. By the definition of N, we have $T' = \kappa N$, where $\kappa \geq 0$ is the curvature. It follows that $\vec{\omega} = \kappa B + \tau T$, for some $\tau : \mathcal{C} \to \mathbb{R}$ the *torsion* of the curve. The torsion is signed as the normal B was chosen by 'right hand rule' from T and N. Substituting this expression for $\vec{\omega}$ yields the *spatial Frenet-Serret equations:*

$$T' = \kappa N, \quad N' = -\kappa T + \tau B, \quad B' = -\tau N,$$

so that the torsion measures changes in the osculating plane (B is normal to the osculating plane).

We may establish the following equations for computing the curvature and torsion of spatial curves. Given a (regular) parametrization $c : \mathbb{R} \to \mathbb{E}^3$:

- the curvature at $p = c(a)$ is given by:

$$\kappa(p) = \frac{|\dot{c}(a) \times \ddot{c}(a)|}{|\dot{c}(a)|^3},$$

- the torsion at $p = c(a)$ is given by:

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

By Taylor expansion, we arrive at the following *local normal forms* for planar or spatial curves around a point, in terms of the curvature and torsion of the curve. Consider a curve parametrized by arc-length, $c(s)$, with $c(0) = p$ and let $\kappa_o = \kappa(c(0)), \kappa_o' = \frac{d}{ds}|_0 \kappa(c(s)), \tau_o = \tau(c(0))$. Then there are Cartesian coordinates centered at p with:

$$c(s) = (s, \frac{s^2}{2}\kappa_o) + o(s^2) \qquad \text{(planar curve)}$$

$$c(s) = (s - \frac{s^3}{6}\kappa_o^2, \frac{s^2}{2}\kappa_o + \frac{s^3}{6}\kappa_o', \frac{s^3}{6}\kappa_o\tau_o) + o(s^3) \qquad \text{(spatial curve).}$$

Local refers to these expansions depending only on the values of κ, τ near the point p. In fact, if one continued computing more terms of the Taylor expansion all terms would depend only on the curvature, torsion and their derivatives at p (this is a consequence of the Frenet-Serret equations).

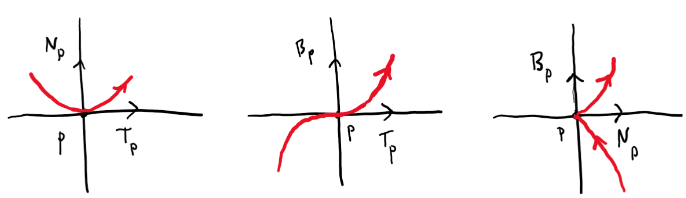

Figure 5. The curvature and torsion at the point p determine the local (near p) form of the curve. Here for a spatial curve with $\kappa(p) \neq 0, \tau(p) > 0$ we draw its projections onto coordinate planes determined by its Frenet-frame at p.

EXERCISES:

1. Show c_p is the limit of circles tangent to ℓ_p at p and passing through $a, p \in \mathcal{C}$ as $a \to p$.
2. Show the radius, r, of a circle circumscribed around a triangle Δ with side lengths a, b, c is $r = \frac{abc}{4 \cdot Area(\Delta)}$.
3. Determine the curvature function of a parabola $y = x^2$ (using the 'geometric' definition of curvature, and say exercises # 1 or # 2).
4. Show $|\kappa| = k$ (justifying the terminology of signed/unsigned curvature).
5. Establish the formula[1] $\kappa_p = \pm \frac{d_p^2 f(T,T)}{|\nabla_p f|}$ for an implicitly defined plane curve $p \in f^{-1}(cst.)$, where the sign depends on whether the normal (a counterclockwise rotation of T) is chosen as proportional to ∇f or $-\nabla f$.
6. For a parametrization $t \mapsto c(t)$ of a plane curve, show the center of the osculating circle at $c(t)$ is given by: $c(t) + \frac{1}{\kappa(t)|\dot{c}(t)|}(-\dot{y}(t), \dot{x}(t))$.
7. For a spatial curve, with regular parametrization $t \mapsto c(t)$, show:
   (a) the curvature at $p = c(t)$ is given by: $\kappa(p) = \frac{|\dot{c}(t) \times \ddot{c}(t)|}{|\dot{c}(t)|^3}$,
   (b) the torsion at $p = c(t)$ is given by: $\tau(p) = \frac{\det(\dot{c}(t), \ddot{c}(t), \dddot{c}(t))}{|\dot{c}(t) \times \ddot{c}(t)|^2}$.
8. Show the Frenet-Serret equations, $\frac{d\mathcal{F}}{ds} = \left(\frac{dR}{ds} R^{-1}\right) \mathcal{F}$, do not depend on the choice of initial reference frame, $\mathcal{F}_o$ (with $\mathcal{F}(s) = R(s)\mathcal{F}_o$).

[1] The *Hessian*, $d_p^2 f$, of f is the quadratic form defined by $d_p^2 f(v, v) := \frac{d^2}{dt^2}|_{t=0} f(p + tv) = \frac{d}{dt}|_{t=0} \nabla_{p+tv} f \cdot v$.

§2 some global results

We consider some classical global results on curves. First, an 'integrated' version of the local normal form result above:

Equivalence theorem for curves: Two curves $\mathcal{C}_1, \mathcal{C}_2$ may be taken to one another by an isometry iff they have the same curvature and torsion functions, ie there exists arclength parametrizations: $s \mapsto c_j(s) \in \mathcal{C}_j$ with $\kappa_1(s) = \kappa_2(s), \tau_1(s) = \tau_2(s)$.

proof: Let $\mathcal{F}_j$ be the Frenet-Serret frames along the curves, with $\mathcal{F}_j' = R_j' R_j^{-1} \mathcal{F}_j = \Omega_j \mathcal{F}_j$ the Frenet-Serret equations with say reference configuration $\mathcal{F}_1(0)$. Let $s \mapsto R_s$ be the curve of rotations defined by $\mathcal{F}_2(s) = R_s \mathcal{F}_1(s)$. We will show that R_s is constant. The assumption that the curvatures and torsions are equal implies that the matrix representations of Ω_j in the $\mathcal{F}_j$ bases are the same, ie:

$$R\Omega_1 = \Omega_2 R.$$

On the other hand, $R_2 = RR_1$, with $R_j' = \Omega_j R_j$ so that

$$\Omega_2 R_2 = R_2' = R'R_1 + RR_1' = R'R_1 + R\Omega_1 R_1 = R'R_1 + \Omega_2 R_2 \Rightarrow R' = 0.$$

Hence $R_s \equiv R_o$ is a constant rotation. In particular integrating $c_2'(s) = R_o c_1'(s)$ yields $c_2(s) = R_o c_1(s) + a_o$ for a constant vector $a_o \in \mathbb{R}^3$, so that the isometry $p \mapsto R_o p + a_o$ takes $\mathcal{C}_1$ to $\mathcal{C}_2$. $\square$

Now we consider planar curves, starting by relating curvature to 'turning number'. Let $s \mapsto c(s)$ be an arc-length parametrized planar curve and write $s \mapsto c'(s) \in S^1$ as $c'(s) = (\cos\theta(s), \sin\theta(s))$. Then:

$$\kappa(s) = \theta'(s).$$

So that $\int_{s_o}^{s_1} \kappa(s)\, ds = \theta(s_1) - \theta(s_o)$, called the *turning angle* of the curve from $c(s_o)$ to $c(s_1)$.

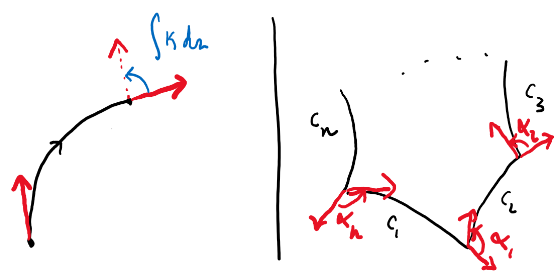

Figure 6. The turning angle over a segment of the curve may be expressed as a 'total curvature' over the segment. For piecewise smooth curves, one has *outer angles* α_j at the vertices. Note that $\theta_{j+1}(a_{j+1}) - \theta_j(b_j) \equiv_{2\pi} \alpha_j$, or $\theta_j(b_j) - \theta_{j+1}(a_{j+1}) + \alpha_j \equiv_{2\pi} 0$.

A curve is *closed* if it admits a parametrization $c : [0,T] \twoheadrightarrow \mathcal{C} = im(c)$ with $\frac{d^k c}{dt^k}(0) = \frac{d^k c}{dt^k}(T),\ k = 0, 1$[1]. For any smooth closed curve:

$$\oint_{\mathcal{C}} \kappa\, ds = \int_0^T \kappa(t)|\dot{c}(t)|\, dt = 2\pi n_{\mathcal{C}},$$

where $n_{\mathcal{C}} \in \mathbb{Z}$ is the *turning number* of the curve. Likewise, for *piecewise* closed curves: $\mathcal{C} = \cup im(c_j)$ with $c_j : [a_j, b_j] \to \mathbb{R}^2, j = 1, ..., n$ having $c_j(b_j) = c_{j+1}(a_{j+1})$ (indeces mod n), we have:

$$\oint_{\mathcal{C}} \kappa\, ds + \sum_{j=1}^{n} \alpha_j = \sum_{j=1}^{n} \left(\int_{c_j} \kappa_j\, ds + \alpha_j \right) = 2\pi n_{\mathcal{C}}$$

[1] We assume this condition does not hold for any $T' \in (0,T)$.

where $\alpha_j = \angle(\dot{c}_j(b_j), \dot{c}_{j+1}(a_{j+1})) \in [0,\pi]$ are the outer angles and $n_C \in \mathbb{Z}$ is the turning number.[1] Next, a closed curve is *simple* if it admits a parametrization $c : [0,T] \twoheadrightarrow C$ that is injective on $[0,T)$. Then:

Turning theorem: The turning number of a simple closed curve is ± 1.

proof: We consider the case of a smooth simple closed curve (the proof may be modified to treat piecewise simple closed curves). Choose an arc-length parametrization, $s \mapsto c(s), \quad s \in [0,\ell]$, with the curve contained entirely on one side of its tangent line at $c(0)$. Since the curve is simple, $f : T \to S^1, (s_1,s_2) \mapsto \begin{cases} \frac{c(s_2)-c(s_1)}{|c(s_2)-c(s_1)|} & s_2 \neq s_1, (s_1,s_2) \neq (0,\ell) \\ c'(s) & s_1 = s_2 = s \\ -c'(0) & s_1 = 0, s_2 = \ell \end{cases}$ for $T = \{0 \le s_1 \le s_2 \le \ell\} \subset \mathbb{R}^2$ is continuous, and may be written as $f(s_1,s_2) = (\cos\theta(s_1,s_2), \sin\theta(s_1,s_2))$ for some continuous $\theta : T \to \mathbb{R}$ (and any two such θ's differ by a constant integer multiple of 2π, so that differences are well defined). Now

$$2\pi n_c = \int_c \kappa \; ds = \theta(\ell,\ell) - \theta(0,0) = \theta(\ell,\ell) - \theta(0,\ell) + \theta(0,\ell) - \theta(0,0)$$

and we will show that the two terms $\theta(\ell,\ell)-\theta(0,\ell), \;\; \theta(0,\ell)-\theta(0,0)$ take the same value $\pm\pi$, so that $n_c = \pm 1$. Consider $\theta(s,\ell) - \theta(0,\ell)$ for $s \in (0,\ell)$. It is the angle between $f(s,\ell) = \frac{c(\ell)-c(s)}{|c(\ell)-c(s)|}$ and $f(0,\ell) = -c'(0)$. Observe that because c is contained in one of the half planes divided by $c'(0)$, we have $f(s,\ell)$ is contained in a semi-circle, divided by $f(0,\ell)$. Hence $|\theta(s,\ell) - \theta(0,\ell)| \le \pi$ and so $\theta(\ell,\ell) - \theta(0,\ell) = \angle(c'(0), -c'(0)) = \pm\pi$. Likewise $\theta(0,\ell) - \theta(0,0) = \lim_{s\to\ell} \theta(0,s) - \theta(0,0) = \pm\pi$. $\square$

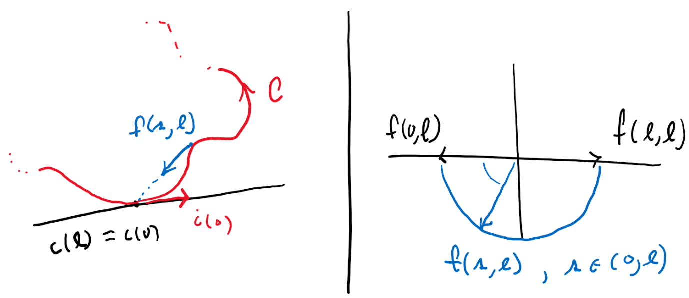

Figure 7. Proof of turning number theorem: the turning number of a simple closed curve is ± 1.

Simple curves also have:

4-vertex theorem: The curvature function of a smooth simple closed curve has at least 4 critical points.

proof: Write $c'(s) = (\cos\theta(s), \sin\theta(s))$, and think of $\rho(s) = \frac{1}{\kappa(s)} = \frac{ds}{d\theta}$ as a mass density over the circle. When the circle has this mass distribution, its center of mass: $\int \rho c' \; d\theta = \int c'(s) \; ds = 0$ is at the origin. Suppose that ρ has only 2-critical points, a maximum and minimum (so κ as well has only 2 critical points). Taking axes through the origin with one parallel to the line joining these maximim and minimum points, we see that the circle cannot be balanced, contradiction. $\square$

The *vertices* of a curve are points where the curvature has a critical point. They are related to optical properties of light emitted from the curve as well as balancing properties of the curve. We have as well:

[1] The equality between the two sides of these formulas is a rather remarkable relation illustrating a common 'local to global' theme: on the one hand we have a 'total' of locally determined quantities while on the other a quantity depending on the global form of the curve. Moreover, the global side is invariant under differentiable deformations of the curve, which in general change individual values of the local quantities (but not their total!).

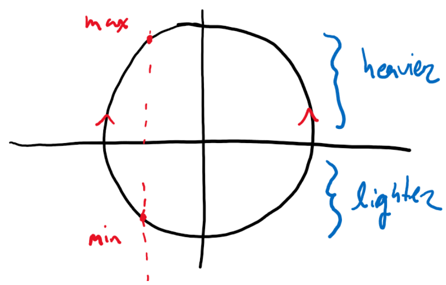

Figure 8. A 'mechanical' proof (due to W. Blaschke) of the 4-vertex theorem (for strictly convex, $\kappa > 0$, curves).

Tait-Kneser theorem: The osculating circles on any vertex free segment of a curve are nested.

proof: Consider a vertex free segment from $c(s_o)$ to $c(s_1)$. The centers of the osculating circles are parametrized by $e(s) = c(s) + \rho(s)N(s)$, where $\rho(s) = 1/\kappa(s)$ is the (signed) radius of the osculating circle at $c(s)$. Then $\ell(e_{s_o,s_1}) = \int_{s_o}^{s_1} |e'(s)|\ ds = |\rho(s_1) - \rho(s_o)| \geq |e(s_1) - e(s_o)|$. $\square$

Another famous classical result on plane curves is:

Isoperimetric inequality: The area, A, enclosed by a simple plane curve of length ℓ satisfies:

$$4\pi A \leq \ell^2$$

with equality iff the curve is a circle (of radius $\ell/2\pi$).

As for spatial curves there are:

Fenchel's theorem: The *total curvature*, $\oint_{\mathcal{C}} \kappa\ ds$, of a closed spatial curve is bounded below by 2π with equality occuring only when $\mathcal{C}$ is simple, planar and convex.

A *convex* plane curve being one which always lies on one side of its tangent line at every point.

Fairy-Milnor theorem: If a simple closed spatial curve is 'knotted' then $\oint_{\mathcal{C}} \kappa\ ds \geq 4\pi$.

the proofs of which we refer to the exercises and references.

Exercises:

1. If two oriented curves may be taken to eachother by an (orientation preserving) isometry, show they have the same curvature and torsion values.

2. (a) For $a, b \geq 0$ show that $\sqrt{ab} \leq \frac{a+b}{2}$ with equality only when $a = b$.

   (b) For $\vec{a}, \vec{b} : [t_o, t_1] \to \mathbb{R}^3$, show that $\int_{t_o}^{t_1} \vec{a}(t) \cdot \vec{b}(t)\ dt \leq \int_{t_o}^{t_1} |\vec{a}(t)||\vec{b}(t)|\ dt$ with equality iff $\vec{a} = \lambda \vec{b}$, some $\lambda : [t_o, t_1] \to \mathbb{R}_+$.

   (c) Let C be closed simple planar curve of length ℓ with arc-length parametrization $c(s) = (x(s), y(s)), s \in [0, \ell]$. Sandwich the curve between two lines parallel to a y-axis, and let $\bar{c}(s) = (x(s), \pm\sqrt{r^2 - x(s)^2})$ parametrize the inscribed circle between the the two lines ($2r$ is the distance between the two lines). For $A = \oint_C x dy$ the area enclosed by C, show that $A + \pi r^2 \leq r\ell$ where $\pi r^2 = \oint_{\bar{c}} -y dx$. Deduce the isoperimetric inequality:

$$4\pi A \leq \ell^2$$

3. With the set-up of # 2, show that there is equality in the isoperimetric inequality iff C is a circle of radius $\ell/2\pi$.

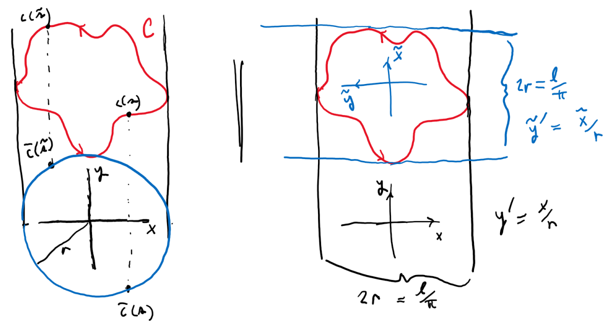

Figure 9. The isoperimetric inequality (left for # 2, right for # 3).

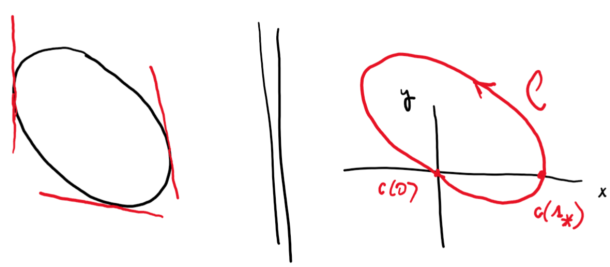

Figure 10. A convex curve and a choice of coordinates which may be used to prove a 4-vertex theorem for convex simple curves.

4. A simple plane curve is convex when at each point it is contained on one side of its tangent line at that point. Show a simple closed curve is convex iff κ has fixed sign (ie depending on the curves orientation we have always $\kappa \geq 0$ or $\kappa \leq 0$).

5. Give an example of a closed curve with $\kappa \geq 0$ at all points which is not convex.

6. Consider a convex simple plane curve and suppose by way of contradiction that it has *only* two vertices (a maximum and minimum of κ). Let $s \mapsto c(s)$ be an arc-length parametrization of the curve, with the two vertices at $s = 0, s = s_* \in (0, \ell)$.

(a) Show there is an appropriate system of coordinates, $c(s) = (x(s), y(s))$ with $y(s) = 0$ only for $s = 0, s = s_*$.

(b) Show that $0 \neq \int_0^\ell \kappa'(s) y(s)\ ds$.

(c) Integrating by parts, show that $\int_0^\ell \kappa'(s) y(s)\ ds = \int_0^\ell x''(s)\ ds = 0$, a contradiction of part b.

7. The *index* or *winding number* (with respect to the origin) of a closed curve $\gamma(t) \in \mathbb{R}^2 \backslash \{(0,0)\}$ is $\frac{1}{2\pi} \oint_\gamma\ d\theta$. Identifying the plane with the complex numbers, show this index is given by $\frac{1}{2\pi i} \oint_\gamma \frac{dz}{z}$.

8. For a closed curve parametrized by arc-length, $s \mapsto c(s), s \in [0, \ell]$, let $s \mapsto \gamma(s) = c'(s) \in S^2$ be its *tangent indicatrix*. Show that for any $\vec{v} \in \mathbb{R}^3$ that $\oint_\gamma \vec{v} \cdot \gamma(s)\ ds = 0$. Conclude that γ intersects every great circle on the sphere in at least two points.

9. Show *Crofton's formula*:[1] for a closed curve with tangent indicatrix $\gamma \subset S^2$, one has:

$$\ell(\gamma) = \frac{1}{4} \int_{\xi \in S^2} \#(\gamma \cap \xi^\perp) dA_\xi.$$

10. There is a planar formula[2] similar to the previous exercise (also called Crofton's formula). Let γ be a planar curve. Coordinatize the lines in the plane by $(p, \varphi) \in \mathbb{R}_+ \times S^1$ corresponding to the line $\ell_{p,\varphi} \subset \mathbb{R}^2$ at distance p from the origin whose normal through the origin makes angle φ. Then:

$$\ell(\gamma) = \frac{1}{2} \int_{\mathbb{R}_+ \times S^1} \#(\gamma \cap \ell_{p,\varphi})\ dp d\varphi.$$

[1] This formula may be used to prove Fenschel's theorem and the Fairy-Milnor theorem. See eg Shifrin's notes, in particular ex. 12 pg. 35.

[2] See for example doCarmo, pgs. 41-46.

§3 calculus of variations

The most fundamental curves in Euclidean geometry –lines– are often described by the following *variational* characterization: for any two points p, q on a line, the arc-length of the line segment is minimal among curves from p to q.

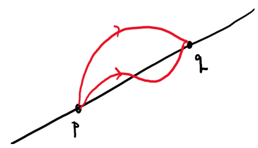

Figure 11. For two points on a line, the segment of the line between the two points minimizes arc-length of curves between the points.

We may prove that lines are in fact the unique curves in Euclidean space satisfying the above variational property as an application of the *calculus of variations*. The general sort of problem considered by such variational methods (as applied to curves) may be phrased as follows. One considers a *functional*:

$$A : \Gamma \to \mathbb{R}$$

where Γ is a certain 'class' of curves, and the functional A assigns a number to each curve in the class. The goal is to describe 'critical points' or *extremal curves* of the functional. For example a curve, $\gamma_* \in \Gamma$, which minimizes A, that is: $A(\gamma_*) \leq A(\gamma), \quad \forall \gamma \in \Gamma$.

The specific definition of Γ and A is determined based on the problem under consideration. For example, the variational property characterizing a line (segment) may be set-up by taking

$$\Gamma := \{\gamma : [0,1] \to \mathbb{E}^n : \gamma(0) = p, \gamma(1) = q \text{ and } \gamma \text{ smooth}\},$$

$$A(\gamma) := \int_0^1 |\dot{\gamma}| \, dt.$$

Our claim about lines being characterized by their variational property amounts to the assertion that the functional A takes its minimum value only when $\gamma_* \in \Gamma$ parametrizes a line segment from p to q.

One may progress on variational problems provided the class of curves and functional satisfy certain 'smoothness' properties. Namely, let us call a *variation* of a parametrized curve, $[a,b] \ni t \mapsto \gamma(t) \in \mathbb{R}^n$, a family of curves $[a,b] \ni t \mapsto \gamma_\varepsilon(t)$ with $\varepsilon \in (-\delta, \delta)$ satisfying $\gamma_0(t) = \gamma(t)$. The variation is smooth when the application $(-\delta, \delta) \times [a,b] \to \mathbb{R}^n, \quad (\varepsilon, t) \mapsto \gamma_\varepsilon(t)$ is smooth (continuous partial derivatives). All the types of variational problems we will consider satisfy: for any smooth variation $\gamma_\varepsilon \in \Gamma$ of $\gamma \in \Gamma$, the function $\mathbb{R} \ni \varepsilon \mapsto A(\gamma_\varepsilon) \in \mathbb{R}$ is differentiable. Moreover, the functionals A we consider will all admit an expression in the following 'integral' form: $A(\gamma) = \int_a^b L(\gamma(t), \dot{\gamma}(t), t) \, dt$ where $L : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ is some (differentiable) function (a *Lagrangian* of the variational problem).

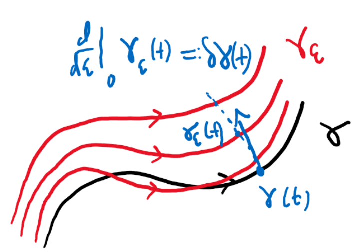

Figure 12. A variation, γ_ε, of a curve γ.

Now, an *extremal* curve of such a 'smooth' variational problem is defined as a curve $\gamma_* \in \Gamma$ such that for any (smooth) variation $\gamma_\varepsilon \in \Gamma$ of γ_* it holds that:

$$\frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} A(\gamma_\varepsilon) = 0.$$

When A is given by a Lagrangian, then for a variation, $[a,b] \ni t \mapsto \gamma_\varepsilon(t)$, of $\gamma_* : [a,b] \to \mathbb{R}^n$, we compute:

$$\frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} A(\gamma_\varepsilon) = \int_a^b \left(\partial_\gamma L - \frac{d}{dt}\partial_{\dot\gamma} L\right) \cdot \delta\gamma \; dt + \partial_{\dot\gamma} L \cdot \delta\gamma|_a^b$$

where $\delta\gamma(t) = \frac{d}{d\varepsilon}|_{\varepsilon=0}\gamma_\varepsilon(t)$ is a vector field along γ_*. In our examples the class of curves will typically consist of: *fixed time* and *fixed endpoints*, meaning all curves in Γ are parametrized on a fixed interval and begin and end at the same endpoints. For such variational problems, we have:

Euler-Lagrange equations: Consider a variational problem given by a Lagrangian defined over smooth curves with fixed time and fixed endpoints. Then γ_* is an extremal iff over γ_* it holds that:

$$\partial_\gamma L = \frac{d}{dt}\left(\partial_{\dot\gamma} L\right).$$

proof: Γ consists of smooth curves parametrized on a fixed interval $[a,b]$ with $\gamma(a), \gamma(b)$ fixed. By our computation above, we have that γ_* is an extremal iff $0 = \int_a^b (\partial_\gamma L - \frac{d}{dt}\partial_{\dot\gamma} L) \cdot \delta\gamma \; dt$ for *any* vector field $\delta\gamma$ along γ_* which vanishes at the endpoints. Using a bump function, it can be shown that this condition only holds when $\partial_\gamma L - \frac{d}{dt}\partial_{\dot\gamma} L = 0$ over γ_*. $\square$

It is important not to forget that extremals (eg curves satisfying Euler-Lagrange equations) do *not* need to be minima of the functional, merely are analogous to critical points. If one seeks minima of the functional, they must be found among the extremals, but further work is needed to assert that the extremals one has found are indeed minima.

Example:

- Let Γ consist of all smooth curves $\gamma : [0,1] \to \mathbb{R}^n$ with $\gamma(0) = p, \gamma(1) = q$ fixed and $L = |\dot\gamma|$. Then $\partial_\gamma L = 0, \partial_{\dot\gamma} L = \frac{\dot\gamma}{|\dot\gamma|}$. The condition to be an extremal is, by the Euler-Lagrange equations, that $\frac{\dot\gamma}{|\dot\gamma|} = cst$. In this case, $A(\gamma)$ does not depend on the parametrization of γ, so we may assume wlog that the extremals are parametrized by constant speed, and hence lines, $\gamma(t) = tq + (1-t)p$.
- There is a convenient 'trick' (getting rid of square roots in the Lagrangian) to convert the variational characterization of lines as length minimizers to 'energy minimizers' . Namely, let $A(\gamma) = \int |\dot\gamma| \; dt$ and $E(\gamma) = \int |\dot\gamma|^2 \; dt$ considered over the class of all smooth curves $\gamma(a) = p, \gamma(b) = q$ with fixed endpoints and time. By Cauchy-Schwarz, $A(\gamma) \leq (b-a)E(\gamma)$ with equality iff $|\dot\gamma| = cst$. From this inequality, it follows that:

  $$\gamma_* \text{ minimizes } E \iff |\dot\gamma_*| = cst. \text{ and } \gamma_* \text{ minimizes } A.$$

  In particular to find the minimizers of length (minimizers of A), it suffices to determine the minimizers of E. The Euler-Lagrange equations for the extremals of E are (with Lagrangian $L = |\dot\gamma|^2$), $0 = \partial_\gamma L = \frac{d}{dt}\partial_{\dot\gamma} L = \ddot\gamma$, so extremals of E are straight lines (with constant velocity).

As for determining whether given extremals are minimizers, one somewhat general method is the following: suppose that $q_o \in \mathbb{R}^n$ is fixed and for each $q \in \mathbb{R}^n$ one has determined a 'candidate' curve, γ_q, for a minimizer going from q_o to q (these candidate curves would in practice by determined by solving the Euler-Lagrange equations to determine an extremal curve connecting q_o to q). Define the function $S : \mathbb{R}^n \to \mathbb{R}$ by $S(q) := A(\gamma_q)$. Now, if $S(q_o) = 0$ and it holds that $L \geq \Phi$, where $\Phi(\gamma, \dot\gamma) = d_\gamma S(\dot\gamma)$ then the candidate curves are in fact minimizers. The proof amounts to that for any curve, γ, joining q_o to q, we have $A(\gamma) = \int_\gamma L\ dt \geq \int_\gamma \Phi\ dt = S(q) - S(p) = S(q) = A(\gamma_q)$.

This method may be applied to check that line segments minimize lengths (as we would hope!). Indeed, let say q_o be the origin and our candidate extremal from q_o to q be the curve $tq,\ t \in [0,1]$. Then $S(q) = |q|$, and $dS(\dot\gamma) = \frac{\gamma\cdot\dot\gamma}{|\gamma|}$. The Cauchy-Schwarz inequality, $|\gamma \cdot \dot\gamma| \leq |\gamma||\dot\gamma|$ implies that indeed $dS(\dot\gamma) \leq |\dot\gamma| = L(\dot\gamma)$, so that line segments realize distances (shortest lengths) between two fixed points.

Many interesting curves arise from variational descriptions. For example, some 'classics' are:

- The *brachistochrone* curve: consider two fixed points in the plane (with a uniform gravitational force field). Let the value of the functional on a smooth curve connecting p to q be the time it takes a bead of fixed mass to fall along the curve (think of say a wire) from p to q when subject to this uniform gravitation. The brachistochrone curve is the curve from p to q which minimizes this time of descent. It is given by an arc of a cycloid passing through p and q.

- The *catenary* curve: consider two fixed points, p, q, in the plane (with a uniform gravitational field) and a 'chain' of fixed length $\ell > |pq|$ of homogeneous mass. The chain will be at rest when its potential energy has a critical point. Thus the shape of the 'hanging chain' or catenary curve is determined by extremizing the functional sending a curve of length ℓ connecting p to q to its gravitational potential energy. In appropriate coordinates, it is given by a section of the graph $y = \frac{1}{c}\cosh cx$.

- an *elastica* curve: consider two fixed points, p, q, with two fixed directions, u, v, at these points and curves (elastic 'wires') of fixed length $\ell > |pq|$ from p to q with initial and final tangents along u, v. The bending energy of the wire bent into the shape of a given curve is $\int \kappa^2\ ds$, and equilibrium configurations of the wire are extremals of this functional. They are characterized by the condition that the curvature at a point on the curve is proportional to the distance of this point to some fixed line in the plane.

Some other useful properties of variational problems are:

Energy constant: Suppose a fixed time and fixed endpoint variational problem is given by a time independent Lagrangian, $L(\gamma, \dot\gamma)$. Then $E := \dot\gamma \cdot \partial_{\dot\gamma} L - L$ is constant along extremals.

Lagrange multipliers v1: Let Γ be smooth fixed time and endpoint curves and $\Gamma_o \subset \Gamma$ a 'subclass' of curves defined by a condition of the form $B = cst.$ for $B : \Gamma \to \mathbb{R}$. Consider a functional $A : \Gamma \to \mathbb{R}$.
If $\gamma_* \in \Gamma_o$ is an extremal of $A + \lambda B : \Gamma \to \mathbb{R}$ for some $\lambda \in \mathbb{R}$ then $\gamma_* \in \Gamma_o$ is an extremal of $A|_{\Gamma_o} : \Gamma_o \to \mathbb{R}$.

Lagrange multipliers v2: Let Γ be smooth fixed time and endpoint curves and $\Gamma_o \subset \Gamma$ defined by a condition of the form $b(\gamma(t), \dot\gamma(t), t) = cst.$ where $b : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$. Consider a functional $A : \Gamma \to \mathbb{R}$.
If $\gamma_* \in \Gamma_o$ is an extremal of $\Gamma \ni \gamma \mapsto A(\gamma) + \int_{dom(\gamma)} \lambda(t) b(\gamma, \dot\gamma, t)\ dt$ for some $\lambda : \mathbb{R} \to \mathbb{R}$ then $\gamma_* \in \Gamma_o$ is an extremal of $A|_{\Gamma_o} : \Gamma_o \to \mathbb{R}$.

§4 examples

We present some 'classic' curves obtained through various constructions or physical properties.

String constructions: Given a plane curve, $\mathcal{C}$, one may associate to it a family of curves called *involutes* of $\mathcal{C}$:

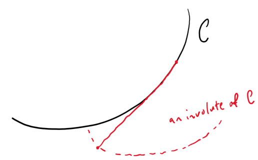

Figure 13. To construct an involute, one fixes an end of a tightly stretched 'string' of fixed length to a point of $\mathcal{C}$ and traces the other end of the string as it is 'wrapped along' $\mathcal{C}$. Involutes arise in the study of certain pendulums.

When $c(s)$ is an arc-length parametrization of $\mathcal{C}$, an involute determined by a string of length ℓ with an end fixed to $c(0)$ is then parametrized by:

$$\gamma(s) = c(s) + (\ell - s)c'(s).$$

The circle may be thought of as a degenerating case of an involute when an end of a string is fixed to a point in the plane. The *conic sections* also may be characterized (defined) by string constructions.

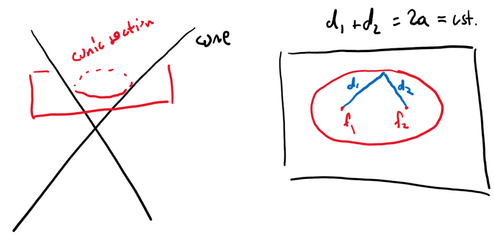

Figure 14. Conic sections are planar curves obtained by intersecting a cone with a plane. They come in three types: ellipses, parabolas, hyperbolas, depending on the relative angle of the plane with the cone. By the 'method of Dandelin spheres' one may show that conic sections may equivalently be defined by certain 'string constructions' (see for example this video), eg an ellipse is the set of points whose sum of distances to two fixed points (foci) is constant. One observes conic sections for instance in the profile of a circle (eg the rim of ones mug) seen from an angle, or in the boundary curve cast by a flashlight on the ground.

Conic sections may be used to define a system of coordinates, *elliptic coordinates*, useful in various physical problems and whose generalization to 3-dimensions are useful in the study of quadratic surfaces.

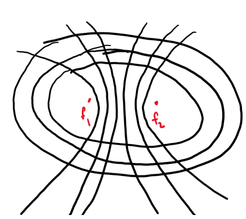

Figure 15. Elliptic coordinates on the plane are determined by fixing two points, f_1, f_2, in the plane and considering the set of *confocal conics*: having foci at f_1, f_2. Such conics may be parametrized by a single number, ρ, as $\frac{x^2}{a^2-\rho} + \frac{y^2}{b^2-\rho} = 1$, with $a > b > 0$ fixed and x, y appropriate Cartesian coordinates. To the parameters $(\mu, \nu) \in (-\infty, b^2) \times (b^2, a^2)$ one assigns the points in the plane given by intersection of the conics $\rho = \mu$ and $\rho = \nu$.

Explicitely, these elliptic coordinates, (μ,ν), coordinatize quadrants of the plane, related to Cartesian coordinates through:

$$x^2=\frac{(a^2-\mu)(a^2-\nu)}{a^2-b^2},\quad y^2=\frac{(b^2-\mu)(\nu-b^2)}{a^2-b^2}.$$

They have the important property of being an *orthogonal coordinate system*, meaning $(x_\mu,y_\mu)\cdot(x_\nu,y_\nu)=0$, or geometrically that confocal ellipses and hyperbolas intersect orthogonally.

Rod constructions: When a rod of fixed length is 'dragged' along a given curve (free to pivot about its point of attachment to the curve) its end traces out another curve. Such curves are called *bicycling curves* [1] since the point of attachment along the given curve may be thought of as the front wheel of a bicycle, and the dragged and as the track traced by the bicycle's rear wheel.

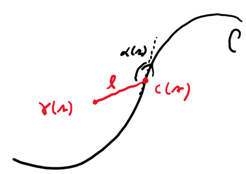

Figure 16. As a rod of fixed length ℓ is 'dragged' along a fixed curve $\mathcal{C}$, its other end traces out another curve. One may think of driving the front wheel of a bicycle of fixed frame length ℓ along $\mathcal{C}$, and the resulting curve traced by the rear wheel.

Let $c(s)$ be an arclength parametrization of the given 'front wheel' curve $\mathcal{C}$, for a 'rod' or bicycle of fixed frame length ℓ. Then the bicycling curves 'rear wheel tracks' are parametrized as:

$$\gamma(s)=c(s)+\ell(\cos(\theta(s)+\alpha(s)),\sin(\theta(s)+\alpha(s)))$$

where $c'(s)=(\cos\theta(s),\sin\theta(s))$ and the angle, $\alpha(s)$, between the rod and the tangent to $\mathcal{C}$ solves:

$$\alpha'=\frac{1}{\ell}\sin\alpha-\kappa$$

for κ the curvature of $\mathcal{C}$.

A *tractrix* is the resulting 'rear track curve' when $\mathcal{C}$ is a straight line. Parametrizing the line as $c(s)=(s,0)$ we parametrize a tractrix by $\gamma(s)=(s,0)+\ell(\cos\alpha,\sin\alpha)$, with $\alpha'(s)=\frac{1}{\ell}\sin\alpha(s)$. This ode has solutions: $\int\frac{d\alpha}{\sin\alpha}=\frac{s}{\ell}\Rightarrow e^{s/\ell}|\tan\frac{\alpha_o}{2}|=|\tan\frac{\alpha}{2}|$. Eliminating s we find the relation:

$$x=\ell\log\frac{\ell+\sqrt{\ell^2-y^2}}{y}-\sqrt{\ell^2-y^2}$$

giving the tractrix explicitely as a graph $(x(y),y)$ for $y\in(0,\ell)$. Moreover, using $e^{s/\ell}=\cosh\frac{s}{\ell}+\sinh\frac{s}{\ell}$, we obtain the explicit parametrization:

$$x(s)=s-\ell\tanh\frac{s}{\ell},\quad y(s)=\frac{\ell}{\cosh\frac{s}{\ell}}.$$

Bicycling curves are special cases of *pursuit curves*: a point moves along a given curve $\mathcal{C}$, as given by the parametrization $t\mapsto c(t)$ while a 'pursuing point' moves at each time in a direction towards $c(t)$ with speed $v(t)$. The trajectory of the pursuing point is determined by its initial position $\gamma(0)$ and satisfying the ode:

$$\dot\gamma(t)=v(t)\frac{c(t)-\gamma(t)}{|c(t)-\gamma(t)|}.$$

[1] see for example: G. Bor, M. Levi, R. Perline, S. Tabachnikov, *Tire tracks and integrable curve evolution.* International Mathematics Research Notices, (9), 2698-2768 (2020).

In this generality, it is typically only possible to find (or plot) pursuit curves by computer integration of the ode's. When say c moves with unit speed, then letting $\alpha = \angle(c', \gamma - c)$ and $d = |c - \gamma|$, we find the system of ode's:

$$d' = -v - \cos\alpha, \quad \alpha' = \frac{1}{d}\sin\alpha - \kappa$$

(κ the curvature of $\mathcal{C}$) to determine pursuit curves $\gamma(s) = c(s) + d(s)e^{i\alpha(s)}c'(s)$. Still in this case, one may typically only precede numerically (although one may still determine certain properties of such pursuit curves, eg when $v > 1$ that $d' < 0$ so distance is 'closing in').

A slight variant on pursuit curves, which may be solved explicitely, is the following situation: a point moves on a curve $c(s)$ with constant (say unit) velocity $|c'(s)| = 1$, so that its tangent makes a fixed angle α wrt the line of sight to a fixed point. In symbols: $\angle(c'(s), c(s)) = \alpha = cst.$ (taking the fixed point as the origin). The solutions are *logarithmic spirals*:

$$r = ae^{b\theta}$$

in polar coordinates for some constants a, b.

Some more curves (useful in design of mechanical devices) can be produced by a *linkage system* of rods. These curves are sometimes called *coupler curves*: consider a system of rods connecting two fixed points and take the curve traced by a point on one of the rod's segments during the motion of the system.

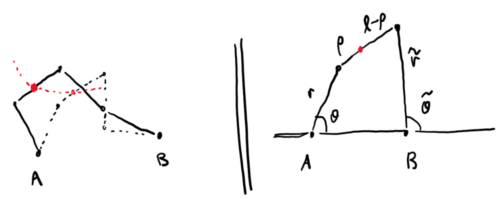

Figure 17. A linkage system of rods may be used to draw curves (the trace of the red point fixed to a rod). The system on the right (usually with $r = \tilde{r}, \rho = \ell/2$) generates what are called *Watt curves*, due to their study by J. Watt in his work on steam engines. In principle they may be found by determining $\tilde{\theta}(\theta)$ through $|1 + \tilde{r}e^{i\tilde{\theta}} - re^{i\theta}| = \ell$ and the Watt curve by $\gamma(\theta) = re^{i\theta} + \frac{\rho}{\ell}(1 + \tilde{r}e^{i\tilde{\theta}(\theta)} - re^{i\theta})$.

Rolling curves: One may generate curves through 'rolling processes'. Rolling a given plane curve $\mathcal{C}$ over the (fixed) plane curve $\tilde{\mathcal{C}}$, may in general be described by giving a (smooth) curve of isometries $\varphi_t : \mathbb{E}^2 \to \mathbb{E}^2$, with the property that for each $t \in \mathbb{R}$, $\varphi_t(\mathcal{C}) =: \mathcal{C}_t$ is tangent to $\tilde{\mathcal{C}}$ at some point $\tilde{p}_t \in \tilde{\mathcal{C}}$.

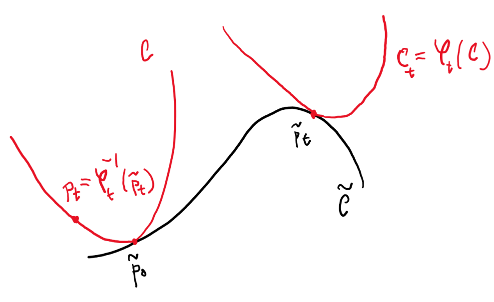

Figure 18. A general 'rolling' of $\mathcal{C}$ over $\tilde{\mathcal{C}}$ is given by a 'curve of isometries' φ_t (wlog one may consider $\varphi_o = id$ and that initially $\mathcal{C}$ is tangent to $\tilde{\mathcal{C}}$ at some point).

Observe that rolling is a symmetric process: if φ_t determines a rolling of $\mathcal{C}$ over $\tilde{\mathcal{C}}$, then so does $\tilde{\varphi}_t = (\varphi_t)^{-1}$ determine a rolling of $\tilde{\mathcal{C}}$ over $\mathcal{C}$.

The above description of rolling is more general than what we typically have in mind, namely, *rolling without slipping*, which may be defined by either of the following two (equivalent) additional conditions:

- For $p_t := \varphi_t^{-1}(\tilde{p}_t) \in \mathcal{C}$, the arc-length from p_o to p_t along $\mathcal{C}$ is the same as the arc-length from $\tilde{p}_o$ to $\tilde{p}_t$ along $\tilde{\mathcal{C}}$.

- The point of contact has zero velocity, meaning for each t we have: $\frac{d}{d\tau}|_{\tau=t}\varphi_\tau(p_t) = 0$.

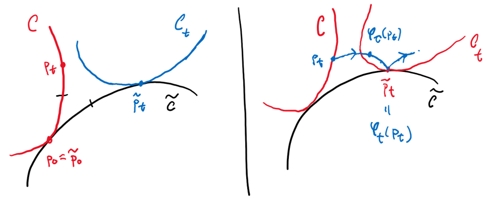

Figure 19. Two characterizations of rolling without slipping (what we will usually just call rolling).

A rolling without slipping of C over $\tilde{C}$ determines a 1-parameter family of isometries $\varphi_s,\ s \in \mathbb{R}$. The traces of a given point in the plane: $\gamma(s) = \varphi_s(q)$ under this family are curves called *roulettes*.

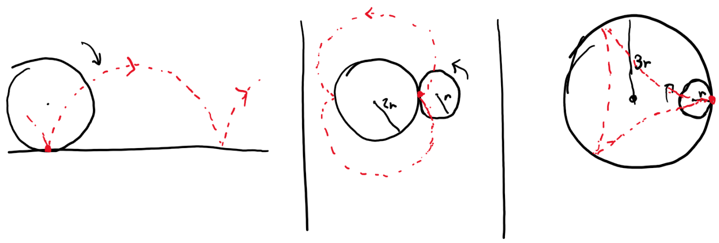

Figure 20. Some rolling curves: trace the red point fixed to a circle as it rolls. A circle along a line gives a cycloid, rolling an exterior circle along a circle gives an epicycloid (whose form will depend on the ratio of the radii of the two circles, here we have drawn what is called a *nephroid*), rolling an interior circle along a circle gives a hypocycloid (we have drawn here what is called a *deltoid*).

One may generate in this way various examples of curves. For example: a *cycloid* is obtained by tracing a point fixed to a circle as the circle rolls along a line, an *epicycloid* is obtained by tracing a point attached to an (exterior) circle rolling along another circle, a *hypocycloid* is obtained by tracing a point attached to an (interior) circle rolling along another circle.

The *Delaunay roulettes* result from tracing a focus of a conic section rolled along a line. Later we see that the surfaces of revolution obtained from these curves have a remarkable property: they (along with a sphere) are exactly the surfaces of revolution with constant mean curvature[1]. They come in three types: *catenary, undulary,* or *nodary* curves by rolling parabolas, ellipses, or hyperbolas (resp.) along the line.

One may derive some equations/properties of these Delaunay roulettes, making use of the geometric properties of conics and rolling depicted in the following figures (21-23):

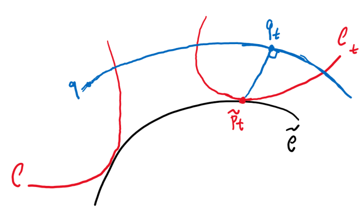

Figure 21. A roulette, $q_t = \varphi_t(q)$, has the property that its tangent at q_t is perpendicular to the line joining q_t to the point of contact, $\tilde{p}_t$. One may see this from: $0 = \frac{d}{d\tau}|_{\tau=t}(q - p_t)\cdot(q - p_t) = \frac{d}{d\tau}|_{\tau=t}(q_t - \varphi_\tau(p_t))\cdot(q_t - \varphi_\tau(p_t)) = 2\dot{q}_t \cdot (q_t - \tilde{p}_t)$ (using the no slip condition $\frac{d}{d\tau}|_{\tau=t}\varphi_\tau(p_t) = 0$).

[1] See for instance J. Eells' article: *The surfaces of Delaunay.*

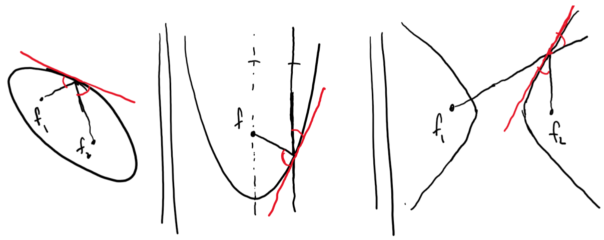

Figure 22. *Optical properties* of conic sections, eg the tangent to an ellipse makes equal angles with the lines to the foci.

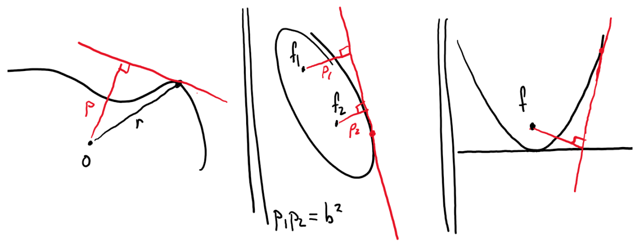

Figure 23. The *pedal equation* of a curve (with respect to an origin o) is a relation between the distance, p, from o to the tangent line to the curve at q, and the radial distance, $r = |qo|$, from o to q. The pedal equations of conic sections may be computed explicitely, from which some useful *pedal properties* of conic sections follow, eg the product of distances from the foci to a tangent line of the ellipse is the semi-minor axis of the ellipse squared.

From these properties, we find that the catenary, undulary and nodary curves arc length parametrizations $(x(s), y(s))$ (with the rolling line as x-axis), satisfy the equations:

$$(y')^2 = 1 - \frac{c^2}{y^2}, \quad c = \frac{\lambda}{2} \quad \text{(catenary, } \lambda = \text{semi-latus rectum of parabola),}$$

$$y^2 - 2ayx' + b^2 = 0 \quad \text{(undulary, } a, b \text{ semi-major/minor axes),}$$

$$y^2 - 2ayx' - b^2 = 0 \quad \text{(nodary, } a, b \text{ semi-major/minor axes),}$$

In particular, from $x'' = -\kappa y'$, it follows that the undulary and nodary's curvatures satisfy the equation:

$$\kappa = \frac{x'}{y} - \frac{1}{a}$$

which will be useful later.

Note that a rolling without slipping of $\mathcal{C}$ along $\tilde{\mathcal{C}}$ may be given explicitely in terms of arc-length parametrizations, $c(s), \tilde{c}(s)$, of the curves (with $c(0) = \tilde{c}(0)$ the initial point of contact). Writing $c'(s) = e^{i\theta(s)}, c(s) = e^{i\tilde{\theta}(s)}$ (with $\theta(0) = \tilde{\theta}(0)$) we have $\varphi_s(p) = e^{i(\tilde{\theta}(s) - \theta(s))}(p - c(s)) + \tilde{c}(s)$, so in particular, Roulettes are the curves $\gamma(s) = \varphi_s(p)$ for some fixed p in the plane.

Lastly, we mention that there is a connection between rolling a closed convex curve along a line and vertices of the curve. If the curve is given a uniform (constant density) mass distribution and subject to constant vertical acceleration than the number of vertices of the curve is the same as the number of balance configurations of the curve.[1]

Optical curves: classical or *geometric optics* models light in a homogeneous medium as (1) steady streams of particles moving along rays (Fermat's principle of least time), or (2) wave fronts consisting of the points reached after a given time by light emitted from a source (Hyugen's principle). Some interesting curve constructions may be generated through such optical models.

[1]See, eg appendix B of P.L. Várkonyi and G. Domokos' article: *Static equilibria of rigid bodies: dice, pebbles, and the Poincaré-Hopf theorem.*

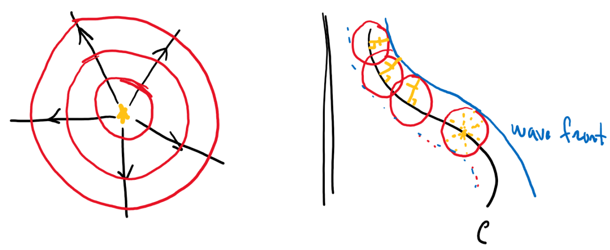

Figure 24. A point light source emits light rays, along which one imagines a stream of light particles moving with constant speed. The wave fronts (circles centered at the point source) are the points reached by particles emitted from the source after a fixed time. Each point of a curve may be viewed as a light source. The wave front consists of a curve tangent to all the wave fronts after a fixed time (circles of some fixed radius) from the individual points of the curve.

When a (plane) curve, $c(t)$, is considered as a light source, its *wave fronts* are the families of curves, Φ_u, parametrized as:

$$t \mapsto \varphi_u(t) = c(t) + uN(t)$$

for each $u \in \mathbb{R}$ and $N(t)$ the unit normal to the curve at $c(t)$. Equivalently, the wave-fronts are (at least for 'small' u-values) curves at distance $|u|$ from c. Observe that the family of wave fronts are in a sense 'dual' to the set of normal lines to the curve, parametrized by

$$u \mapsto \varphi_u(t) = c(t) + uN(t).$$

The curve c as the light source, has, in particular, a constant 'intensity' of light emitted along it. In general, one imagines the intensity of light along each wave front to vary, being represented by a 'density of normal lines' or some 'collapse rate' of the wave front.

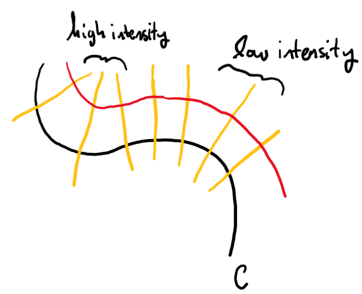

Figure 25. The strength or intensity of the light emitted from a curve $\mathcal{C}$ will in general vary along a given wave front. It may be thought of as a 'density of normal lines' along the wave front.

The *caustic* curve to a given curve, $\mathcal{C}$, may be defined as the set of points obtained as 'intersections of infinitesimally close normal lines to $\mathcal{C}$', that is: $p(t) = \lim_{\varepsilon \to 0} n(t) \cap n(t+\varepsilon)$ where $n(t)$ is the normal line to $\mathcal{C}$ at $c(t)$.

To compute this intersection, we have $c(t) + u_1 N(t) = c(t+\varepsilon) + u_2 N(t+\varepsilon) \Rightarrow 0 = \varepsilon(1 - \kappa(t)u_2)\dot{c}(t) + (u_2 - u_1)N(t) + O(\varepsilon^2) \Rightarrow u_2 = u_1 + O(\varepsilon^2), u_2 = \frac{1}{\kappa(t)} + O(\varepsilon)$. Sending $\varepsilon \to 0$, we find the caustic of the curve $\{c(t)\}$ is parametrized as:

$$t \mapsto c(t) + \frac{1}{\kappa(t)} N(t)$$

ie the centers of the osculating circles to c (also called the *evolute* of c). Observe that the vertices of a curve correspond to 'cusps' (zero velocity points) of its caustic, since velocity along the caustic is: $-\frac{\dot{\kappa}}{\kappa^2}N$.

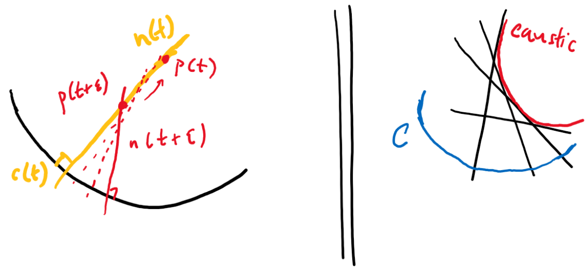

Figure 26. The points on the caustic of a curve are given as a limit of intersections of normal lines to the curve. The caustic of a given curve is tangent to normal lines of the curve at their points of contact.

Caustic curves are special cases of *envelopes*. A 1-parameter family of curves is a collection C_u of curves, $u \in \mathbb{R}$. The envelope of the family may then defined as: the set of points $p(u) = \lim_{v \to u} C_v \cap C_u$. Hence the caustic is the envelope of normal lines to a curve. In formulas, we find:

- If the family of curves is defined implicitly, $f(x, y, u) = 0$, then its envelope is the set of points satisfying $f(x, y, u) = 0, \partial_u f(x, y, u) = 0$.
- If the family of curves is defined parametrically, $t \mapsto c_u(t)$, then its envelope is determined as $c_{u_*}(t_*)$ where (u_*, t_*) is a critical point of $(u, t) \mapsto c_u(t)$ (ie $\det(\partial_u(c_u(t)), \partial_t(c_u(t))) = 0$).

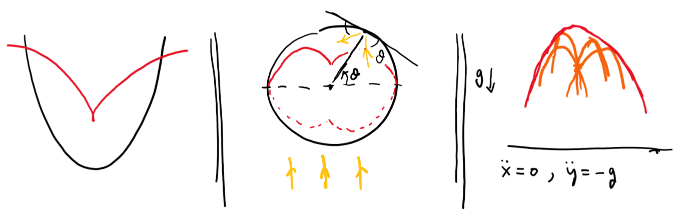

Figure 27. The caustic of a parabola may be computed explicitely. From the parametrization $c(t) = (t, t^2)$, we find $c(t) + \frac{1}{\kappa(t)} N(t) = c(t) + \frac{|\dot{c}(t)|^2}{2}(-2t, 1) = (-4t^3, \frac{1}{2} + 3t^2)$. Eliminating t, we obtain the implicit equation $16(y - \frac{1}{2})^3 = 27x^2$ for the parabola's caustic. Caustics of other optical systems may be found by determining envelopes of certain families of lines. For example, the envelope of the family of lines obtained by reflecting a parallel family of lines off a (unit) circle is a Nephroid curve. Also interesting is a 'fireworks' property. The family of vertical parabolas emanating from a point with fixed energy (in constant gravitational field, so with constant speed) envelope another vertical parabola, the profile one sees in a fireworks explosion.

The *Bezier curves*, are related to envelopes. First recall a line segment between two points may be parametrized by $(1 - t)p_o + tp_1,\ t \in [0, 1]$ (line segments are 'degree 1 Bezier curves'). Given three points, p_o, p_1, p_2 we consider the 1-parameter family of line segments joining $(1-u)p_o + up_1$ to $(1-u)p_1 + up_2$. The envelope of this 1-parameter family of line segments is a degree 2 Bezier curve.
Explicitely, the family of line segments are parametrized by $c_u(t) = (1-t)\left((1-u)p_o + up_1\right) + t\left((1-u)p_1 + up_2\right)$, and their envelope is determined by $0 = \det\left(\partial_u c_u(t), \partial_t c_u(t)\right) = \left(t(1-u) - u(1-t)\right)\det\left(p_1 - p_o, p_2 - p_1\right) \Rightarrow u = t$, and so:

$$b(t) = (1-t)^2 p_o + 2t(1-t)p_1 + t^2 p_2, \quad t \in [0, 1]$$

parametrizes the degree 2 Bezier curve generated by the three points p_o, p_1, p_2.
Similarly, degree k Bezier curves may be generated by $k+1$ points, $p_o, ..., p_k$, as:

$$b(t) = \sum_{j=0}^{k} b_{k,j}(t) p_j, \quad t \in [0, 1]$$

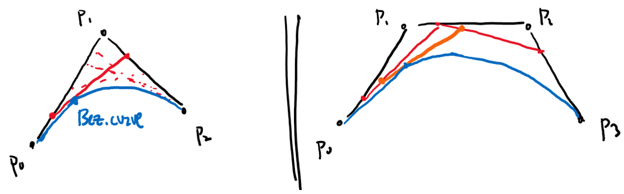

Figure 28. Bezier curves are generated by a given set of points. They were used by the car company Renault (P. Bezier was an engineer for the company) in the designs of the exteriors of their cars. They are commonly used in computer graphics to generate curves interpolating certain points.

where $b_{k,j}(t) = \binom{k}{j}(1-t)^{k-j}t^j$ are the *Bernstein polynomials*. They may iteratively be defined as envelopes of 1-parameter families of line segments: consider the k line segments joining the sequence of points $p_{j,1}(u) := (1-u)p_j + up_{j+1}$, the $k-1$ line segments joining the points $p_{j,2}(u) := (1-u)p_{j,1}(u) + up_{j+1,1}(u)$,...until one has a 1-parameter family of line segments joining the points $p_{o,k-1}(u)$ and $p_{1,k-1}(u)$. The envelope of this family of segments is then parametrized by u as $(1-u)p_{o,k-1}(u) + up_{1,k-1}(u)$.

Variational curves: We consider some examples mentioned in §I.3.

- The *brachistochrone* curve: let $(0,0), (x_o, y_o)$ be two points in the plane (with $y_o < 0$). For planar curves connecting these two points assign the time it takes a 'bead' to fall along the curve from point to point (subject to constant vertical acceleration $(0, -g)$). Then $\frac{|v|^2}{2} + gy = cst.$ by energy conservation, so that for the bead falling along the curve, we have: $|v|^2 = -2gy$ (recall $y_o < 0$). When the curve is given by a graph, $(x, y(x))$, we have then:

$$y(x) \mapsto \int_0^\ell \frac{ds}{|v|} = \int_0^{x_o} \sqrt{\frac{1+(y')^2}{-2gy}}\ dx$$

is our variational problem. One could at this point write out the Euler-Lagrange equations (a 2nd-order ode for $y(x)$), however it is simpler to use the energy integral (a 1st-order ode for $y(x)$, depending on a constant parameter). Namely, we have:

$$cst. = y'\partial_{y'}L - L$$

over extremals, which may be re-arranged as:

$$(y')^2 = -\frac{k^2+y}{y}, \quad k = cst.$$

This may be integrated explicitely, through the following substitutions: $\sqrt{\frac{-y}{k^2+y}}dy = -dx \overset{y=-k^2Y}{\Rightarrow} \sqrt{\frac{Y}{1-Y}}dY = \frac{dx}{k^2} \overset{Y=\sin^2\frac{\theta}{2}}{\Rightarrow} \frac{k^2}{2}(1-\cos\theta)\ d\theta = dx \Rightarrow x = \frac{k^2}{2}(\theta - \sin\theta), y = -\frac{k^2}{2}(1-\cos\theta)$. So the extremals are cycloids!

- The *catenary* curve: let $(x_o, y_o), (x_1, y_1)$ two points in the plane (with $x_o < x_1$ and $y_j > 0$) and consider the form of a hanging chain between these two points. For chains given by graphs $(x, y(x))$, the length of the chain is fixed, $\ell = \int_{x_o}^{x_1}\sqrt{1+(y')^2}dx$ and the potential energy of the chain is given (upto constant multiples) by $\int_{x_o}^{x_1} y\sqrt{1+(y')^2}dx$. By Lagrange multipliers, we seek extremals of:

$$y(x) \mapsto \int_{x_o}^{x_1} (y-\lambda)\sqrt{1+(y')^2}\ dx$$

over $y : [x_o, x_1] \to \mathbb{R}$ with $y(x_j) = y_j$ fixed, and $\lambda \in \mathbb{R}$ some constant. By energy conservation, such extremals satisfy $cst. = y'\partial_{y'}L - L$ which may be rewritten as:

$$\left(\frac{y-\lambda}{k}\right)^2 = 1 + (y')^2$$

for some constant $k \in \mathbb{R}$. The substitution $y - \lambda = k \cosh u$ leads to the equation: $y = k \cosh \frac{x+a}{k} + \lambda$ for the extremals (a, k, λ constants).

- We consider a model for a 'loaded chain' or 'suspension bridge' as a slight variation of the catenary problem.

  Assume that a load along the x-axis is suspended by a chain of fixed length, ℓ, hanging between two fixed points $(x_o, y_o), (x_1, y_1)$ and moreover the load is much more massive than the chain, so that we only consider the mass distribution along the chain due to the load. That is, for $(x(s), y(s))$ an arc-length parametrization of the loaded chain, the mass from $(x(0), y(0))$ to $(x(s_*), y(s_*))$ is $\int_0^{s_*} \rho(s)\, ds = x(s_*) - x(0)$ where $\rho(s)$ gives the mass density along the chain. Observe that

  $$\rho(s) = x'(s).$$

  As with the catenary, the potential energy of the loaded chain arc-length parametrized by $\gamma(s) = (x(s), y(s))$ is now:

  $$V(\gamma) = \int_0^\ell \rho(s) y(s)\; ds$$

  with the constraint of arc-length parametrization $(x')^2 + (y')^2 = 1$. The problem may be treated with Lagrange multipliers as seeking extremals of:

  $$\gamma \mapsto \int_0^\ell \rho(s)y(s) + \frac{\lambda(s)}{2}((x'(s))^2 + (y'(s))^2)\; ds$$

  for some $\lambda : [0, \ell] \to \mathbb{R}$ and over curves $\gamma : [0, \ell] \to \mathbb{R}^2$ with $\gamma(0), \gamma(\ell)$ fixed.

  The Euler-Lagrange equations read:

  $$(\lambda x')' = 0, \quad (\lambda y')' = \rho = x'$$

  Hence $\lambda x' = c \neq 0$ is constant (assuming $x_o \neq x_1$) and $\lambda y' = x + b$ for b some constant. Now:

  $$x + b = \lambda \frac{dy}{ds} = \lambda x' \frac{dy}{dx} = c\frac{dy}{dx} \Rightarrow y = \frac{(x+b)^2}{2c} + a$$

  for a, b, c some constants (chosen to satisfy boundary conditions and fixed length condition).

- The *elastica* curves[1] arise from the following physical situation: a (thin) beam of fixed length is represented by a plane curve with fixed length ℓ. Given two points in the plane and two lines through these points, what shape will the beam bend into when its ends are held fixed to the given points and tangent to the given lines?

  The curves may be characterized variationally: they are extremals of

  $$\gamma(s) \mapsto \int_0^\ell \kappa(s)^2\; ds$$

[1]For more details on the physical principles involved in deriving elastica properties, see §38 of R. Feynman, R. Leighton, M. Sands. *The Feynman lectures on physics*, vol. II. American Journal of Physics 33.9 (1965). On the history of these curves, see R. Levien, *The elastica: a mathematical history.* Electrical Engineering and Computer Sciences University of California at Berkeley (2008). One can find variational derivations and generalizations in the lecture notes of D. Singer, *Lectures on elastic curves and rods.* AIP Conference Proceedings. Vol. 1002. No. 1. American Institute of Physics, 2008.

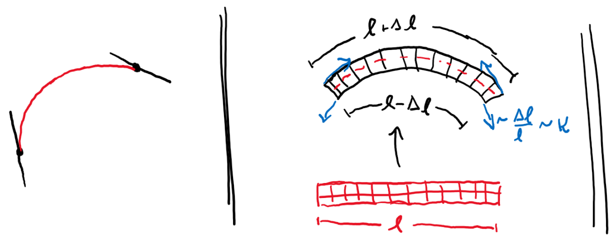

Figure 29. An elastica curve is determined by fixing two points and two lines through these points. The variational characterization may be derived by considering a thin rectangular beam. When the beam is bent its extremities are stretched and compressed. By Hooke's law, forces act on the edges of the beam proportional to the stretching/compressing distance and leading in the limit to an infinitesimal torque, $\tau(s) = \lambda\kappa(s)$, proportional to the curvature acting at each point of the curve. The work (potential energy) done by a torque in twisting an element from the flat straight line configuration to a bent configuration is $\tau(s)d\theta$, where $\theta(s)$ is the angle of the tangent. So the total energy of a bent configuration is $\int_0^\ell \tau(s)\theta'(s)\, ds = \lambda \int_0^\ell \kappa(s)^2\, ds$ with λ a constant (depending on the material properties of the beam).

over arc-length parametrized curves with length ℓ with fixed endpoints and tangent directions at the endpoints. Let $\gamma'(s) = e^{i\theta(s)}$. The function $\theta(s)$ determines the curve γ upto translations. Hence we may rewrite the variational problem in terms of $\theta(s)$ as seeking extremals of: $\theta \mapsto \int_0^\ell (\theta'(s))^2\, ds$ over $\theta : [0, \ell] \to \mathbb{R}$ with $\theta(0), \theta(\ell)$ and $\int_0^\ell e^{i\theta(s)}\, ds$ all fixed. By Lagrange multipliers, we seek extremals of:

$$\theta \mapsto \int_0^\ell (\theta')^2 + \lambda_1 \cos\theta + \lambda_2 \sin\theta\ ds$$

over $\theta : [0, \ell] \to \mathbb{R}$ with $\theta(0), \theta(\ell)$ fixed and $\lambda_j \in \mathbb{R}$ some fixed multipliers. Applying the Euler-Lagrange equations, we arrive at: $\theta'' = \vec{a} \cdot \gamma'$ for some fixed vector $\vec{a}$. An integration yields:

$$\kappa(s) = \vec{a} \cdot \gamma(s) + \vec{b}$$

for some fixed vectors $\vec{a}, \vec{b}$ (chosen to satisfy the boundary conditions). Alternately, the elastica curves are characterized by the curvature at a given point being proportional to the distance of this point to some fixed line in the plane.

EXERCISES:

1. Derive equations for the (signed) curvature of a plane curve in polar coordinates.

2. Show that a spatial curve is a helix[1] iff it has constant torsion and constant curvature.

3. Show that a spatial curve is a generalized helix[2] iff its curvature and torsion satisfy $\tau = \lambda\kappa$ for some constant $\lambda \in \mathbb{R}$.

4. Show that a spatial curve (with non-zero curvature) is contained in some sphere iff its curvature and torsion satisfy: $\frac{\tau}{\kappa} + \left(\frac{1}{\tau}(\frac{1}{\kappa})'\right)' = 0$ (here, as usual, $' = \frac{d}{ds}$ denotes derivatives wrt arc-length).

5. (a) Find the arc length of one 'arc' of a cycloid (the curve traced after the circle completes one revolution).

   (b) Determine the involute of a cycloid when a string with half the length of an 'arc' of the cycloid is attached at one of the cycloids 'cusps' (point of contact with the line upon which the circle rolls).

6. (a) Let $a > b > 0$. Show the conics $\frac{x^2}{a^2-\rho} + \frac{y^2}{b^2-\rho} = 1$ are confocal (for $\rho \in (-\infty, a^2)$).

   (b) For $\mu \in (-\infty, b^2), \nu \in (b^2, a^2)$, show that the points (x, y) given by intersecting the ellipsoid $\frac{x^2}{a^2-\mu} + \frac{y^2}{b^2-\mu} = 1$ and hyperboloid $\frac{x^2}{a^2-\nu} - \frac{y^2}{\nu-b^2} = 1$ are given by:

$$x^2 = \frac{(a^2-\mu)(a^2-\nu)}{a^2-b^2}, \quad y^2 = \frac{(b^2-\mu)(\nu-b^2)}{a^2-b^2}.$$

   (c) Show that elliptic coordinates are an orthogonal system of coordinates (you may want to try using the 'optical properties' of conics depicted in fig. 22).

7. For a parametrized curve $t \mapsto c(t)$ with $\angle(c(t), \dot{c}(t)) = \alpha = cst.$, show that $c(t)$ parametrizes a logarithmic spiral ($r = ae^{b\theta}$ in polar coordinates, with a, b constants).

8. Consider three points in the plane located at the vertices of an equilateral triangle (say $\gamma_1(0) = 1, \gamma_2(0) = e^{2\pi i/3}, \gamma_3(0) = e^{4\pi i/3}$ as complex numbers) and subject to the pursuit rule that at each instant they move with unit speed in the direction of their counterclockwise neighbor.

   (a) Show the resulting pursuit curves, $\gamma_1(s), \gamma_2(s), \gamma_3(s)$, remain vertices of an equilateral triangle for all time: $\gamma_2(s) = e^{2\pi i/3}\gamma_1(s) = e^{4\pi i/3}\gamma_3(s)$ (suggestion: apply uniqueness of solutions to ode's with given initial conditions)

   (b) Show the resulting pursuit curves follow logarithmic spirals ($\angle(\gamma_j(s), \gamma_j'(s)) = cst.$).

9. Show the optical properties of the conic sections (illustrated in fig. 22).

10. (a) For the parameters, r, p involved in the pedal equation of a parametrized plane curve, $t \mapsto \vec{c}(t)$, verify that at $\vec{c}(t)$, we have: $r = |\vec{c}(t)|, p = \frac{|\det(\vec{c}(t), \dot{\vec{c}}(t)|}{|\dot{\vec{c}}(t)|}$.

    (b) Show the pedal equation of a parabola from its focus is $p^2 = \frac{\lambda}{2}r$, where λ is the semi-latus rectum of the parabola.

    (c) Show the pedal equation of an ellipse from a focus is $\frac{b^2}{p^2} = \frac{2a}{r} - 1$ and the pedal equation of an hyperbola from a focus is $\frac{b^2}{p^2} = \frac{2a}{r} + 1$ (here a, b are the semi-major, semi-minor axes).

    (for these exercises, you may want to use the polar equation for a conic from its focus: $r = \frac{\lambda}{1+e\cos\theta}$)

11. (a) Show that for an ellipse or hyperbola that the distances, p_1, p_2 from the foci to a tangent line satisfy $p_1 p_2 = b^2$ (for b the semi-minor axis).

    (b) Show that for a general tangent line to a parabola, the perpendicular dropped from the focus of the parabola to this tangent line lies on the tangent line to the parabola at its vertex (see figure 23).

[1] Helices are curves admitting parametrizations of the form $t \mapsto (a\cos t, a\sin t, bt)$ for some constants $a, b \in \mathbb{R}$.
[2] A generalized helix is a curve for which there exists some fixed unit vector $\hat{v}$ such that $T \cdot \hat{v} = cst.$.

12. Describe (via a parametrization or implicit equation) the roulette curves obtained by tracing a point attached to a circle of radius r as it rolls along:

    (a) the outside of a circle of radius r (a cardioid)

    (b) the inside of a circle of radius $4r$ (an astroid).

13. Show the two definitions of rolling without slipping (fig. 19) are equivalent.

14. Let γ be an involute of $\mathcal{C}$. Show that the evolute of γ is $\mathcal{C}$.

15. (a) Consider a family of curves given in implicit form: $\mathcal{C}_u = \{f(x, y, u) = 0\}$. Show that points of the envelope satisfy the system of equations $f(x, y, u) = 0, \partial_u f(x, y, u) = 0$.

    (b) Consider a family of curves given in parametric form: $\mathcal{C}_u$ is parametrized by $t \mapsto c(t, u)$. Show that points of the envelope are given by $c(t, u)$ s.t. $0 = \det(\partial_t c, \partial_u c)$.

16. Suppose the family of curves $\mathcal{C}_u$ envelope the curve γ. Show that γ is tangent to each curve of the family which it intersects.

17. (a) Consider a family of parallel lines reflected once off a circle. Show the envelope of this reflected family is a nephroid curve (obtained by rolling a circle of radius r along the outside of a circle of radius $2r$).

    (b) Consider the family of vertical parabolas which are trajectories in a uniform gravitational field: $\ddot{x} = 0, \ddot{y} = -g$. Determine the envelope of the 1-parameter family of such parabolas passing through a point and with fixed initial speed.

18. For curves, $t \mapsto (x(t), 0, z(t))$, in the xz-plane with fixed length and connecting two fixed points $(x_o, z_o), (x_1, z_1)$ (with $x_j > 0$) determine extremals of the functional sending such a curve to the area of the surface of revolution generated by revolving around the z-axis.

§5 remarks

In this section we remark on some 'odds and ends' that have been omitted or glossed over in the notes above.

Differentials: Consider a function $f : \mathcal{C} \to M$ defined on a curve $\mathcal{C} \subset \mathbb{E}^2$ and taking values in some 'smooth space'[1], $M \subset \mathbb{R}^n$, eg $M = \mathbb{R}^n$ for concreteness.

Intuitively, the differential df of f measures 'infinitesimal changes' in the values of f as follows: fix $p \in \mathcal{C}$ and let $q = p + dp \in \mathcal{C}$ be close to p, ie $|dp| = |q - p|$ is small. Then $f(p + dp) - f(p)$ is the resulting change in f, and df will provide a measure of this resulting change as $dp \to 0$ becomes 'infinitesimal'.

To begin making this more precise, we assume that $\mathcal{C}$ is a smooth plane curve, so that intuitively as $q \to p$, $dp \to 0$ approaches the direction of a tangent vector to $\mathcal{C}$ at p. If $c(t)$ is a parametrization of $\mathcal{C}$ with say $c(0) = p, c(\varepsilon) = q$ then $dp = \varepsilon \dot{c}(0) + o(\varepsilon)$, where the 'correction' or 'remainder' term $o(\varepsilon)$ satisfies

$$\lim_{\varepsilon \to 0} \frac{o(\varepsilon)}{\varepsilon} = 0$$

and $\dot{c}(0) \in T_p\mathcal{C} = \ell_p$ is a tangent vector to $\mathcal{C}$ at p. We capture $dp \to 0$ becoming infinitesimal by considering:

$$dp = \varepsilon \vec{v} + o(\varepsilon), \quad \mathbb{R} \ni \varepsilon \to 0, \quad \vec{v} \in T_p\mathcal{C}.$$

The resulting infinitesimal change in f we may now write as: $f(p + \varepsilon\vec{v} + o(\varepsilon)) - f(p)$ as $\varepsilon \to 0$. When:

$$f(p + \varepsilon\vec{v} + o(\varepsilon)) - f(p) = \varepsilon d_p f(\vec{v}) + o(\varepsilon)$$

for some vector $d_p f(\vec{v}) \in T_pM \subset \mathbb{R}^n$, then this is the 'infinitesimal change' in f at p in the direction $\vec{v}$. Now, we give a definition. The *differential* of f (at p) is the map:

$$df = d_p f : T_p\mathcal{C} \to T_{f(p)}M, \quad \dot{c}(0) \mapsto \frac{d}{dt}|_{t=0} f(c(t))$$

where $\dot{c}(0) \in T_p\mathcal{C}$, for $c(t)$ some parametrization of $\mathcal{C}$ with $c(0) = p$. The function f is *differentiable* when its differential is well-defined (for every $p \in \mathcal{C}$, $\vec{v} \in T_p\mathcal{C}$).

Isometries: The Euclidean plane (or space), $\mathbb{E}^n$, $n = 2$ (or $n = 3$) is a Euclidean vector space, $\vec{\mathbb{E}}^n$, 'upto choice of origin', what is usually called an (Euclidean) affine space.[2] That is, for any $p \in \mathbb{E}^n$, $\vec{v} \in \vec{\mathbb{E}}^n$ there is a *unique* point $p+\vec{v} \in \mathbb{E}^n$. Moreover, every point of $\mathbb{E}^n$ may be obtained in this way, and $(p+\vec{v})+\vec{u} = p+(\vec{v}+\vec{u})$.

Now, $\mathbb{E}^n$ is a metric space: for any two points $p, q \in \mathbb{E}^n$ there is a unique $q - p := \vec{v} \in \vec{\mathbb{E}}^n$ with $q = p + \vec{v}$. The distance between p and q is then $dist(p, q) := \|\vec{v}\| = \sqrt{\vec{v} \cdot \vec{v}}$.

The isometries of $\mathbb{E}^n$ are transformations $f : \mathbb{E}^n \to \mathbb{E}^n$ preserving distance between points. They may all be written in the form:

$$f(p) = A(p - p_o) + f(p_o)$$

with $A : \vec{\mathbb{E}}^n \to \vec{\mathbb{E}}^n$ an orthogonal transformation (reflection/rotation) and $p_o \in \mathbb{E}^n$ a fixed point.

The *orthogonal transformations*, denoted by O_n, are linear transformations of $\vec{\mathbb{E}}^n$ characterized by $AA^T = id$. Those with determinant one (preserving orientation) are called *rotations* or *special orthogonal transformations* and denoted by SO_n.

- Planar rotations, SO_2, may be identified with angles. They are determined by their action on an orthonormal basis: $e_1 \mapsto \cos\theta e_1 + \sin\theta e_2, e_2 \mapsto -\sin\theta e_1 + \cos\theta e_2, \quad \theta \in S^1 = \mathbb{R}/2\pi\mathbb{Z}$. In matrix form: $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \in \mathrm{SO}_2$.

[1] What we have in mind are called *manifolds* (embedded in $\mathbb{R}^n$), whose precise definition we will not worry about. The surfaces in space we consider in the following section are examples of 2-dimensional manifolds. Intuitively $M \subset \mathbb{R}^n$ is a k-dimensional manifold if it may be locally parametrized smoothly by k-parameters. In particular it follows that there is a well defined tangent space to M at each point, $T_mM \subset \mathbb{R}^n$, a k-dimensional affine subspace of $\mathbb{R}^n$ directed by all possible velocity vectors of curves in M passing through $m \in M$.

[2] A Euclidean vector space is a vector space with an inner product (dot product). In different language, a Euclidean affine space is a set (of points) upon which a Euclidean vector space acts freely and transitively.

- Spatial rotations, SO_3, may be visualized with axes of rotation and angles. Any element $A \in \mathrm{SO}_3$ admits an axis of rotation: $A\hat{a} = \hat{a}$ for $\hat{a} \in S^2$ a unit vector and consequently preserves the plane $\hat{a}^\perp$ on which it acts as a planar rotation by some angle $\theta \in S^1$. Note however that $(\hat{a}, \theta) \in S^2 \times S^1$ is not a global parametrization of SO_3, since for instance $\pm\hat{a}$, $\theta \equiv_{2\pi} \pi$ represent the same transformation, and $\hat{a}, \theta \equiv_{2\pi} 0$ for any $\hat{a} \in S^2$ all represent the identity transformation. The situation is similar to using spherical coordinates to parametrize the sphere (where one has problems of 1-1'ness at the poles).

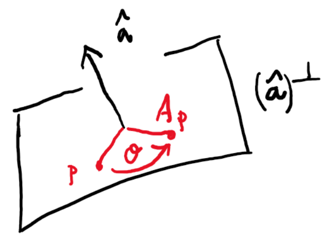

Figure 30. A spatial rotation may be described by an axis of rotation, $\hat{a} \in S^2$ and an angle $\theta \in S^1$. The angle θ is oriented according to the direction of $\hat{a}$ by 'right hand rule'. The points $(\hat{a}, \theta), (-\hat{a}, \tilde{\theta}) \in S^2 \times S^1$ with $\theta \equiv_\pi \tilde{\theta}$ represent the same transformation.

An *infinitesimal rotation* is a transformation $\Omega := \dot{R}(0) : \vec{\mathbb{E}}^n \to \vec{\mathbb{E}}^n$ where $t \mapsto R(t) \in \mathrm{SO}_3,\ R(0) = id$ is a 'smooth curve of rotations'. Infinitesimal rotations are skew-symmetric linear maps $\Omega^T = -\Omega$, and denoted by $\mathfrak{so}_3$. Every infinitesimal rotation ($n = 3$) may be represented by an *infinitesimal axis of rotation*, $\vec{\omega} \in \mathbb{R}^3$ defined through the relation: $\Omega(\vec{v}) = \vec{\omega} \times \vec{v}, \quad \forall \vec{v} \in \mathbb{R}^3$.

The set of isometries of $\mathbb{E}^n$ form a group, called the *Euclidean group* and denoted E_n. Upon fixing a basepoint, $p_o \in \mathbb{E}^n$, they are identified with translations and reflections/rotations (orthogonal transformations): $f \in \mathrm{E}_n \leftrightarrow (A, \vec{b}) \in \mathrm{O}_n \times \vec{\mathbb{E}}^n$ through $f(p) = A(p - p_o) + f(p_o), \quad A \in \mathrm{O}_n, \vec{b} = f(p_o) - p_o \in \vec{\mathbb{E}}^n$. The group structure (composition of transformations) becomes that of a semi-direct product: $\mathrm{E}_n \cong \mathrm{O}_n \ltimes \vec{\mathbb{E}}^n$ with $(A_2, \vec{b}_2) \cdot (A_1, \vec{b}_1) = (A_2 A_1, A_2\vec{b}_1 + \vec{b}_2)$.

The group structure and its action is often represented by considering $\vec{\mathbb{E}}^n \hookrightarrow \mathbb{R}^{n+1},\ \vec{v} \mapsto \begin{pmatrix} \vec{v} \\ 1 \end{pmatrix}$ and $\mathrm{E}_n \cong \mathrm{O}_n \ltimes \vec{\mathbb{E}}^n \hookrightarrow \mathrm{GL}_{n+1}(\mathbb{R}), \quad (A, \vec{b}) \mapsto \begin{pmatrix} A & \vec{b} \\ 0 & 1 \end{pmatrix}$, for which:

$$\begin{pmatrix} A & \vec{b} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \vec{v} \\ 1 \end{pmatrix} = \begin{pmatrix} A\vec{v} + \vec{b} \\ 1 \end{pmatrix},$$

$$\begin{pmatrix} A_2 & \vec{b}_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A_1 & \vec{b}_1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} A_2 A_1 & A_2\vec{b}_1 + \vec{b}_2 \\ 0 & 1 \end{pmatrix}.$$

Order of contact: the definition of osculating circle may be given in the language of 'order of contact'. We will consider the case of plane curves (the definitions may be generalized to 'order of contact' between embedded manifolds in an ambient space).

One may approach order of contact 'iteratively'. Two plane curves, $\mathcal{C}_1, \mathcal{C}_2 \subset \mathbb{E}^2$ are said to have *0'th-order contact* at a point p when they intersect at p, and *1'st-order contact* at p when they intersect at p and are tangent at p. So 1'st-order contact curves at p are a subset of 0'th-order contact curves at p.

A concrete analytic way to define higher order contact in the plane is: consider two curves with first order contact at p, take Cartesian coordinates with p as the origin and x-axis as their common tangent. Then the two curves are given locally as graphs: $y = f_1(x), y = f_2(x)$ (with $f_j(0) = f_j'(0) = 0$) in these coordinates. Since the curves are smooth, the functions $f_j(x)$ may be Taylor expanded around $x = 0$:

$$f_1(x) = a_2x^2 + a_3x^3 + ..., \quad f_2(x) = b_2x^2 + b_3x^3 + ...$$

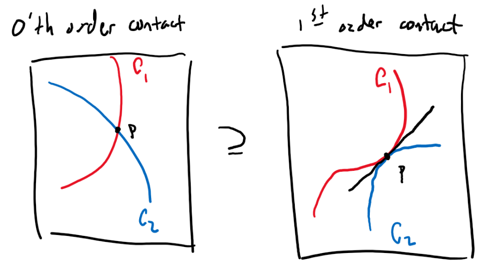

Figure 31. Plane curves with 0-order contact at p (intersecting) and 1st-order contact at p (tangent).

The curves are said to have *k'th-order contact* at p when $a_j = b_j$ for $j = 2, ..., k$.

An equivalent definition to osculating circles to the one we have given above is that the osculating circle to a curve at a point p is the circle with 2nd-order contact to the curve at p.

There is yet another more 'high brow' way in which to phrase order of contact which generalizes more easily to situations other than plane curves. Let us call a *contact element* in the plane a 'pointed line' in the plane, ie a pair $p \in \mathbb{E}^2$ and line $\ell \subset \mathbb{E}^2$ with $p \in \ell$.

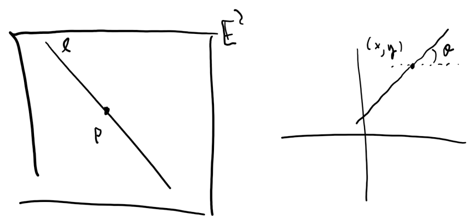

Figure 32. A planar contact element. One may coordinatize contact elements by $(x, y, \theta) \in \mathbb{R}^2 \times S^1$.

The set of planar contact elements is 3-dimensional, since we may coordinates it with $(x, y, \theta) \in \mathbb{R}^2 \times S^1$ where (x, y) are coordinates of a point on the plane, and θ is the angle from the x-axis of a line passing through (x, y). We will denote the set of contact elements by $J^1(\mathbb{E}^2)$.

Now we may rephrase the definition of first order contact in a way that will generalize naturally to k'th-order contact. First, observe that any (smooth) plane curve, $\mathcal{C} \subset \mathbb{E}^2$ lifts to a smooth curve, $\mathcal{C}^1 \subset J^1(\mathbb{E}^2)$, consisting of the pairs $p \in \mathcal{C}$ and ℓ the tangent line to $\mathcal{C}$ at p. Two plane curves intersecting at a point p then have 1'st-order contact at p iff their lifts to $J^1(\mathbb{E}^2)$ have 0'th order contact (intersect) in $J^1(\mathbb{E}^2)$ over p.

The iterative process to define higher order contact proceeds as follows: a contact element in $J^1(\mathbb{E}^2)$ is a 'pointed line'[1] in $J^1(\mathbb{E}^2)$. The set of all contact elements in $J^1(\mathbb{E}^2)$ is denoted $J^2(\mathbb{E}^2)$. Every (smooth) curve in $J^1(\mathbb{E}^2)$ may be lifted to a curve in $J^2(\mathbb{E}^2)$. Hence given a plane curve $\mathcal{C}$, we may lift it to a curve $\mathcal{C}^1 \subset J^1(\mathbb{E}^2)$ and lift this to a curve $\mathcal{C}^2 \subset J^2(\mathbb{E}^2)$. Two curves in $\mathbb{E}^2$ are then said to have 2nd-order contact at a point, when their lifts to $J^1(\mathbb{E}^2)$ have first order contact at a point over p, ie, their lifts to $J^2(\mathbb{E}^2)$ have 0-order contact (intersect) at a point over p.

All this may be summarized in the following 'tower':

$$... \to J^3(\mathbb{E}^2) \to J^2(\mathbb{E}^2) \to J^1(\mathbb{E}^2) \to \mathbb{E}^2$$

with each $J^{k+1}(\mathbb{E}^2)$ being the set of contact elements of $J^k(\mathbb{E}^2)$. A smooth plane curve, $\mathcal{C} \subset \mathbb{E}^2$, may be lifted to a curve $\mathcal{C}^k \subset J^k(\mathbb{E}^2)$, and two plane curves have k'th-order contact at p when their lifts to $J^k(\mathbb{E}^2)$ intersect at a point over p.

[1] More precisely, a point $\xi \in J^1(\mathbb{E}^2)$ and a line in the tangent space to $J^1(\mathbb{E}^2)$ at ξ.

Exercises:

1. Let $f : \mathbb{E}^n \to \mathbb{E}^n$ be an isometry and $p_o \in \mathbb{E}^n$. Show that we have $f(p) = A(p - p_o) + f(p_o)$ for $A : \vec{\mathbb{E}}^n \to \vec{\mathbb{E}}^n$ a transformation preserving norms: $\|A\vec{v}\| = \|\vec{v}\|, \ \ \forall \vec{v} \in \vec{\mathbb{E}}^n$.

2. Show that the transformation A defined in the previous problem is linear and preserves dot products ($A\vec{u} \cdot A\vec{v} = \vec{u} \cdot \vec{v}, \ \ \forall \vec{u}, \vec{v} \in \vec{\mathbb{E}}^n$).

3. Suppose $A : \vec{\mathbb{E}}^n \to \vec{\mathbb{E}}^n$ is a linear transformation preserving dot products. Show that $AA^T = A^TA = id$. Deduce that $\det(A) = \pm 1$.

4. Let $t \mapsto R(t) \in \mathrm{SO}_3$ be a smooth curve of rotations (this just means that when expressed in a basis, the matrix entries are smooth functions). Show that for each t, the matrices $R(t)^{-1}\dot{R}(t)$ and $\dot{R}(t)R(t)^{-1}$ are skew-symmetric.

5. Let $\vec{\omega} \in \mathbb{R}^3$. Show that the transformation $\vec{v} \mapsto \vec{\omega} \times \vec{v}$ is linear and skew-symmetric. Moreover, show that every infinitesimal rotation may be represented uniquely by such an infinitesimal rotation axis.

6. (a) Let $R \in \mathrm{SO}_3$. Show that there exists $\hat{a} \in \vec{\mathbb{E}}^3$ such that $R(\hat{a}) = \hat{a}$ (an axis of rotation for R).

   (b) For $t \mapsto R(t)$ a smooth curve of rotations with $R(0) = id$ and axes of rotation $\hat{a}(t)$, show that $\dot{R}(0)\hat{a}(0) = 0$. When $\dot{R}(0) \neq 0$, deduce that the infinitesimal axis representing $\dot{R}(0)$ is parallel to $\hat{a}(0)$.

7. For $t \in \mathbb{R}$, let $R(t)$ be rotation about the $\hat{k} = (0, 0, 1)$ axis by angle ωt (where $\omega \in \mathbb{R}$ is a fixed 'angular speed'). Show that the skew-symmetric matrices $R(t)^{-1}\dot{R}(t), \ \dot{R}(t)R(t)^{-1}$ are equal and represented by the infinitesimal rotation axis $\omega\hat{k}$.

§6 charts

A (smooth) surface in Euclidean space, is a subset $\Sigma \subset \mathbb{E}^3$ described (locally) through either

- a parametrization or *coordinate patch*: the image, $\sigma(U) = V \cap \Sigma$, of $\sigma : U \overset{open}{\subset} \mathbb{R}^2 \hookrightarrow \mathbb{E}^3$
- implicitly, as a level set $f^{-1}(a)$, of a map $f : \mathbb{E}^3 \to \mathbb{R}$

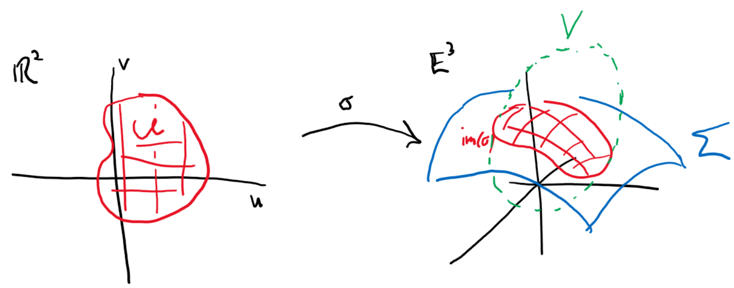

Figure 33. A coordinate chart σ of a surface $\Sigma \subset \mathbb{E}^3$. The 1:1 mapping $\sigma : U \subset \mathbb{R}^2 \to \mathbb{E}^3$ is differentiable and non-degenerate ($\partial_u\sigma, \partial_v\sigma$ span a plane) with $im(\sigma) = V \cap \Sigma$ for some open subset $V \subset \mathbb{E}^3$.

For example:

- The *graph*, $\Gamma(f)$, of a function $f : \mathbb{R}^2 \to \mathbb{R}$. Is globally parametrized by $(x, y) \mapsto (x, y, f(x, y))$, or implicitly as $z = f(x, y)$ (the level set $F(x, y, z) := z - f(x, y) = 0$).
- The *spheres*, $S^2(r)$, with center $c \in \mathbb{E}^3$ are the set of points at distance r from c. Given implicitly (in Cartesian coordinates centered at c) by $x^2 + y^2 + z^2 = r^2$, or parametrized (away from the poles) in spherical coordinates: $(\varphi, \theta) \mapsto (\sin\varphi\cos\theta, \sin\varphi\sin\theta, \cos\varphi)$, for $(\varphi, \theta) \in (0, \pi) \times [0, 2\pi)$.
- The *surfaces of revolution*, are generated by revolving a plane curve about an axis (line) contained in its plane[1]. Taking the axis as a z-axis, and parametrizing the generating curve as $t \mapsto (x(t), 0, z(t))$, the surface of revolution is parametrized by $(t, \theta) \mapsto (x(t)\cos\theta, y(t)\sin\theta, z(t))$.

Smooth surfaces determine a *tangent plane* or best linear approximation to the surface at each point of the surface: the limit of planes passing through $p, a, b \in \Sigma$ as $a, b \to p$. For $p \in \Sigma$, we write $T_p\Sigma$ for the tangent plane to Σ at p. In coordinates, we have:

$$T_p\Sigma = \{p + \vec{v} \;:\; \nabla_p f \cdot \vec{v} = 0\} = p + span\{\partial_u\sigma(u, v), \partial_v\sigma(u, v)\}$$

when Σ is defined around p implicitly by f or parametrized by $p = \sigma(u, v)$.

The *area* of a smooth surface may be defined analogous to lengths of curves. Let $p_{i,j}$ be a discrete 'mesh' or grid of points on Σ, and set:

$$A(\Sigma) = \sup \sum_{i,j} |(p_{i+1,j} - p_{i,j}) \times (p_{i,j+1} - p_{i,j})|.$$

The area of a parametrized patch, $im(\sigma)$, $\sigma : [u_o, u_1] \times [v_o, v_1] \to \mathbb{E}^3$ may be computed by the double integral:

$$\int_{v_o}^{v_1} \int_{u_o}^{u_1} |\partial_u\sigma \times \partial_v\sigma| \; dudv.$$

[1] To generate a smooth surface of revolution, one should have at least that any intersections of the curve with the axis are perpendicular.

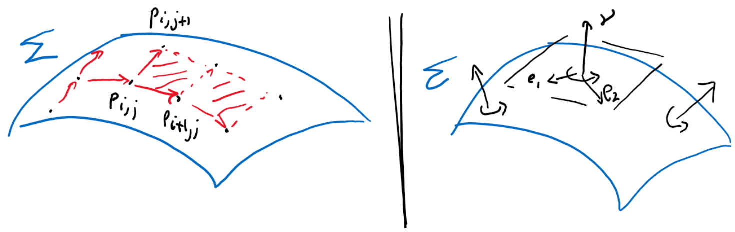

Figure 34. Surface area may be defined as a limit of sums of areas of parallelograms determined by a discrete grid of points along the surface as the distance between the points of the grid goes to zero. Surfaces may be oriented via the choice of a directed normal, ν, along the surface (so that ordered bases e_1, e_2 for the tangent spaces are oriented according to the ordered basis e_1, e_2, ν giving the standard –right hand rule– orientation of space).

Surfaces may also often be equipped with an *orientation*: a class of ordered basis for each tangent space. One typically represents an orientation by choice of directed normal to the surface. One may always orient a surface locally through say a coordinate patch: $\sigma(u,v)$ determines the orientation $\partial_u\sigma, \partial_v\sigma, \nu := \partial_u\sigma \times \partial_v\sigma$ over $im(\sigma)$ or when given implicitly by the normal $\nu := \nabla f$.

§7 curves on surfaces

We introduce two *fundamental forms* associated to a surface embedded in space, which serve to measure curvatures of the surface.

Observe that $\Sigma \subset \mathbb{E}^3$ inherits an 'intrinsic' geometry from the ambient space: given two curves $C_1, C_2 \subset \Sigma$ in the surface, intersecting at $p \in \Sigma$, we may define the *angle* between the two curves as the angle (measured with $\mathbb{E}^3$'s standard 'dot product') between their tangents:

$$\angle_p(C_1, C_2) := \arccos\left(\frac{v_1 \cdot v_2}{|v_1||v_2|}\right) \in [0, \pi)$$

where $v_j \neq 0$ span the tangent lines to the curves at p. As well the *length* of a curve $C \subset \Sigma$ parametrized by $t \mapsto c(t) \in \Sigma$ is just its usual Euclidean length:

$$\ell(C) := \int |\dot{c}(t)|\, dt.$$

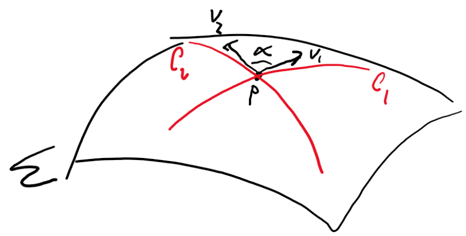

Figure 35. The surface inherits a geometry (angles, lengths) from the ambient space. The measurement of lengths and angles are said to be part of the surfaces *intrinsic* geometry, meaning they could be determined by a being living only on the surface and unaware of the ambient structure.

That is, angle and length are determined by the structure on Σ consisting of inner products on each tangent space to the surface obtained by restriction of the standard dot product of $\mathbb{E}^3$. This collection of inner products on the tangent spaces to the surface is called the *first fundamental form* of the surface. We write, for $p \in \Sigma$ and $p + \vec{u}, p + \vec{v} \in T_p\Sigma \subset \vec{\mathbb{E}}^3$:

$$I_p(\vec{u}, \vec{v}) := \vec{u} \cdot \vec{v}$$

for this first fundamental form, I, of Σ.

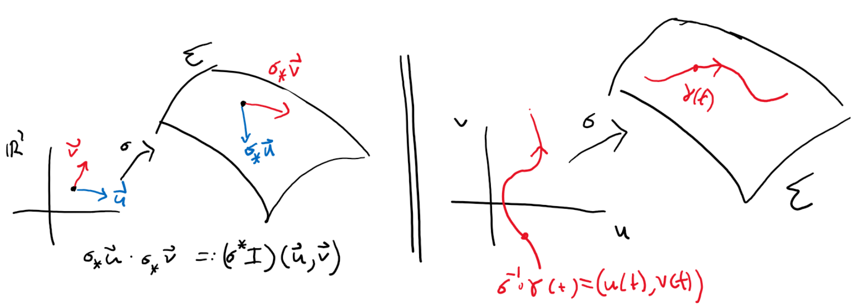

Figure 36. The first fundamental form may be expressed in coordinates and used to compute lengths of curves represented in coordinates: $\sigma_*(\dot{u}, \dot{v}) = \dot{\gamma}$.

In a parametrization $\sigma(u,v)$ of a patch on the surface, we may represent I as inner products on $U \subset \mathbb{R}^2$ *with coefficients depending on* u,v through:

$$(\sigma^* I)_{(u,v)}(\vec{a},\vec{b}) := I_{\sigma(u,v)}(\sigma_*\vec{a},\sigma_*\vec{b})$$

with $\vec{a},\vec{b} \in \mathbb{R}^2$ and $\sigma_*(\vec{a}) = d\sigma_{(u,v)}(\vec{a}) := \frac{d}{dt}|_{t=0}\sigma((u,v)+t\vec{a}) \in T_p\Sigma$ and $p = \sigma(u,v) \in \Sigma$, or for short,

$$\sigma^* I = d\sigma \cdot d\sigma.$$

In the standard basis $(1,0),(0,1)$ of $\mathbb{R}^2$, the coordinate representation,[1] $\sigma^* I$, of I is written:

$$(*) \qquad Edu^2 + 2Fdudv + Gdv^2$$

with $E(u,v) := \partial_u\sigma \cdot \partial_u\sigma$, $F(u,v) := \partial_u\sigma \cdot \partial_v\sigma$, $G(u,v) := \partial_v\sigma \cdot \partial_v\sigma$. The meaning of the coordinate expression $(*)$ is for instance that to find the length of a curve on the surface, $t \mapsto \gamma(t) \in \Sigma$, represented in coordinates as a curve $t \mapsto \sigma^{-1}(\gamma(t)) = (u(t),v(t)) \in U \subset \mathbb{R}^2$, one integrates:

$$\int \sqrt{E\ \dot{u}^2 + 2F\ \dot{u}\dot{v} + G\ \dot{v}^2}\ dt = \int |\dot{\gamma}|\ dt.$$

To introduce the second fundamental form of a surface, we consider the following scheme to measure how the surface is curved in space: at a given point $p \in \Sigma$, let n be the normal line to the surface at p and consider a slice of the surface by a plane, π, containing n. This plane curve $\pi \cap \Sigma$ has an osculating circle whose center lies on n. As one rotates the slicing plane π around n one thus sweeps out a locus of points on the normal line.

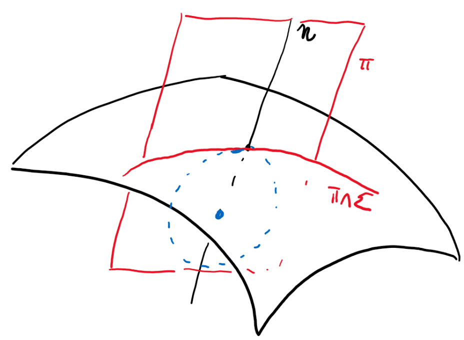

Figure 37. By intersecting the surface with planes containing the normal line, one obtains plane curves in the surface whose osculating circles may be defined. The locus of centers of these osculating circles as the plane is rotated around the normal line sweeps out some intervals on the normal line.

To derive a formula for the curvatures of such 'vertical plane slices' of the surface, we consider the function:

$$\hat{I}I_p(\hat{v}) := \kappa_p(\hat{v})$$

where $\hat{v} \in T_p\Sigma$ is a unit vector and $\kappa_p(\hat{v})$ is the curvature of the (plane) curve obtained by intersecting Σ with the plane through p containing the normal line and $\hat{v}$. The curvature $\kappa_p(\hat{v})$ may be given a sign by locally orienting the surface around p by choice of a unit normal ν to the surface around p. Then:

$$\hat{I}I_p(\hat{v}) = \nu(p) \cdot c''(0)$$

[1] In coordinate computations, the (Einstein) *sum convention* is often used: the coordinates are written with superscripts, (u^1,u^2) in place of (u,v), and any expression with the same index appearing in a subscript and superscript is to be summed over. For example, one would write $\sigma^* I = g_{ij}du^i du^j = g_{11}(du^1)^2 + 2g_{12}du^1du^2 + g_{22}(du^2)^2$, with $g_{11} = E, g_{12} = F, g_{22} = G$ in place of our expression here. Provided one takes care from the context to distinguish between superscripts and powers, the convention is extremely efficient (particularly in higher dimensions).

where $s \mapsto c(s)$ is an arc-length parametrization of the curve $\Sigma \cap span(\nu(p), \hat{v})$ with $c(0) = p, c'(0) = \hat{v}$. We may in fact rewrite $\hat{II}$ to depend only on tangent vectors (and not accelerations) by observing that $0 = \nu(c(s)) \cdot c'(s)$ so that $0 = d\nu_p(\hat{v}) \cdot \hat{v} + \nu(p) \cdot c''(0)$, ie:

$$\hat{II}_p(\hat{v}) = -d\nu_p(\hat{v}) \cdot \hat{v}.$$

The *second fundamental form* of a (oriented) surface with unit normal $\nu : \Sigma \to S^2$ is the structure on the surface consisting of bilinear forms on each tangent space defined by:

$$II_p(\vec{u}, \vec{v}) := -d\nu_p(\vec{u}) \cdot \vec{v}$$

where $\vec{u}, \vec{v} \in T_p\Sigma$ and $d\nu_p(\vec{u}) = \frac{d}{dt}|_{t=0}\nu(c(t))$ for $c(t)$ a curve in Σ with $c(0) = p, \dot{c}(0) = \vec{u}$.

In coordinates, $\sigma(u, v)$, we consider the unit normal $n(u, v) = \nu(\sigma(u, v)) := \frac{\partial_u\sigma \times \partial_v\sigma}{|\partial_u\sigma \times \partial_v\sigma|}$ corresponding to the induced orientation on the patch. Then:

$$\sigma^* II = Ldu^2 + 2Mdudv + Ndv^2$$

where $L(u, v) := -\partial_u n \cdot \partial_u\sigma = n \cdot \partial_u^2\sigma$, $M(u, v) := -\partial_u n \cdot \partial_v\sigma = n \cdot \partial_u\partial_v\sigma$, $N(u, v) := -\partial_v n \cdot \partial_v\sigma = n \cdot \partial_v^2\sigma$. In particular, II is a symmetric bilinear form ($II(\vec{u}, \vec{v}) = II(\vec{v}, \vec{u})$).

Every bilinear form on an inner product space has a representation as a linear operator (matrix). For II this representation is called the *shape operator* or *Weingarten map*, $S_p : T_p\Sigma \to T_p\Sigma$. It is defined through $II_p(\vec{u}, \vec{v}) = I_p(S_p\vec{u}, \vec{v})$, so that:[1]

$$S_p = -d\nu_p.$$

Now we are in position to define certain functions along the surface which measure its curvature. Recall from the geometric contruction leading to II that the centers of curvature of vertical plane slices of the surface sweep out a locus on the normal line (some intervals on this line). The signed curvatures corresponding to the extremal radii of curvature correspond to eigenvalues of the shape operator.

The eigenvalues $\kappa_1(p), \kappa_2(p)$ of S_p are called the *principal curvatures* of the surface. The *Gaussian curvature* of Σ is $K(p) := \kappa_1(p)\kappa_2(p) = \det S_p$ and the *mean curvature* of Σ is $H(p) := \frac{\kappa_1(p)+\kappa_2(p)}{2} = \frac{1}{2}tr(S_p)$.

Much of our efforts in surface theory will be directed towards geometric interpretations and effects of the Gaussian and mean curvature values. For now, let us derive the following formulas for their computation in coordinates:

$$K = \frac{LN - M^2}{EG - F^2}, \quad H = \frac{LG - 2MF + NE}{2(EG - F^2)}.$$

derivation (curvature formulas): For any $\vec{u}, \vec{v} \in \mathbb{R}^2$, we compute:

$$(S\sigma_*\vec{u}) \cdot \sigma_*\vec{v} = II(\sigma_*\vec{u}, \sigma_*\vec{v}) = (\sigma^*II)(\vec{u}, \vec{v}),$$
$$(S\sigma_*\vec{u}) \cdot \sigma_*\vec{v} = I(S\sigma_*\vec{u}, \sigma_*\vec{v}) = (\sigma^*I)(\sigma_*^{-1}S\sigma_*\vec{u}, \vec{v}).$$

Now, in coordinates the bilinear forms σ^*I, σ^*II have matrix representations in the standard basis $(1, 0), (0, 1)$ of $\mathbb{R}^2$ given by

$$g := \begin{pmatrix} E & F \\ F & G \end{pmatrix}, h := \begin{pmatrix} L & M \\ M & N \end{pmatrix}.$$

Equating our two expressions for $(S\sigma_*\vec{u}) \cdot \sigma_*\vec{v}$, we obtain: $(h\vec{u}) \cdot \vec{v} = (gs\vec{u}) \cdot \vec{v}$, where $s := \sigma_*^{-1}S\sigma_*$ is the coordinate representation for the shape operator. Since $\vec{u}, \vec{v}$ were arbitrary, we have: $s = g^{-1}h$, and so $K = \det S = \det s = \frac{\det h}{\det g}$, $H = \frac{1}{2}tr(S) = \frac{1}{2}tr(g^{-1}h)$, yielding the stated equations. □

We now define some fundamental clases of curves on a surface.

The *geodesics*, of Σ may be defined variationally as the extremals of the surfaces length functional. Namely, let $p, q \in \Sigma$ be two points and call the curve $\gamma_* \subset \Sigma$ a geodesic if it is an extremal of $\gamma \mapsto \int |\dot{\gamma}| \, dt$ among curves $\gamma \subset \Sigma$ connecting p to q. One computes that a curve γ_* is a geodesic iff it admits a parametrization with $\gamma_*'' \in (T_{\gamma_*}\Sigma)^\perp$, ie its acceleration is normal to the surface.

[1] Observe that $d\nu_p : T_p\Sigma \to T_pS^2 = \nu(p)^\perp = T_p\Sigma$.

The principal curvatures (eigenvalues of S) have as well associated eigenspaces which – away from the *umbillic* points (where $\kappa_1 = \kappa_2$) – give perpendicular line fields on $\Sigma\backslash\{p \in \Sigma : \kappa_1(p) = \kappa_2(p)\}$. These line fields are called the *principal directions* of the surface, and their integral curves the *lines of curvature* of the surface. Similarly, the null directions of II, $\vec{v} \in T_pM$ s.t. $II_p(\vec{v},\vec{v}) = 0$, determine line fields on $\Sigma\backslash\{p : K(p) > 0\}$ called the *asymptotic directions* of the surface with integral curves called the *asymptotic lines* of the surface.

The above curves may be used to define special coordinate systems (local normal forms of the next section), in which geometric properties of the surface may be more easily described. They may also be characterized via certain 'surface curvature' properties, generalizing our Frenet-frame derivations for spatial curves. Namely, let $C \subset \Sigma$ be an oriented curve on the (oriented) surface with unit normal ν, and consider the frame

$$T, \quad N = \nu \times T, \quad \nu$$

along the curve – called the curves' *Darboux frame* adapted to Σ – where T is the unit tangent to the curve. Then we have the surfaces structural equations:

$$T' = \omega \times T, \quad N' = \omega \times N, \quad \nu' = \omega \times \nu$$

where $\omega = \tau_{geo}T - \kappa_{nor}N + \kappa_{geo}\nu$ defines the *geodesic torsion*, τ_{geo}, the *normal curvature*, κ_{nor}, and the *geodesic curvature*, κ_{geo} of the curve $C \subset \Sigma$.

Observe that when the curve in the surface is parametrized by arc-length, $s \mapsto c(s)$, we have:

$$c'' = \kappa_{geo}N + \kappa_{nor}\nu.$$

In particular, a curve in Σ is a geodesic iff its geodesic curvature vanishes: $\kappa_{geo} \equiv 0$. Likewise, we have:

$$S(c') = -\nu' = \tau_{geo}N + \kappa_{nor}T$$

so that a curve in Σ is a line of curvature iff its geodesic torsion vanishes and is an asymptotic line iff its normal curvature vanishes.

Finally, note that the fundamental forms and their derived notions here are all invariants under isometries of $\mathbb{E}^3$. That is, if two surfaces $\Sigma_1, \Sigma_2 \subset \mathbb{E}^3$ may be taken to eachother by an isometry of $\mathbb{E}^3$ then, for instance, the corresponding points must have the same principal curvature values.

Exercises:

1. Show that the 'area element', $dA := |\partial_u \sigma \times \partial_v \sigma|\ dudv$, of a surface Σ in the coordinate patch $(u,v) \mapsto \sigma(u,v)$ may be written as $dA = \sqrt{EG - F^2}\ dudv$, where E, F, G are the coefficients of the first fundamental form of Σ in the (u,v)-coordinates.[1]

2. Consider the coordinates on the unit sphere, $x^2 + y^2 + z^2 = 1$, through stereographic projection: a point $p \neq (0,0,1)$ of the sphere corresponds to the point in the xy-plane where the line through p and $(0,0,1)$ intersects the xy-plane.

   (a) Show that the stereographic projection is given in coordinates by $(x,y,0) \mapsto \frac{(2x,2y,r^2-1)}{r^2+1}$ where $r^2 = x^2 + y^2$.

   (b) In these (x,y) coordinates of stereographic projection, show the first fundamental form of the sphere[2] is given by:
$$\frac{4}{(1+x^2+y^2)^2}\left(dx^2 + dy^2\right)$$

3. Consider 'Lambert's cylindrical projection' of the unit sphere, $x^2 + y^2 + z^2 = 1$, onto the cylinder $x^2 + y^2 = 1$: a point $p = (x,y,z) \neq (0,0,\pm 1)$ of the sphere corresponds to the intersection of the cylinder with the ray from $(0,0,z)$ to p. Show that areas are preserved by Lambert's projection.

4. Determine the geodesics on a sphere.

5. Consider a 'gnomic' or 'central projection' between a sphere and a plane π not passing through the center of the sphere: a point p of the sphere corresponds to the point on the plane π where the line through p and the center of the sphere intersects π. Show that geodesics of the sphere correspond to straight lines in the plane π.

6. Determine the geodesic curvature of a latitude on the unit sphere.

7. Suppose a curve in a surface, $\mathcal{C} \subset \Sigma \subset \mathbb{E}^3$, is both a geodesic and an asymptotic line. Show that $\mathcal{C}$ is a straight line in $\mathbb{E}^3$.

8. Consider a 'torus of revolution': the surface of revolution obtained by revolving a circle of radius r in the xz-plane with center $z = 0, x = R$ around the z-axis (here $R > r$). Determine an implicit representation of this torus of revolution.

9. Calculate the Gaussian and mean curvature functions of the torus of revolution from the preceding exercise.

10. Let the surface $\Sigma \subset \mathbb{E}^3$ be defined implicitly by $f(x,y,z) = 0$. For $p \in \Sigma$, define the Hessian of f at p as the operator $\vec{v} \mapsto \frac{d}{dt}|_{t=0} \nabla_{p+t\vec{v}} f =: d_p^2 f(\vec{v})$. Show that the Gaussian curvature of Σ is given by[3]:
$$K = -\frac{\det\begin{pmatrix} d^2 f & \nabla f \\ \nabla f^T & 0 \end{pmatrix}}{|\nabla f|^4}.$$

11. Consider the quadratic surface Σ defined implicitly by: $\frac{x^2}{a} + \frac{y^2}{b} + \frac{z^2}{c} = 1$, with $a, b, c \neq 0$ constants. Show its Gaussian curvature at $(x,y,z) \in \Sigma$ is given by:
$$K = \frac{1}{abc\left(\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2}\right)^2}.$$

[1] Suggestion: the vector identity $|\vec{a} \times \vec{b}|^2 + (\vec{a} \cdot \vec{b})^2 = |\vec{a}|^2 |\vec{b}|^2$ is useful.

[2] This expression shows that the stereographic projection between the sphere and the plane is a conformal (angle preserving) mapping.

[3] Suggestion: the determinant of a matrix is independent of the basis in which the matrix is expressed. Consider the 4×4 matrix given here in a basis $(e_1, 0), (e_2, 0), (\nu, 0), (0, 1) \in \mathbb{R}^3 \times \mathbb{R}$, where e_1, e_2, ν are an orthonormal basis for space (so e_1, e_2 are a basis for $T_p\Sigma = \nu^\perp$). One can similarly show that the mean curvature is given by: $H = \frac{d^2 f(\nabla f) \cdot \nabla f - |\nabla f|^2\ tr(d^2 f)}{2|\nabla f|^3}$, when oriented with normal $\nu = \nabla f / |\nabla f|$. Note that $tr(d^2 f) = \Delta f$ is the Laplacian of f.

§8 local normal forms

We consider some particular coordinate expressions and expansions in order to understand the relations of curvature values to local geometry of the surface. Unless mentioned otherwise, consider that an orientation of the surface has been chosen.

Graph expansion: Let $p \in \Sigma$ with principal curvature values $\kappa_1(p), \kappa_2(p)$. Then there exists Cartesian coordinates, (x, y, z), centered at p in which Σ is given in some neighborhood of p as a graph of the form:

$$z = \frac{1}{2}\left(\kappa_1(p)x^2 + \kappa_2(p)y^2\right) + O_3(x, y).$$

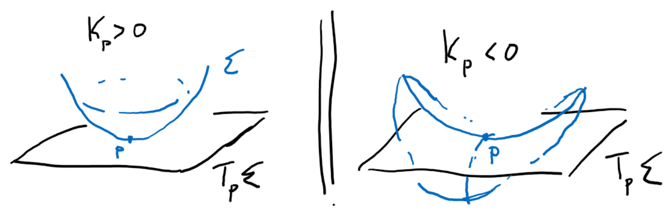

Figure 38. Surfaces may be locally seen as graphs over their tangent plane. When $K_p > 0$ the surface is locally convex (contained on one side of its tangent plane near p), while when $K_p < 0$ the surface is locally 'saddle shaped' around p.

proof: Take Cartesian coordinates centered at p with the xy-plane as $T_p\Sigma$. Then Σ is given around p as some graph $(x, y) \overset{\sigma}{\mapsto} (x, y, f(x, y))$. Since $p = (0, 0, 0)$ in these coordinates, we have $f(0, 0) = 0$ and since $T_p\Sigma$ is the xy-plane we have as well that[1] $f_x(0, 0) = f_y(0, 0) = 0$. We compute:

$$\sigma^* I = (1 + f_x^2)\ dx^2 + 2 f_x f_y\ dxdy + (1 + f_y^2)\ dy^2$$

$$\sigma^* II = \frac{1}{\sqrt{1 + f_x^2 + f_y^2}}\left(f_{xx}\ dx^2 + 2 f_{xy}\ dxdy + f_{yy}\ dy^2\right)$$

when the orientation is given by the 'upward' unit normal $n = \sigma_x \times \sigma_y = \frac{(-f_x, -f_y, 1)}{\sqrt{1+f_x^2+f_y^2}}$.[2] Hence, at p, the shape operator is given in these coordinates by:

$$S_p = g(0,0)^{-1} h(0,0) = d^2 f_{(0,0)} = \begin{pmatrix} f_{xx}(0,0) & f_{xy}(0,0) \\ f_{xy}(0,0) & f_{yy}(0,0) \end{pmatrix}$$

Now, the symmetric matrix $d^2 f_{(0,0)}$ may be diagonalized in an orthonormal basis with diagonal entries the eigenvalues. Hence an appropriate rotation of the xy-coordinates gives (wlog) that $f_{xx}(0, 0) = \kappa_1(p)$, $f_{yy}(0, 0) = \kappa_2(p)$, $f_{xy}(0, 0) = 0$, and the Taylor expansion of $z = f(x, y)$ around $(0, 0)$ yields the stated formula. □

To consider other expansions, we will need to develope some 'natural' coordinate systems. First we derive explicit equations for the geodesics in coordinates.

Geodesic equations: Let $\sigma(u, v)$ be local coordinates on Σ. Then geodesics of Σ are represented by the plane curves $(u(s), v(s))$ satisfying:

$$u'' + \Gamma^u_{uu}(u')^2 + 2\Gamma^u_{uv}u'v' + \Gamma^u_{vv}(v')^2 = 0, \quad v'' + \Gamma^v_{uu}(u')^2 + 2\Gamma^v_{uv}u'v' + \Gamma^v_{vv}(v')^2 = 0$$

[1] We will use often subscripts for partial derivatives, eg $\partial_x f = f_x$.

[2] For a general graph, one finds the expressions $K = \frac{f_{xx}f_{yy} - f_{xy}^2}{1+f_x^2+f_y^2}$, $H = \frac{f_{xx}(1+f_y^2) - 2f_{xy}f_xf_y + f_{yy}(1+f_x^2)}{2\sqrt{1+f_x^2+f_y^2}}$ for the Gaussian and mean curvatures. Given a function, $K(x, y)$, solutions, $f(x, y)$, of the pde $K(x, y) = \frac{\det d^2 f}{1+|\nabla f|^2}$ yield graphs $z = f(x, y)$ with such prescribed Gaussian curvature. Such pde's are examples of *Monge-Ampere equations.*

where

$$\begin{pmatrix} \Gamma^u_{uu} & 2\Gamma^u_{uv} & \Gamma^u_{vv} \\ \Gamma^v_{uu} & 2\Gamma^v_{uv} & \Gamma^v_{vv} \end{pmatrix} = \begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{2}E_u & E_v & F_v - \frac{1}{2}G_u \\ F_u - \frac{1}{2}E_v & G_u & \frac{1}{2}G_v \end{pmatrix}$$

are the *Christoffel symbols* in the coordinates $\sigma(u, v)$.

proof: From the above, an *unparametrized* geodesic is an extremal of the length functional among fixed endpoint curves on the surface and when parametrized with constant speed they are exactly the curves whose acceleration is entirely normal to the surface ($\kappa_{geo} = 0$).[1] First, set:

$$(*) \qquad \sigma_{uu} = \Gamma^u_{uu}\sigma_u + \Gamma^v_{uu}\sigma_v + Ln, \quad \sigma_{vv} = \Gamma^u_{vv}\sigma_u + \Gamma^v_{vv}\sigma_v + Nn$$

$$\sigma_{uv} = \Gamma^u_{uv}\sigma_u + \Gamma^v_{uv}\sigma_v + Mn$$

where $n = \frac{\sigma_u \times \sigma_v}{|\sigma_u \times \sigma_v|}$. Then $\gamma(s) = \sigma(u(s), v(s))$ parametrizes a geodesic iff $\gamma'' \sim n$ is normal to Σ, ie:

$$\sigma_u u'' + \sigma_v v'' + \sigma_{uu}(u')^2 + 2\sigma_{uv}u'v' + \sigma_{vv}(v')^2 \sim n$$

$$\Rightarrow (u'' + \Gamma^u_{uu}(u')^2 + 2\Gamma^u_{uv}u'v' + \Gamma^u_{vv}(v')^2)\sigma_u + (v'' + \Gamma^v_{uu}(u')^2 + 2\Gamma^v_{uv}u'v' + \Gamma^v_{vv}(v')^2)\sigma_v = 0$$

Since σ_u, σ_v are independent, we have the stated equations provided we derive the explicit expressions for the Christoffel symbols. These may be obtained by some algebra and taking inner products of $(*)$ with σ_u, σ_v, to obtain, eg $\frac{1}{2}E_u = \sigma_{uu}\cdot\sigma_u = E\Gamma^u_{uu} + F\Gamma^v_{uu}$ (the first entry above upon applying $\begin{pmatrix} E & F \\ F & G \end{pmatrix}$ to both sides). $\square$

In general the Christoffel symbols are cumbersome and not very practical for computations. However the mere fact that the geodesics are given as solutions to a 2nd order ode allows us to observe for instance that: from a given point, $p \in \Sigma$ with an initial velocity $v \in T_p\Sigma$ there exists a unique geodesic through p with initial velocity v. Moreover, note that if $\gamma(s) \in \Sigma$ parametrizes a geodesic then so does $\gamma(\lambda s)$ for any $\lambda \neq 0$.

Continuing on with our coordinate computations, we obtain:

Gauss' theorema egregium: The Gaussian curvature is an *intrinsic* quantity, that is it can be computed from only the first fundamental form. Explicitely, with $\Delta := \det g = EG - F^2$, we have:

$$4\Delta^2 K = E\left(E_v G_v - 2F_u G_v + G_u^2\right) + F\left(E_u G_v - E_v G_u - 2E_v F_v + 4F_u F_v - 2F_u G_u\right)$$

$$+G\left(E_u G_u - 2E_u F_v + E_v^2\right) - 2\Delta\left(E_{vv} - 2F_{uv} + G_{uu}\right).$$

proof: Continuing from $(*)$, the 3rd derivatives of $\sigma(u, v)$ are:

$$\sigma_{abc} = \Gamma^u_{ab,c}\sigma_u + \Gamma^v_{ab,c}\sigma_v + h_{ab,c}n + \Gamma^u_{ab}\sigma_{uc} + \Gamma^v_{ab}\sigma_{vc} + h_{ab}n_c,$$

for $a, b, c \in \{u, v\}$ (here we write eg $\partial_c\Gamma^u_{ab} = \Gamma^u_{ab,c}$ and $h = \begin{pmatrix} L & M \\ M & N \end{pmatrix} = \begin{pmatrix} h_{uu} & h_{uv} \\ h_{uv} & h_{vv} \end{pmatrix}$). Substituting the definition $(*)$ of the Christoffel symbols for the σ_{uc}, σ_{vc} terms and that $n_c = n^u_c\sigma_u + n^v_c\sigma_v$ with[2] $n^u_c = \frac{Fh_{vc} - Gh_{uc}}{EG - F^2}, n^v_c = \frac{Fh_{uc} - Eh_{vc}}{EG - F^2}$ one has:

$$\sigma_{abc} = \left(\Gamma^u_{ab,c} + \Gamma^u_{ab}\Gamma^u_{uc} + \Gamma^v_{ab}\Gamma^u_{vc} + h_{ab}n^u_c\right)\sigma_u + \left(\Gamma^v_{ab,c} + \Gamma^u_{ab}\Gamma^v_{uc} + \Gamma^v_{ab}\Gamma^v_{vc} + h_{ab}n^v_c\right)\sigma_v$$

$$+ \left(h_{ab,c} + \Gamma^u_{ab}h_{uc} + \Gamma^v_{ab}h_{vc}\right)n.$$

Thus we find relations between the coefficients of the first and second fundamental forms upon equating mixed partials of $\sigma(u, v)$ taken in different orders: $\sigma_{abc} = \sigma_{acb} \Rightarrow$

[1] One may at this point apply the Euler-Lagrange equations to the energy functional, $\gamma \mapsto \int |\dot\gamma|^2\, dt$, to obtain the coordinate expression stated above. However we will proceed here with an alternate method.

[2] Called the *Weingarten equations*.

Codazzi-equations: (equating normal components)

$$L_v - M_u = \Gamma^u_{uv}L + (\Gamma^v_{uv} - \Gamma^u_{uu})M - \Gamma^v_{uu}N,$$

$$N_u - M_v = \Gamma^v_{uv}N + (\Gamma^u_{uv} - \Gamma^v_{vv})M - \Gamma^u_{vv}L.$$

Gauss-equations: (equating tangential components)

$$EK = \Gamma^v_{uu,v} - \Gamma^v_{uv,u} + \Gamma^u_{uu}\Gamma^v_{uv} + \Gamma^v_{uu}\Gamma^v_{vv} - \Gamma^u_{uv}\Gamma^v_{uu} - (\Gamma^v_{uv})^2$$

$$FK = \Gamma^u_{uv,u} - \Gamma^u_{uu,v} + \Gamma^v_{uv}\Gamma^u_{uv} - \Gamma^v_{uu}\Gamma^u_{vv}$$

$$GK = \Gamma^u_{vv,u} - \Gamma^u_{uv,v} + \Gamma^u_{vv}\Gamma^u_{uu} + \Gamma^v_{vv}\Gamma^u_{uv} - \Gamma^v_{uv}\Gamma^u_{vv} - (\Gamma^u_{uv})^2$$

Now observe that in the Gauss equations, one has the Gaussian curvature expressed in terms of the coefficients, E, F, G, of $\sigma^* I$ and its derivatives (using the expressions for the Christoffel symbols in terms of the first fundamental form). □

As a first application of these equations, we have the 'fundamental theorem of surfaces':

Equivalence theorem: Two (oriented) surfaces, $\Sigma_1, \Sigma_2 \subset \mathbb{E}^3$, may be taken to one another by an isometry iff their fundamental forms agree in some parametrization.
Moreover, given any two fundamental forms in coordinates with components satisfying the Codazzi-Gauss equations, there exists a unique –upto isometries– surface in space with the given fundamental forms.

proof: If two surfaces may be taken to eachother by an isometry then their fundamental forms will agree at the corresponding points. More interesting is the converse direction. Let the functions $(E, F, G) = (g_{11}, g_{12}, g_{22}), (L, M, N) = (h_{11}, h_{12}, h_{22})$ of (u^1, u^2) be given, and suppose they satisfy the Codazzi-Gauss equations. We aim to show that there exists a unique –up to isometries– surface in space with the given functions as coefficients of its fundamental forms in some parametrization, $\sigma(u^1, u^2)$.
The existence of the surface is shown by two 'integrations' of pde's, using a 'Frobenius condition'[1] to guarantee solutions. We consider first the pde's that would need to be satisfied by $\sigma_1, \sigma_2, n := \frac{\sigma_1 \times \sigma_2}{|\sigma_1 \times \sigma_2|}$. Set $\mathscr{F} := (\sigma_1\ \sigma_2\ n)$ (a 3×3 matrix). Our sought surface would then need to yield an $\mathscr{F}(u^1, u^2)$ satisfying the system of pde's:

$$(*) \quad \mathscr{F}_1 = \mathscr{F}A, \quad \mathscr{F}_2 = \mathscr{F}B$$

with $A = \begin{pmatrix} \Gamma^1_{11} & \Gamma^1_{12} & n^1_1 \\ \Gamma^2_{11} & \Gamma^2_{12} & n^2_1 \\ L & M & 0 \end{pmatrix}, B = \begin{pmatrix} \Gamma^1_{12} & \Gamma^1_{22} & n^1_2 \\ \Gamma^2_{12} & \Gamma^2_{22} & n^2_2 \\ M & N & 0 \end{pmatrix}$ and $\begin{pmatrix} n^1_j \\ n^2_j \end{pmatrix} = -\begin{pmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{pmatrix}^{-1} \begin{pmatrix} h_{1j} \\ h_{2j} \end{pmatrix}$ given. Now, the Gauss-Codazzi equations[2] are exactly the conditions on the coefficients of the fundamental forms so that 3'rd order partials of σ commute, ie $\mathscr{F}_{12} = \mathscr{F}_{21}$, this 'Frobenius condition' guarantees that given initial conditions there exists a unique solution, $\mathscr{F}(u^1, u^2)$, of $(*)$. Consider a solution of $(*)$ with initial condition $\sigma_1(0), \sigma_2(0), n(0)$ satisfying $\sigma_i(0) \cdot \sigma_j(0) = g_{ij}(0), \sigma_j(0) \cdot n(0) = 0, n(0) \cdot n(0) = 1$. By satisfying $(*)$, we have $\sigma_{ij} = \sigma_{ji}$, so the 'Frobenius condition' again guarantees the existense of $\sigma(u^1, u^2)$ having these first partial derivatives[3], and one may verify that indeed this surface has the desired fundamental forms: $\sigma_i \cdot \sigma_j = g_{ij}, n \cdot \sigma_{ij} = h_{ij}$. Finally, note that *any* other solution satisfying the initial conditions, $\tilde{\sigma}_i(0) \cdot \tilde{\sigma}_j(0) = g_{ij}(0), \tilde{\sigma}_j(0) \cdot \tilde{n}(0) = 0, \tilde{n}(0) \cdot \tilde{n}(0) = 1$ is given from $\sigma_i(0), n(0)$ by applying a fixed rotation, R, ie all solutions of $(*)$ satisfying the initial conditions are given by $\tilde{\mathscr{F}} = R\mathscr{F}$ for a fixed rotation R, hence all surfaces realizing the given fundamental forms are given by $(u^1, u^2) \mapsto R\big(\sigma(u^1, u^2)\big) + p_o$. □

Essentially, this theorem says that *all* the geometry of a surface may be derived from its two fundamental forms. Let us now examine some properties of this geometry in 'special' coordinate systems.

[1] For example: $(A(x, y), B(x, y))$ has $A_y = B_x$ for $(x, y) \in \mathbb{R}^2$ iff there exists a function $f(x, y)$ with $f_x = A, f_y = B$.
[2] They may be written more compactly than how we gave them above in matrix form: $A_2 - B_1 = AB - BA$.
[3] Defined uniquely upto initial value by, eg, $\sigma(u, v) = \int_{(0,0)}^{(u,v)} \sigma_1\, du^1 + \sigma_2\, du^2$ (a path independent integral).

Fermi coordinates: Let $t \mapsto \gamma(t) \in \Sigma$ be a unit speed geodesic, and for each t, let $s \mapsto \sigma(s,t) \in \Sigma$ parametrize the unit speed geodesic with initial condition $\gamma(t)$ and initial velocity perpendicular to $\dot\gamma(t)$. Then $\sigma(s,t)$ parametrizes Σ around $p := \gamma(0)$ and for K_p the Gaussian curvature of Σ at p:

$$\sigma^* I = ds^2 + \left(1 - s^2 K_p + O_3(s,t)\right) dt^2.$$

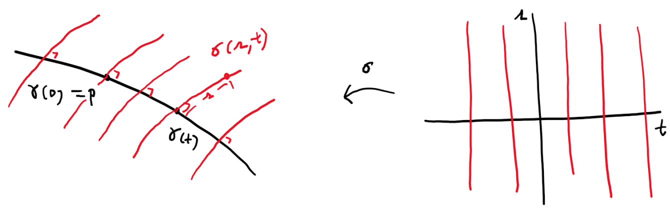

Figure 39. The Fermi coordinates are based off (unit speed) geodesics in the surface perpendicular to a given geodesic.

proof: For each fixed t value, $s \mapsto \sigma(s,t)$ parametrizes a unit speed geodesic, so that $E = \partial_s\sigma \cdot \partial_s\sigma = 1$. Also, the curves $s \mapsto (s,t)$ are geodesics in coordinates, in particular satisfy the geodesic equations so that:

$$\Gamma^s_{ss} \equiv \Gamma^t_{ss} \equiv 0.$$

By coordinate expression of the Christoffel symbols: $0 = \Delta\Gamma^s_{ss} = \frac{1}{2}GE_s - F(F_s - \frac{1}{2}E_t) = -\frac{1}{2}FF_s$, since $E \equiv 1$. Hence $FF_s \equiv 0$, so that $F = \sigma_s \cdot \sigma_t$ is a function only of t. However, when $s = 0$, we have $F = 0$ since $\sigma_s(t,0)$ is perpendicular to $\dot\gamma(t) = \sigma_t(t,0)$. Hence $F \equiv 0$.
Since $\gamma(t) = \sigma(0,t)$ parametrizes a (unit speed) geodesic, we have as well, $G(0,t) = \sigma_t \cdot \sigma_t|_{(0,t)} \equiv 1$. Similar to before, that $t \mapsto (0,t)$ is a geodesic implies $\Gamma^t_{tt}|_{(0,t)} = \Gamma^s_{tt}|_{(0,t)} = 0$, which when written in terms of the coefficients of $\sigma^* I$ read as $G_t(0,t) = G_s(0,t) = 0$. Finally, note that $\Gamma^t_{st} = \frac{G_s}{2G}$, so that:

$$K(0,t) = -\Gamma^t_{st,s}|_{(0,t)} = -\left(\frac{G_{ss}}{2G} - \frac{G_s^2}{2G^2}\right)|_{(0,t)} = -\frac{1}{2}G_{ss}(0,t).$$

Expansion of G now gives: $G(s,t) = G(0,t) + sG_s(0,t) + \frac{s^2}{2}G_{ss}(0,t) + O(s^3;t) = 1 - s^2K(0,t) + O(s^3;t)$. $\square$

This expansion of the metric allows us to describe the Gaussian curvature *geometrically*. For example:

$$d(s,t) = |t|\left(1 - \frac{s^2}{2}K_p + O(s^3;t)\right)$$

where $d(s,t)$ is the distance between the points $\gamma_1(s), \gamma_2(s)$ of two unit speed geodesics emanating perpendicularly from a geodesic $\gamma(t)$ with $\gamma(0) = \gamma_1(0) = p$ and $\gamma_2(0) = \gamma(t)$. Observe as well that it follows from our Fermi coordinates that a sufficiently short – enough to be contained in the image of the Fermi parametrization– geodesic indeed minimizes length between its endpoints:

$$\ell(c) = \int_{\tau_o}^{\tau_1} \sqrt{\dot s^2 + G\dot t^2}\, d\tau \geq \int_{s_o}^{s_1} ds = s_1 - s_o = \ell(\gamma_o)$$

where $c(\tau)$ is any curve in the surface from $\sigma(s_o,0)$ to $\sigma(s_1,0)$ and γ_o is the geodesic $\sigma(s,0), s \in [s_o, s_1]$.
Similarly, one has:

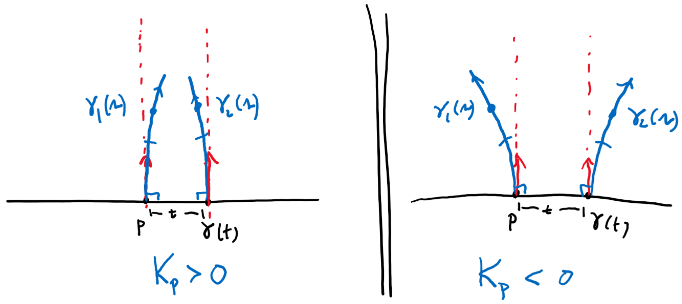

Figure 40. When the Gaussian curvature is positive, distances between nearby geodesics are shorter in comparison with Euclidean distance, while when negative greater.

Geodesic (normal) coordinates: Let $p \in \Sigma$ and for each $\vec{v} \in T_p\Sigma$ let $\gamma_{\vec{v}}(t)$ be the geodesic in Σ with initial conditions $\gamma_{\vec{v}}(0) = p, \dot{\gamma}_{\vec{v}}(0) = \vec{v}$. Then for an orthonormal basis $e_1, e_2 \in T_p\Sigma$,

$$(x, y) \overset{\exp_p}{\mapsto} \gamma_{\vec{v}}(1), \quad \vec{v} = xe_1 + ye_2$$

parametrizes Σ near p. Moreover, with K_p the Gaussian curvature of Σ at p, the first fundamental form in these coordinates is:

$$dx^2 + dy^2 - \frac{1}{3}K_p\,(x\;dy - y\;dx)^2 + O_3(x, y).$$

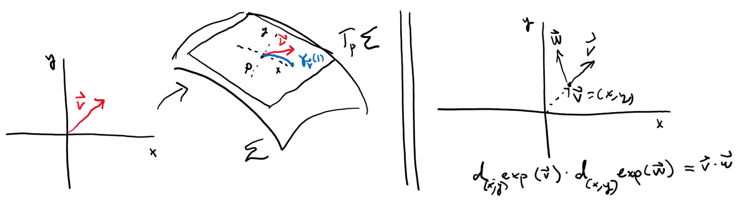

Figure 41. A system of normal coordinates around $p \in \Sigma$ is defined by sending $\vec{v} \in T_p\Sigma$ to $\gamma_{\vec{v}}(1)$ where $\gamma_{\vec{v}}(t)$ is a geodesic through p with initial velocity $\vec{v}$. In these coordinates, one has the *Gauss Lemma*, which states that inner products with radial vectors – $\vec{v} = (x, y)$ at the point (x, y)– are preserved.

proof: First we note that for $\vec{v} \in T_p\Sigma$ and fixed $t \in \mathbb{R}$, the curve $\gamma(\tau) := \gamma_{\vec{v}}(\tau t)$ is a geodesic through p with initial velocity $t\vec{v}$. Hence $\exp_p(t\vec{v}) = \gamma(1) = \gamma_{\vec{v}}(t)$. Let us write $\sigma(x, y) = \exp_p(x, y)$ for the local parametrization. First note that at $p = \sigma(0, 0)$, we have:

$$d_{(0,0)}\sigma(\vec{v}) \cdot d_{(0,0)}\sigma(\vec{w}) = \dot{\gamma}_{\vec{v}}(0) \cdot \dot{\gamma}_{\vec{w}}(0) = \vec{v} \cdot \vec{w}.$$

Hence $E(0, 0) = e_1 \cdot e_1 = 1, G(0, 0) = e_2 \cdot e_2 = 1, F(0, 0) = e_1 \cdot e_2 = 0$. Similarly, in 'radial' directions, we have $d_{\vec{v}}\sigma(\vec{v}) \cdot d_{\vec{v}}\sigma(\vec{v}) = \dot{\gamma}_{\vec{v}}(1) \cdot \dot{\gamma}_{\vec{v}}(1) = |\vec{v}|^2$ since geodesics have constant speed. We now compute the general inner products with a radial vector at a point in the domain of σ. Namely let $\vec{v}, \vec{w} \in T_p\Sigma$. We aim to find $d_{\vec{v}}\sigma(\vec{v}) \cdot d_{\vec{v}}\sigma(\vec{w})$. Consider the coordinates on $\mathbb{R}^2$ by $(t, \theta) \mapsto t\vec{v}_\theta$, with $\vec{v}_\theta := \vec{v}\cos\theta + \vec{w}\sin\theta$. In Σ we set $\alpha(t, \theta) := \sigma(t\vec{v}_\theta)$. Then:

$$d_{\vec{v}}\sigma(\vec{v}) \cdot d_{\vec{v}}\sigma(\vec{w}) = \partial_t\alpha(1, 0) \cdot \partial_\theta\alpha(1, 0).$$

First, we compute, at general (t, θ)

$$\partial_t\alpha(t, \theta) = \partial_t\left(\sigma(t\vec{v}_\theta)\right) = \partial_t\left(\gamma_{\vec{v}_\theta}(t)\right) = \dot{\gamma}_{\vec{v}_\theta}(t).$$

Next, using that $\gamma_{\vec{v}_\theta}(t)$ is a geodesic, we have:

$$\partial_t\left(\partial_t\alpha \cdot \partial_\theta\alpha\right) = \ddot{\gamma}_{\vec{v}_\theta}(t) \cdot \partial_\theta\alpha(t, \theta) + \dot{\gamma}_{\vec{v}_\theta}(t) \cdot \partial_t\partial_\theta\alpha(t, \theta) = \dot{\gamma}_{\vec{v}_\theta}(t) \cdot \partial_\theta\left(\dot{\gamma}_{\vec{v}_\theta}(t)\right)$$

$$= \frac{1}{2}\partial_\theta \left(\dot{\gamma}_{\vec{v}_\theta}(t) \cdot \dot{\gamma}_{\vec{v}_\theta}(t)\right) = \frac{1}{2}\partial_\theta |\vec{v}_\theta|^2 = \frac{1}{2}\left(|\vec{w}|^2 - |\vec{v}|^2\right)\sin 2\theta + (\vec{v}\cdot\vec{w})\cos 2\theta.$$

In particular, along $\theta = 0$, we have: $\partial_t \alpha \cdot \partial_\theta \alpha|_{(t,0)} = t(\vec{v}\cdot\vec{w})$, so that:

$$(*) \quad d_{\vec{v}}\sigma(\vec{v}) \cdot d_{\vec{v}}\sigma(\vec{w}) = \partial_t\alpha(1,0)\cdot\partial_\theta\alpha(1,0) = \vec{v}\cdot\vec{w}.$$

Now, we return to computing the expansion of $\sigma^* I$ around $x = y = 0$. We have already observed the '0th order terms' are $E(0,0) = G(0,0) = 1, F(0,0) = 0$. The Gauss Lemma, $(*)$, is a useful tool for this expansion. First, since $F(x,y) = d_{(x,y)}\sigma(e_1)\cdot d_{(x,y)}\sigma(e_2)$, we have by the Gauss Lemma:

$$F(x,0) = e_1 \cdot e_2 \equiv 0, \quad F(0,y) \equiv 0 \Rightarrow F_x(0,0) = F_{xx}(0,0) = F_y(0,0) = F_{yy}(0,0) = 0.$$

Next, since $x \mapsto (x,0), y \mapsto (0,y)$ represent geodesics we have $E(x,0) = G(0,y) \equiv 1$, so that:

$$E_x(0,0) = E_{xx}(0,0) = G_y(0,0) = G_{yy}(0,0) = 0.$$

The remaining terms may be found by writing the Gauss Lemma in coordinate form:

$$E(x,y)x + F(x,y)y = x, \quad F(x,y)x + G(x,y)y = y$$

Differentiating the first with respect to y, we obtain $E_y(x,0)x = 0 \Rightarrow E_y(x,0) = 0$ for $x \neq 0$ and so by continuity of E, we conclude[1] $E_y(0,0) = 0$. Similarly, one finds:

$$G_x(0,0) = E_{xy}(0,0) = G_{xy}(0,0) = 0,$$

$$E_{yy}(0) = -2F_{xy}(0), \quad G_{xx}(0) = -2F_{xy}(0).$$

Now by either the Gauss equations or theorema egregium, we obtain:

$$K_p = F_{xy}(0) - \frac{1}{2}\left(E_{yy}(0) + G_{xx}(0)\right) = -\frac{3}{2}E_{yy}(0)$$

and Taylor expansion of $E(x,y)\ dx^2 + 2F(x,y)\ dxdy + G(x,y)\ dy^2$ around $x = y = 0$ yields:

$$dx^2 + dy^2 + \frac{E_{yy}(0)}{2}\left(y^2\ dx^2 - 2xy\ dxdy + x^2\ dy^2\right) + O_3(x,y)$$

$$= dx^2 + dy^2 - \frac{1}{3}K_p\left(x\ dy - y\ dx\right)^2 + O_3(x,y).$$

□

Similarly, there are *geodesic polar coordinates*, related to normal coordinates by $r\cos\theta = x, r\sin\theta = y$. Then the first fundamental form is given by:

$$dr^2 + r^2\left(1 - \frac{r^2}{3}K_p + O(r^3)\right)\ d\theta^2.$$

From this expansion we obtain the following geometric descriptions of Gaussian curvature:

$$d(s,\alpha) = 2s\sin\frac{\alpha}{2} - s^2\frac{K_p}{3}\sin\alpha + O(s^3)$$

$$K_p = \lim_{r\to 0} 3\frac{2\pi r - \ell_\Sigma(C_r)}{\pi r^3} = \lim_{r\to 0} 12\frac{\pi r^2 - A_\Sigma(D_r)}{\pi r^4}$$

where $d(s,\alpha)$ is the distance between the points $\gamma_1(s), \gamma_2(s)$ of two unit speed geodesics emanating from p with initial angle α between their initial velocities. By $\ell_\Sigma(C_r)$ we mean the length measured in the surface of a circle of radius r in Σ centered at p: $C_r = \{q \in \Sigma : dist(q,p) = r\}$, and by $A_\Sigma(D_r)$ the area measured in Σ of a disk of radius r in Σ centered at p: $D_r = \{q \in \Sigma : dist(q,p) \le r\}$.

[1] The vanishing of all first partials of E, F, G at $(0,0)$ is equivalent to the vanishing of all Christoffel symbols at $(0,0)$, $\Gamma^k_{ij}(0,0) = 0$. This can be obtained as well by considering that $t \mapsto (tx, ty)$ are geodesics in coordinates so that $\Gamma^k_{xx}(0,0)x^2 + 2\Gamma^k_{xy}(0,0)xy + \Gamma^k_{yy}(0,0)y^2 = 0$ for any (x,y).

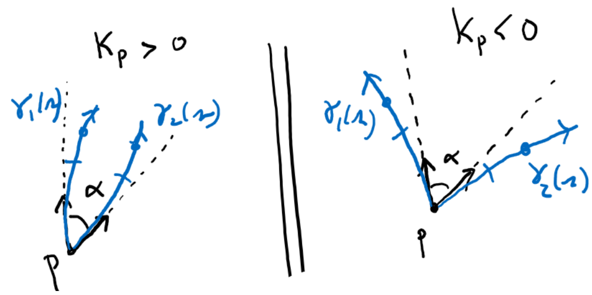

Figure 42. When the Gaussian curvature is positive, the distance in Σ between two geodesic rays emanating from p with angle α grows slower than the Euclidean distance ($dist_{Eucl}(s, \alpha) = 2s \sin \frac{\alpha}{2}$) for small s, while for negative curvature the growth is faster.

Now we describe a fundamental structure on surfaces: the *Levi-Cevita connection*. Geometrically, it is defined via the concepts of *parallel transport* and *covariant derivative*.

Parallel transport is a process to 'transport' tangent vectors along curves in the surface: given $\gamma(t) \in \Sigma$ and $X_o \in T_{\gamma(0)}\Sigma$ tangent to the surface, the parallel transport of X_o is a family of tangent vectors $X(t) \in T_{\gamma(t)}\Sigma$, $X(0) = X_o$ to the surface. When γ is a geodesic of Σ, we take $X(t)$ as the tangent vector to Σ with the same length as X_o and making the same angle with the tangent to the geodesic as X_o. Observe that since the parallel transport $X(t)$ of X_o along a geodesic is a fixed rotation of $\dot{\gamma}(t)$, we have that $\dot{X} \in T\Sigma^{\perp}$ is normal to the surface (as is $\ddot{\gamma}$).

The parallel transport of X_o along a general curve might be imagined as a 'limit' of its transport along piecewise geodesics approximating the given curve. It may also be defined through a 'rolling without slipping or twisting' along the given curve.

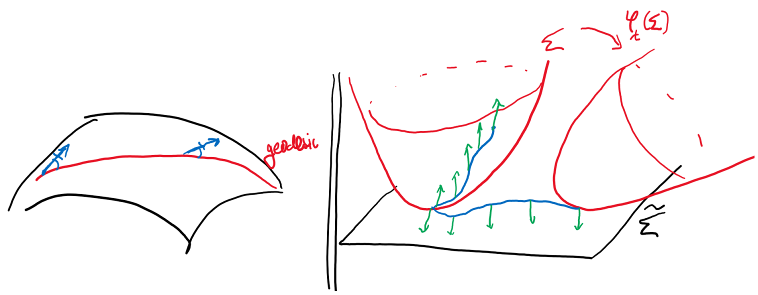

Figure 43. Parallel transport of vectors *along geodesics* may be defined by requiring that the length of the transported vector and its angle with the tangent to the geodesic remain constant. A parallel transport may also be defined through a rolling process: a *rolling* of the surface Σ over the surface $\tilde{\Sigma}$ is a collection of isometries $\varphi_t : \mathbb{E}^3 \to \mathbb{E}^3$ such that $\varphi_t(\Sigma)$ is tangent to $\tilde{\Sigma}$ at some *contact point* $\tilde{\Sigma} \ni \tilde{p}_t = \varphi_t(p_t)$. The curves, $\tilde{\mathcal{C}} = \{\tilde{p}_t\} \subset \tilde{\Sigma}$, $\mathcal{C} = \{p_t\} \subset \Sigma$, are called *developements* of one another. If $\tilde{X}$ is a vector field along $\tilde{\mathcal{C}}$ then one has a corresponding vector field X along the developement, $\mathcal{C}$, of $\tilde{\mathcal{C}}$ in Σ. In this way, taking $\tilde{\Sigma}$ as a plane, one may transfer the usual notion of parallel transport in the plane to the surface. To determine the transport uniquely, the rolling must be *without slipping*: $\frac{d}{ds}|_{s=t}\varphi_s(p_t) = 0$ and *without twisting*: the rotational axis of φ_t is contained in the tangent plane: $T_{\tilde{p}_t}\tilde{\Sigma} = \varphi_t T_{p_t}\Sigma$.

Either of these 'parallel transport schemes' lead to the following differential characterization of parallel transport along a given curve in the surface: $X(t) \in T_{\gamma(t)}\Sigma$ is the parallel transport of $X_o = X(0) \in T_{\gamma(0)}\Sigma$ when $\dot{X}(t)$ is normal to the surface. One writes:

$$\frac{DX}{dt} = \nabla_{\dot{\gamma}} X := pr_{T\Sigma}\dot{X}$$

for the *covariant derivative* of the vectors $X(t) \in T_{\gamma(t)}\Sigma$ along a given curve $\gamma(t) \in \Sigma$. Then $X(t)$ is *parallel* *along* γ iff $\nabla_{\dot\gamma}X = 0$. In coordinates, $X(t) = X^1(t)\sigma_u + X^2(t)\sigma_v, \dot\gamma = \dot u\sigma_u + \dot v\sigma_v$, we have:

$$\nabla_{\dot\gamma}X = \left(\dot X^1 + X^1\dot u\Gamma^u_{uu} + (X^1\dot v + X^2\dot u)\Gamma^u_{uv} + X^2\dot v\Gamma^u_{vv}\right)\sigma_u + \left(\dot X^2 + X^1\dot u\Gamma^v_{uu} + (X^1\dot v + X^2\dot u)\Gamma^v_{uv} + X^2\dot v\Gamma^v_{vv}\right)\sigma_v.$$

Therefore the covariant derivative (projection of usual derivative to tangent space), determines the notion of parallel transport along curves as the (unique) solution $X(t)$ to the ode's $\nabla_{\dot\gamma}X = 0$ with initial condition $X(0) = X_o$. Conversely, a well-defined 'parallel transport scheme' leads to a covariant derivative defined by $\nabla_{\dot\gamma}X(0) := \frac{d}{dt}|_{t=0}X_t$ where $X_t \in T_{\gamma(0)}\Sigma$ is the parallel transport of $X(t) \in T_{\gamma(t)}\Sigma$ along γ to $\gamma(0)$.

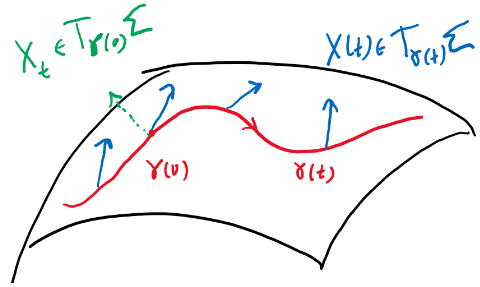

Figure 44. Parallel transport determines a covariant derivative: given a vector field $X(t) \in T_{\gamma(t)}\Sigma$ along a curve, one may take $\nabla_{\dot\gamma}X \in T_{\gamma(0)}\Sigma$ as the derivative of the vectors $X_t \in T_{\gamma(0)}\Sigma$ where X_t is the parallel transport to $\gamma(0)$ along γ of $X(t)$. Conversely a covariant derivative determines a parallel transport by declaring the parallel vector fields to be those with derivative zero.

The structure allowing to differentiate vectors along curves leads to concise expressions for our formulas above and allows one to extend many operations from multivariable calculus to intrinsic operations on the surface (see remarks section).

We finish with a geometric characterization of Gaussian curvature in terms of 'holonomy of parallel transport'. First (see remarks section on the notation of differential forms):

Cartan's structural equations: Let e_1, e_2, e_3 be a (local) orthonormal frame adapted to the surface– so e_1, e_2 are tangent to the surface and e_3 is a unit normal to the surface. For $\omega^1, \omega^2, \omega^3$ the dual basis of e_j, the *connection 1-forms*, $\omega^k_j, \quad j,k \in \{1,2,3\}$, satisfy and are determined by the equations:

$$\omega^k_j = -\omega^j_k, \quad d\omega^k_j = \sum_{\ell=1}^3 \omega^\ell_j \wedge \omega^k_\ell, \quad 0 = \sum_{b=1}^2 \omega^b \wedge \omega^3_b,$$

$$d\omega^a = \sum_{b=1}^2 \omega^b \wedge \omega^a_b, \quad a \in \{1,2\}.$$

proof: Consider $e_j : \Sigma \to S^2$, and $de_j(v) := \frac{d}{dt}|_{t=0}e_j(c(t))$ where $v = \dot c(0) \in T\Sigma$. We write

$$de_j = \omega^1_j e_1 + \omega^2_j e_2 + \omega^3_j e_3$$

defining the connection 1-forms $\omega^k_j : T\Sigma \to \mathbb{R}$. Since $\{e_j\}$ are orthonormal, $e_j \cdot e_k = \delta_{jk} = cst.$, so that

$$0 = de_j \cdot e_k + e_j \cdot de_k \Rightarrow \omega^k_j = -\omega^j_k.$$

Next, note that in a fixed basis $\hat\imath, \hat\jmath, \hat k$, writing $e_j = a_j\hat\imath + b_j\hat\jmath + c_j\hat k$ we have $de_j = da_j\hat\imath + db_j\hat\jmath + dc_j\hat k$, so that $d^2e_j = 0$. In the $\{e_j\}$ basis, this reads:

$$0 = d^2e_j = \sum_k (d\omega^k_j - \sum_\ell \omega^\ell_j \wedge \omega^k_\ell)e_k \Rightarrow d\omega^k_j = \sum_{\ell=1}^3 \omega^\ell_j \wedge \omega^k_\ell.$$

For the last equations, note that in a parametrization $\sigma(u,v)$, we have $d\sigma = \sigma^*(\omega^1 e_1 + \omega^2 e_2)$, so that $d(\omega^1 e_1 + \omega^2 e_2) = 0$, or

$$0 = \sum_{a=1}^{2} (d\omega^a e_a - \omega^a \wedge de_a) = \sum_{a=1}^{2} \left((d\omega^a - \sum_{b=1}^{2} \omega^b \wedge \omega^a_b) e_a \right) - (\omega^1 \wedge \omega^3_1 + \omega^2 \wedge \omega^3_2) e_3$$

□

The computational apparatus of differential forms, in which Cartan's structural equations are given, leads to some efficient and concise formulations of our above coordinate computations. For example:

$$K\omega^1 \wedge \omega^2 = d\omega^1_2$$

is the statement of Gauss' teorema egregium. As well, for a curve $\gamma(t) \in \Sigma, \quad t \in [0,T]$ and $X_o \in T_{\gamma(0)}\Sigma$:

$$\Delta\theta = \theta(T) - \theta(0) = \int_\gamma \omega^1_2$$

where $\theta(t)$ is the angle between the tangent to $\gamma(t)$ and the parallel transport $X(t)$ of X_o along γ. In particular, for a closed loop, $\gamma = \partial D$, the general Stoke's theorem implies:

$$\int_D K \ dA = \Delta\theta$$

where $\Delta\theta$ is the *holonomy of parallel transport* around γ, ie the angle between X_o and X' where X' is the parallel transport of X_o around γ.

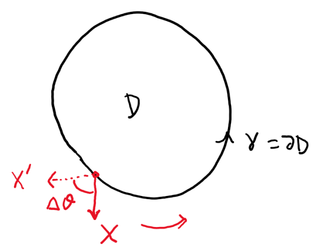

Figure 45. The Gaussian curvature is related to holonomy of parallel transport around loops. For $\gamma = \partial D$ the boundary of a region D, a vector X parallel transported around γ will in general return to a vector X' differing from X by some angle $\Delta\theta$. Then $\int_D K \ dA = \Delta\theta$ (taking care with orientations!).

Exercises:

1. For a surface of revolution: $\sigma(s,\theta) = (x(s)\cos\theta, x(s)\sin\theta, z(s))$ with $s \mapsto (x(s), 0, z(s))$ an arc-length parametrization of the generator ($x(s) > 0$), derive the expressions for the Gaussian and mean curvatures:
$$K = -\frac{x''}{x}, \quad 2H = -\kappa + \frac{z'}{x}$$
where $\kappa = z'x'' - z''x'$ is the curvature of the generator.

2. Describe the surfaces of revolution with constant Gaussian curvature.

3. Let $\sigma(u,v)$ be local coordinates on Σ, with unit normal $n = \frac{\sigma_u \times \sigma_v}{|\sigma_u \times \sigma_v|}$. Derive the Weingarten equations:
$$n_u = \frac{(FM - GL)\sigma_u + (FL - EM)\sigma_v}{EG - F^2}, \quad n_v = \frac{(FN - GM)\sigma_u + (FM - EN)\sigma_v}{EG - F^2}$$

4. In local coordinates, $\sigma(u,v)$, show that the asymptotic lines are given as integral curves[1] of:
$$L\ du^2 + 2M\ dudv + N\ dv^2 = 0.$$

5. In local coordinates, $\sigma(u,v)$, show that the lines of curvature are given as integral curves of:
$$(EM - FL)\ du^2 + (EN - GL)\ dudv + (FN - GM)\ dv^2 = 0.$$

6. Let $t \mapsto \gamma(t) \in \Sigma$ be a curve in the surface Σ. Show that if $\ddot\gamma(t) \in T_{\gamma(t)}\Sigma^\perp$ is normal to the surface then $\gamma(t)$ has constant speed ($|\dot\gamma(t)| = cst.$).

7. Suppose the surface $\Sigma \subset \mathbb{E}^3$ is compact and has no boundary. Show there exists a point $p \in \Sigma$ at which $K_p > 0$.

8. Let $X(t), Y(t)$ be the parallel transports of $X_o, Y_o \in T_{\gamma(0)}\Sigma$ along the curve $\gamma(t) \in \Sigma$. Show that $I(X(t), Y(t)) = I(X_o, Y_o)$, ie parallel transport preserves the first fundamental form.

9. Determine the holonomy of parallel transport along a latitude of the (unit) sphere.

[1] Meaning, eg, graphs $(u, v(u))$ with $v(u)$ satisfying: $L + 2M\frac{dv}{du} + N(\frac{dv}{du})^2 = 0$.

§9 some global results

We state a few 'famous' global results on surfaces.

Umbillic points: Suppose every point of the connected surface Σ is an umbillic point: $\kappa_1(p) = \kappa_2(p), \ \ \forall p \in \Sigma$. Then Σ is a subset of a fixed plane or sphere.

proof: Let $\sigma(u, v)$ be a parametrization with unit normal $n(u, v)$. Then –by assumption–the shape operator is diagonal with diagonal entries $\kappa_1(u, v) = \kappa_2(u, v) =: \kappa(u, v)$. That is:

$$\partial_u n = \kappa \ \partial_u \sigma, \quad \partial_v n = \kappa \ \partial_v \sigma.$$

Differentiating and setting $\partial_v \partial_u n = \partial_u \partial_v n$ yields:

$$\kappa_v \partial_u \sigma = \kappa_u \partial_v \sigma \Rightarrow \kappa_u = \kappa_v = 0$$

since $\partial_u \sigma, \partial_v \sigma$ are independent vectors. Hence the principal curvatures are constant. Now, if $\kappa \equiv 0$, then $n(u, v) \equiv n_o$ is constant, so that $\partial_u \sigma \cdot n_o = \partial_v \sigma \cdot n_o = 0 \Rightarrow \sigma(u, v) \cdot n_o = cst.$ and σ parametrizes some piece of a plane. Since Σ is connected, by covering Σ with charts, we have that Σ is contained in a fixed plane. Likewise, if $\kappa \neq 0$, then $\partial_u(n - \kappa\sigma) = \partial_v(n - \kappa\sigma) = 0$ so that $n = \kappa\sigma - p_o$ for some fixed p_o. Hence $|\sigma(u, v) - \frac{p_o}{\kappa}| = \frac{1}{|\kappa|} = cst.$ and $\sigma(u, v)$ parametrizes some piece of a sphere (center $\frac{p_o}{\kappa}$ and radius $\frac{1}{|\kappa|}$). The same covering argument and connectivity of Σ shows that Σ is contained in a fixed sphere. $\square$

Now we consider a classic local to global result, which may be thought of as generalizing our 'turning number' formulas for plane curves (pg. 7) to surfaces. First:

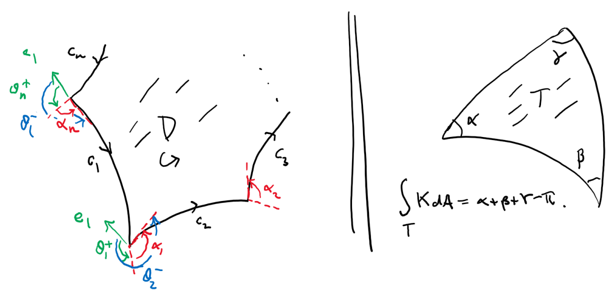

Figure 46. The local Gauss-Bonnet formula generalizes turning number formulas for plane curves. Applied to geodesic triangles (sides are geodesics), one obtains for example that triangles on positively curved surfaces have interior angle sums exceeding π.

Local Gauss-Bonnet: Let ∂D be a piecewise smooth curve in Σ bounding the simply connected region D with exterior angles α_j at its vertices. Then:

$$\int_D K \ dA + \int_{\partial D} \kappa_{geo} \ ds + \sum \alpha_j = 2\pi.$$

proof: We will present a derivation using this same notation of differential forms and Cartan's structural equations.[1] Let $c_j(s)$ be an arc-length parametrization of side j of ∂D, and write $c_j' = \cos\theta_j e_1 + \sin\theta_j e_2$. The geodesic curvature is then: $\kappa_{geo} = c_j'' \cdot Jc_j'$, where $Jc_j' := -\sin\theta_j e_1 + \cos\theta_j e_2$. Then (using $de_j = \omega_j^k e_k$)

$$c_j'' = \left(\theta_j' - \omega_2^1(c_j')\right) Jc_j' + (\cos\theta_j \ \omega_1^3(c_j') + \sin\theta_j \ \omega_2^3(c_j'))\nu$$

$$\Rightarrow \kappa_{geo} = \theta_j' - \omega_2^1(c_j').$$

[1] One may proceed without this formalism, using coordinate expressions. See eg, ch. 3 of Shifrin, or §6.3 of Klingenberg, or §4.5 of doCarmo. A good starting point is to take eg Fermi or geodesic polar coordinates: $(u, v) \mapsto \sigma(u, v)$ with $\sigma^* I = du^2 + G(u, v)dv^2$. Then one may use $e_1 := \partial_u \sigma, e_2 := \frac{\partial_v \sigma}{\sqrt{G}}$ for an orthonormal frame.

From $de_3 = d\nu = \omega_3^k e_k$, we have $K = \det d\nu = \omega_2^3 \wedge \omega_3^1(e_1, e_2) \Rightarrow K\omega^1 \wedge \omega^2 = \omega_2^3 \wedge \omega_3^1 = d\omega_2^1$. Hence:[1]

$$\int_D K\ dA + \int_{\partial D} \kappa_{geo}\ ds = \int_{\partial D} \kappa_{geo} + \omega_2^1(c_j')\ ds = \sum_j \theta_j^+ - \theta_j^-$$

$$= \theta_1^+ - \theta_2^- + \theta_2^+ - \theta_3^- + ... + \theta_n^+ - \theta_1^- = -\alpha_1 - \alpha_2 - ... - \alpha_{n-1} + 2\pi - \alpha_n.$$

□

This formula may be applied to a geodesic polygon: when ∂D consists of geodesic arcs. Then $\kappa_{geo} = 0$ over ∂D and the boundary term drops out. When the curvature, $K \equiv K_o$, is constant (eg on a sphere), one obtains interesting relations for example on geodesic triangles, for α, β, γ are interior angles of the geodesic triangle T:

$$K_o\ Area(T) = \alpha + \beta + \gamma - \pi.$$

Gauss-Bonnet theorem:[2] Let Σ be a compact oriented surface with boundary $\partial\Sigma$. Then:

$$\int_\Sigma K\ dA + \int_{\partial\Sigma} \kappa_{geo}\ ds = 2\pi\chi(\Sigma)$$

where $\chi(\Sigma)$ is the *Euler characteristic* of Σ.

proof: Triangulate $\Sigma = \sqcup T_j$ and apply the local Gauss-Bonnet formula to each face:

$$\int_{T_j} K\ dA + \int_{\partial T_j} \kappa_{geo}\ ds = \sum \beta_j - \pi$$

where β_j are the interior angles of the face T_j. Summing over the triangulation gives:

$$\int_\Sigma K\ dA + \int_{\partial\Sigma} \kappa_{geo}\ ds = \pi\left(2v^o + v^\partial - f\right)$$

where v^o are the number of interior vertices, v^∂ the number of boundary vertices and f the number of faces. For the boundary term, we have used that the triangles are oriented to agree with the orientation of Σ, and so the integrals over the edges of ∂T_j pairwise cancel *except* for those edges contained in the boundary $\partial\Sigma$. Note that every face contains 3 edges, and each interior edge is contained in two faces, while each boundary edge in one face so that $3f = 2e^o + e^\partial$, where e^o are the number of interior edges and e^∂ the number of edges along the boundary. Also, we have $v^\partial = e^\partial = f^\partial$ where e^∂ are the number of edges along $\partial\Sigma$ and f^∂ the number of faces with an edge along $\partial\Sigma$, so that:

$$2v^o + v^\partial - f = 2v - v^\partial - f + 3f - 3f = 2v + 2f - 2e^o - e^\partial - v^\partial = 2v + 2f - 2e$$

where $e = e^o + e^\partial$ are the total number of edges, $v = v^o + v^\partial$ the total number of vertices. The number $\chi(\Sigma) := v - e + f$ is called the *Euler characteristic* of the surface (it follows from this theorem that it does not depend on the triangulation chosen). □

For the next results, we will consider some properties related to the distance function on the surface. For $p, q \in \Sigma$, we take:

$$d(p, q) := \inf length(\gamma)$$

over curves $\gamma \subset \Sigma$ from p to q. For surfaces without boundary –as we have considered for the most part– we have locally around any point p a normal coordinate system, in which we see that the distance from p to sufficiently close points q to p is realized by geodesics from p to q. If Σ is not connected, then for p, q in different connected components one has $d(p, q) = \inf \emptyset = \infty$. We will focus on connected components of the surface –in fact it follows they are path connected– where the distance function is finite.

[1] One uses here the general Stoke's theorem: $\int_D K\ dA = \int_D d\omega_2^1 = \int_{\partial D} \omega_2^1$.

[2] For an interesting proof of this theorem, see the article of M. Levi here.

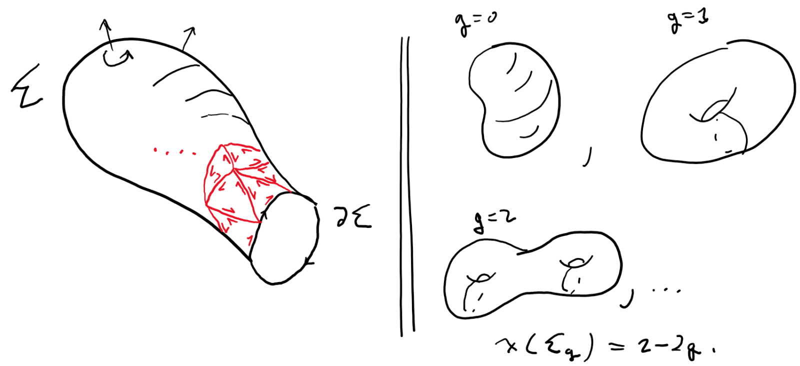

Figure 47. The Gauss-Bonnet theorem may be obtained from its local version applied to a triangulation of the surface – orienting the triangles to agree with that of the surface, the integrals over the interior edges pairwise cancel. Since the angles at each interior vertex sum to 2π and at each boundary vertex sum to π one has that $2\pi v^o + \pi v^\delta$ is the total sum of all interior angles to the triangulation. The compact surfaces without boundary, the *genus* g surfaces, have Euler characteristics $2-2g$.

It follows that with this distance function, Σ may be considered as a *metric space*, and it is in this sense that we may say the surface is *complete*[1]. Moreover,[2] the topology induced by this distance function agrees with the standard subspace topology on the surface as a subset of $\mathbb{E}^3$, and it is in this sense that one speaks of the surface being compact, connected, etc..

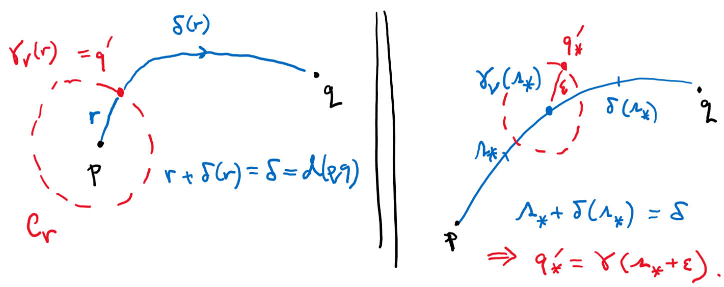

Figure 48. Normal coordinates may be used to show that on a complete surface, there exists a distance realizing geodesic joining any two given points p, q in the surface. One first takes normal coordinates around p and $\exp_p(rv) = q' \in C_r$ in a circle aorund p with $d(q', q) = \min_{x\in C_r} d(x, q)$. When C_r is in a normal coordinate chart (r sufficiently small) then $d(p, q') = r$. To show the geodesic $\gamma := \gamma_v$ goes from p to q and realizes the distance $\delta = d(p, q)$, one shows that $s + d(\gamma(s), q) = \delta$ is an open and closed condition for $s \in [0, \delta]$ so that $d(\gamma(\delta), q) = 0$. Namely, for $s_* + d(\gamma(s_*), q) = \delta$, one takes normal coordinates around $\gamma(s_*)$ containing a circle of radius $\varepsilon > 0$ and $q'_* \in C_\varepsilon(\gamma(s_*))$ with $d(q'_*, q) = \min_{x\in C_\varepsilon(\gamma(s_*))} d(x, q)$. Then $d(\gamma(s_*), q) = \varepsilon + d(q'_*, q) = \delta - s_*$ implies with triangle inequality that $d(p, q'_*) = s_* + \varepsilon$ is the length of the concactenated curve from $p \to \gamma(s_*), \gamma(s_*) \to q'_*$. Since distance realizing curves are smooth geodesics, one has $q'_* = \gamma(s_* + \varepsilon)$, and so $s_* + \varepsilon + d(\gamma(s_* + \varepsilon), q) = \delta$.

Hopf-Rinow theorem: Let Σ be a complete connected surface (without boundary). Then Σ is *geodesically complete*, meaning that any geodesic of Σ is defined for all time. Moreover, any two points of Σ may be joined by (at least one) distance realizing geodesic.

proof: See for example, §5.3 of doCarmo, for more details. The main tool is the normal (geodesic) coordinates we defined in the previous section. For example, to show geodesics are defined for all time, let $\gamma(t), t \in [a, b]$

[1] That is, every Cauchy sequence converges: a sequence of points $p_n \in \Sigma$ s.t. for any $\varepsilon > 0$, $d(p_n, p_m) < \varepsilon$ for $n, m > N$ has $d(p_n, p) \to 0$ for some $p \in \Sigma$.

[2] This is essentially according to our definition of surfaces –what are called *embedded* surfaces. Namely each point $p \in \Sigma$ has some local parametrization $\sigma : U \to im(\sigma) = V \cap \Sigma \ni p$ for $U \subset \mathbb{R}^2, V \subset \mathbb{R}^3$ some open sets. It follows that the local topology of $\mathbb{R}^2$ induced on Σ through charts agrees with the subspace topology on $\Sigma \subset \mathbb{E}^3$ induced from the standard topology of $\mathbb{E}^3$.

be some geodesic, with say $b < \infty$ and take normal coordinates around $\gamma(b)$ to see that γ may be extended to a geodesic defined for $t \in [a, b+\varepsilon]$ some $\varepsilon > 0$. Hence geodesics may always be extended and are defined for all time. Note that in particular this implies that the exponential map $\exp_p : T_p\Sigma \to \Sigma$ is defined over all of $T_p\Sigma$. Similarly, normal coordinates allow one to show that $d(p,q) = 0 \iff p = q$.
To see any two points $p, q \in \Sigma$ may be joined by a distance realizing geodesic –ie the exponential map is onto– consider a circle C_r around p with r sufficiently small so that $\exp_p$ is a (smooth) parametrization of the disk $D_r \subset \Sigma$. Since C_r is compact, we may take $q' \in C_r$ realizing $d(q', q) = \min_{x \in C_r} d(x, q)$. Since, by assumption, r was chosen small enough so that the exponential map parametrizes D_r, we have:

$$q' = \exp_p(rv)$$

for some unit vector $v \in T_p\Sigma$. Now we claim the unit speed geodesic, $\gamma_v(s)$, realizes the distance from p to q. Namely, let $\delta := d(p,q)$ so that we want to show $\gamma_v(\delta) = q$. Set $\delta(s) = d(\gamma_v(s), q)$, then:

$$\delta \leq r + \delta(r)$$

be triangle inequality. For any curve γ from p to q, let $x \in C_r$ be a point of intersection of γ with C_r. Then:

$$length(\gamma) \geq r + d(x,q) \geq r + \delta(r)$$

so that $\delta = \inf length(\gamma) \geq r + \delta(r)$ and so $\delta = r + \delta(r)$. The same argument shows $\delta = s + \delta(s)$ for $s \in [0, r]$. Since the distance function is continuous, the interval $I := \{s \in [0, \delta] : \delta = s + \delta(s)\}$ is closed (and non-empty, $I \supset [0, r]$). Let $s_* \in [0, \delta) \cap I$. Taking normal coordinates around $\gamma_v(s_*)$, one obtains

$$\delta = s_* + \varepsilon + \delta(s_* + \varepsilon) \Rightarrow \quad s_* + \varepsilon \in I$$

for some $\varepsilon > 0$. In particular, one has $\delta \in I \Rightarrow 0 = \delta(\delta) = d(\gamma_v(\delta), q) \Rightarrow q = \gamma_v(\delta)$ as desired. $\square$

Bonnet-Meyers theorem: Let Σ be a complete connected surface having Gaussian curvature bounded below by a positive constant: $K \geq K_o > 0$. Then Σ is compact.

proof: See for example, §5.4 of doCarmo, for more details. The main idea is that the signs of the 'second variation' or 'Hessian' along geodesics (see next section) determines whether the geodesics are minimizing (distance realizing). The signs of this second variation are related to the signs of the curvature, in particular one shows that when $K \geq K_o > 0$ then the second variation along *any* geodesic becomes negative –and so the geodesic is not distance realizing– when the length of the geodesic is greater than $\frac{\pi}{\sqrt{K_o}}$. Hence the *diameter* of Σ, $\delta := \sup_{p,q \in \Sigma} d(p,q) \leq \frac{\pi}{\sqrt{K_o}} < \infty$ is bounded. By the Hopf-Rinow theorem the exponential map $\exp_p : T_p\Sigma \to \Sigma$ is onto, and now the exponential map restricted to a compact disk, $\exp_p : D_p \to \Sigma$, in the tangent space is also onto. Since the exponential map is continuous, one obtains that Σ is compact. $\square$

Cartan-Hadamard theorem: Let Σ be a complete connected surface (without boundary) having non-positive Gaussian curvature, $K \leq 0$. Then for any $p \in \Sigma$, the exponential map:

$$\exp : T_p\Sigma \to \Sigma$$

is a covering map, in particular the universal cover of Σ is $\mathbb{R}^2$.

proof: See for example, §5.6 of doCarmo, for more details. By Hopf-Rinow, the exponential map is onto, and one computes when $K \leq 0$ that the exponential map is locally invertible around any point, so that it is a local diffeomorphism and covering map. $\square$

§10 calculus of variations

With surfaces, there are several interesting variational problems one may consider – either by varying the surface and considering functionals over surfaces, or by keeping the surface fixed and considering variational problems, eg for curves, contained in the surface.

First, we consider some variational properties of geodesics on a surface. We have already remarked that unparametrized geodesics may be characterized as extremals of the length functional on the surface. Analytically, these are curves admitting parametrizations with accelerations normal to the surface: $\gamma'' \in T_\gamma\Sigma^\perp$. Variationally, it is easier to work with the *energy* functional than the length functional:

$$\gamma \mapsto \int |\dot\gamma|^2 \ dt =: E(\gamma).$$

Extremals of the energy functional among curves in the surface connecting two given fixed points in a given time are exactly the geodesics parametrized by constant speed:

Energy extremals: Given $p, q \in \Sigma$ and $t_o < t_1 \in \mathbb{R}$, the extremals of $\gamma \mapsto \int_{t_o}^{t_1} \frac{1}{2}|\dot\gamma|^2 \ dt$ among curves in Σ with $\gamma(t_o) = p, \gamma(t_1) = q$ are constant speed geodesics from p to q. Moreover any extremal is smooth.

proof: Let γ_ε be a variation of a smooth extremal γ_* with $\eta = \frac{d}{d\varepsilon}|_{\varepsilon=0}\gamma_\varepsilon \in T_\gamma\Sigma$ a vector field along γ_*. Then integrating by parts:

$$0 = \int_{t_o}^{t_1} \dot\gamma_* \cdot \dot\eta \ dt = \int_{t_o}^{t_1} \ddot\gamma_* \cdot \eta \ dt$$

Since $\eta \in T_\gamma\Sigma$ is arbitrary, we have $\ddot\gamma_* \in T_\gamma\Sigma^\perp$. If an extremal is not smooth, eg piecewise smooth, in the integration by parts one obtains boundary terms $(\dot\gamma_*^+ - \dot\gamma_*^-) \cdot \eta$ summed over the non-smooth points. If one of these is non-zero, we may make the derivative of energy non-zero, eg taking $\eta = \dot\gamma_*^+ - \dot\gamma_*^-$. $\square$

In particular, it follows that any curve realizing the distance between two points is a minimum of the energy functional, and so in particular an extremal and thus a smooth parametrized geodesic between the points. We already know –from Fermi or normal coordinates– that *sufficiently short* geodesics realize distances between points. As for testing in general whether a given geodesic realizes distance between two points, one has analogues of the second derivative tests.

Namely, the *second variation* of the energy functional along a geodesic γ from p to q is defined by:

$$\eta \mapsto \frac{d^2}{d\varepsilon^2}E(\gamma_\varepsilon) =: \delta_\gamma^2 E(\eta)$$

where $\eta = \frac{d}{d\varepsilon}|_{\varepsilon=0}\gamma_\varepsilon$ is the variational vector field along γ by curves γ_ε from p to q. In particular, if the geodesic γ realizes distance, it minimizes E and so the second variation along γ must be non-negative. Thus if one can show the second variation is negative for some η along the geodesic (vanishing at the endpoints), then the geodesic does not realize distance.[1]

Second variation of energy: Let $\gamma(s)$ be a unit speed geodesic from $p = \gamma(0)$ to $q = \gamma(\ell)$ and γ_ε a variation of γ with the same fixed endpoints. Set $\eta = \frac{d}{d\varepsilon}|_{\varepsilon=0}\gamma_\varepsilon$. Then:

$$\delta_\gamma^2 E(\eta) = \frac{d^2}{d\varepsilon^2}|_{\varepsilon=0}E(\gamma_\varepsilon) = \int_0^\ell |\nabla_{\gamma'}\eta|^2 - K(s)|\eta^\perp|^2 \ ds$$

where $K(s)$ is the Gaussian curvature at $\gamma(s)$ and $\eta^\perp := \eta - (\eta \cdot \gamma')\ \gamma'$.

proof: One may consider $(s, \varepsilon) \mapsto \gamma_\varepsilon(s) =: \sigma(s, \varepsilon)$ as a local parametrization. Let $\eta_\varepsilon = \partial_\varepsilon\sigma = \frac{d}{d\varepsilon}\gamma_\varepsilon$ be the variational vector fields, and $\gamma_\varepsilon' = \partial_t\sigma$. Then:

$$\frac{d}{d\varepsilon}E(\gamma_\varepsilon) = \int_0^\ell \gamma_\varepsilon' \cdot \eta_\varepsilon' \ ds = \int_0^\ell \gamma_\varepsilon' \cdot \nabla_{\gamma_\varepsilon'}\eta_\varepsilon \ ds,$$

[1] A nice reference is Milnor's book *Morse theory*. See as well §5.4 of doCarmo.

$$\frac{d^2}{d\varepsilon^2}E(\gamma_\varepsilon) = \int_0^\ell |\nabla_{\gamma_\varepsilon'}\eta_\varepsilon|^2 + \gamma_\varepsilon' \cdot \nabla_{\eta_\varepsilon}\nabla_{\gamma_\varepsilon'}\eta_\varepsilon \ ds.$$

Using that the Gauss equations may be written $\gamma' \cdot (\nabla_{\gamma'}\nabla_\eta\eta - \nabla_\eta\nabla_{\gamma'}\eta) = |\eta^\perp|^2 K$, or $\gamma' \cdot \nabla_\eta\nabla_{\gamma'}\eta = \gamma' \cdot \nabla_{\gamma'}\nabla_\eta\eta - |\eta^\perp|^2 K$. We have, since γ is a geodesic:

$$\frac{d}{dt}(\gamma' \cdot \nabla_\eta\eta) = \gamma' \cdot \nabla_{\gamma'}\nabla_\eta\eta$$

so that integrating by parts gives:

$$\delta_\gamma^2 E(\eta) = \int_0^\ell |\nabla_{\gamma'}\eta|^2 - K(s)|\eta^\perp|^2 \ ds + (\gamma' \cdot \nabla_\eta\eta)|_0^\ell$$

Since the variation is by fixed endpoints, we have $\eta(0) = \eta(\ell) = 0$, and so the boundary term vanishes. □

One may make a further integration by parts of the first term using that $\frac{d}{dt}(\nabla_{\gamma'}\eta \cdot \eta) = \nabla_{\gamma'}\nabla_{\gamma'}\eta \cdot \eta + |\nabla_{\gamma'}\eta|^2$ to write the second variation as: [1]

$$\delta_\gamma^2 E(\eta) = -\int_0^\ell \left(\nabla_{\gamma'}\nabla_{\gamma'}\eta + K(s)\eta^\perp\right) \cdot \eta \ ds + \nabla_{\gamma'}\eta \cdot \eta|_0^\ell$$

and again, with the fixed endpoints, the boundary term vanishes. In this form, the ode $\nabla_{\gamma'}\nabla_{\gamma'}\eta + K(s)\eta^\perp = 0$ for η is called the Jacobi-equation, vector fields along the geodesic γ satisfying it being called *Jacobi fields.* One may further show that Jacobi fields are exactly the vector fields along γ obtained by taking $\frac{d}{d\varepsilon}|_{\varepsilon=0}\gamma_\varepsilon$ of a variation through geodesics, ie when each γ_ε is a geodesic.[2] The points p and q along a given geodesic γ are said to be *conjugate points* if there exists a (not identically zero) Jacobi field along γ which is zero at p and q. Geometrically, one interprets such points as joined by a family of geodesics (collecting to 'first order') –similarly to antipodal points on a sphere.

Conjugate points: Let $\gamma(s)$ be a unit speed geodesic with $\gamma(0)$ and $\gamma(\ell)$ conjugate points. Then γ does not realize distance from $\gamma(0)$ to $\gamma(\ell+\varepsilon)$ for any $\varepsilon > 0$.

proof: The principal observation is that any vector field along γ which is an extremal of the second variation is smooth. One sees this via the usual integration by parts trick. Suppose η_* is a vector field along γ that is an extremal of $\eta \mapsto \delta_\gamma^2 E(\eta)$. If η_* is say piecewise smooth than boundary terms $(\dot\eta^+ - \dot\eta^-) \cdot \delta\eta$ will appear which must be zero for an extremal. In particular, if $\gamma(0)$ and $\gamma(\ell)$ are conjugate, we consider the piecewise smooth vector field $\eta(s) = v(s), s \in [0,\ell], \eta(s) = 0, s \in [\ell, \ell+\varepsilon]$ with $v(s)$ the Jacobi-field vanishing at $\gamma(0), \gamma(\ell)$. Then we have $\delta_\gamma^2 E(\eta) = 0$, but η is not smooth, in particular cannot be a minimizer of $\delta_\gamma^2 E$ (any minimizer is an extremal so is smooth). Hence there must exist $\tilde\eta$ along γ vanishing at $\gamma(0), \gamma(\ell+\varepsilon)$ with $\delta_\gamma^2 E(\tilde\eta) < 0$. □

In other words, geodesics fail to realize distance –minimize– after their first conjugate point. The equation for Jacobi fields along a given geodesic is often written not in vector form, but rather in tangential and perpendicular components. Namely, write the vector field η along the unit speed geodesic γ as:

$$\eta(s) = \tau(s)\gamma'(s) + y(s)N(s)$$

where $N(s)$ is a unit tangent vector perpendicular to $\gamma'(s)$. Then γ' and N are parallel along γ, so that the Jacobi equation reads:

$$\tau'' = 0, \quad y'' = -K(s)y.$$

The second –normal component– is called a *Hill equation*, and the problem of determining conjugate points is an example of a *Sturm-Liouville problem.*

[1]The Bonnet-Meyers theorem follows from this formula by considering $\eta(s) = \sin(\frac{\pi}{\ell}s)N$ with N and γ' an orthonormal basis at each point. Then, when $K \geq K_o > 0$, $\delta_\gamma^2 E(\eta) \leq ((\frac{\pi}{\ell})^2 - K_o)\int_0^\ell \sin^2\frac{\pi s}{\ell} \ ds < 0$ for $\ell > \frac{\pi}{\sqrt{K_o}}$.

[2]To show such variations satisfy the Jacobi equation, one computes $\nabla_{\gamma_\varepsilon'}\nabla_{\gamma_\varepsilon'}\eta_\varepsilon = \nabla_{\gamma_\varepsilon'}\nabla_{\eta_\varepsilon}\gamma_\varepsilon' = -K\eta_\varepsilon^\perp + \nabla_{\eta_\varepsilon}\nabla_{\gamma_\varepsilon'}\gamma_\varepsilon'$ and uses that the curves are geodesics: $\nabla_{\gamma_\varepsilon'}\gamma_\varepsilon' = 0$.

Now, we consider some variational problems of surfaces as a way to see some more geometric meaning of the mean curvature. Namely we will consider the *area functional*, sending a surface to its total area,

$$\Sigma \mapsto Area(\Sigma) = \int_\Sigma dA.$$

One may ask given a certain class of surfaces, what are those which are extremals –critical points– of this functional, or minimize the total area in the given class.

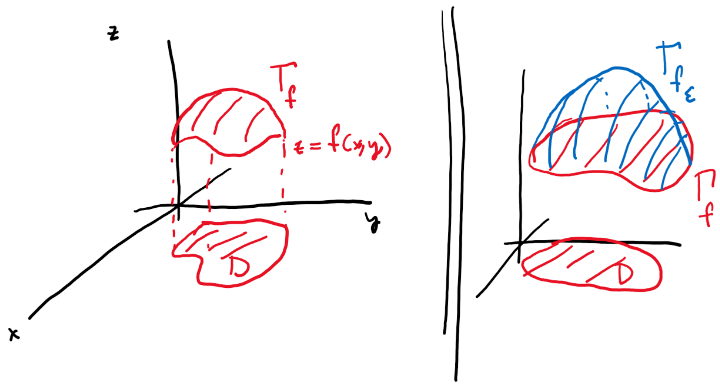

Figure 49. A graph Γ_f over a planar region D is a critical point of the area functional among surfaces with the same fixed boundary iff its mean curvature is zero. In fact one can show that such graphs minimize the area among surfaces with the given fixed boundary $\partial\Gamma_f = \{z = f|_{\partial D}\}$. Since every surface is locally a graph, one has that a minimal surface – one with mean curvature zero– locally minimizes area, ie for any point on a minimal surface there is a neighborhood U of that point s.t. any surface U' with boundary $\partial U' = \partial U$ has: $Area(U') \geq Area(U)$.

Consider first a surface which is a graph, $\Gamma_f := \{z = f(x, y)\}$ for $(x, y) \in D \subset \mathbb{R}^2$ a simply connected compact region (eg a closed disk). We will consider the class of surfaces given by graphs over D and with fixed boundary conditions. Over this class of surfaces, let us determine conditions for f to be a critical point of the area functional.

A 1-parameter family of such graphs is given by $f_\varepsilon : D \to \mathbb{R}$ with $f_o = f$ and $f_\varepsilon|_{\partial D} = f|_{\partial D}$. Then:

$$Area(\Gamma_{f_\varepsilon}) = \iint\limits_D \sqrt{1 + |\nabla f_\varepsilon|^2}\ dxdy \Rightarrow \frac{d}{d\varepsilon}|_{\varepsilon=0} Area(\Gamma_{f_\varepsilon}) = \iint\limits_D \frac{\nabla f \cdot \nabla \eta}{\sqrt{1 + |\nabla f|^2}}\ dxdy$$

where $\eta : D \to \mathbb{R}, (x, y) \mapsto \frac{d}{d\varepsilon}|_{\varepsilon=0} f_\varepsilon(x, y)$. Note that $\eta|_{\partial D} \equiv 0$ by the fixed boundary condition. Setting $X := \frac{\nabla f}{\sqrt{1+|\nabla f|^2}}$, we have: $div(\eta X) = X \cdot \nabla\eta + \eta div(X)$ so that

$$\frac{d}{d\varepsilon}|_{\varepsilon=0} Area(\Gamma_{f_\varepsilon}) = \iint\limits_D X \cdot \nabla\eta\ dxdy = \iint\limits_D -\eta\ div(X) + div(\eta X)\ dxdy.$$

Applying divergence theorem to the last term gives $\iint_D div(\eta X)\, dxdy = \int_{\partial D} \eta\, X \cdot \hat{n}\, ds = 0$ since $\eta|_{\partial D} = 0$. We have derived before a formula for the mean curvature, H, of a graph and one can check that: $2H = div(X)$. Hence, in summary, for the variation of area:[1]

$$\frac{d}{d\varepsilon}|_{\varepsilon=0} Area(\Gamma_{f_\varepsilon}) = -\iint\limits_D 2\eta H\ dxdy$$

where H is the mean curvature of f (oriented with 'upward' normal $\sim (-f_x, -f_y, 1)$).

[1] In slightly more generality, let $\eta : \Sigma \to \mathbb{R}$ be a function on an (oriented with unit normal ν) surface vanishing outside the compact set $D \subset \Sigma$. For any extension of the vector field $\eta\nu$ with flow φ_ε one has: $\frac{d}{d\varepsilon}|_{\varepsilon=0} Area(\varphi_\varepsilon(\Sigma)) = \int_\Sigma \mathscr{L}_{\eta\nu}(i_\nu\omega_{vol}) = \int_\Sigma \eta\ div(\nu)\ i_\nu\omega_{vol} = -\int_\Sigma 2\eta H\ dA$, where $i_\nu\omega_{vol} =: dA$ is the area form on Σ, and $div(\nu) = tr(d\nu) = -2H$.

The mean curvature may thus be interpreted as a density for ‘rate of area change’ of a surface when it is varied. In particular surfaces with zero mean curvature are called *minimal surfaces*. As every surface is locally a graph, sharpening our above computation yields:

Local minimizers: Let Σ be a minimal surface (so $H \equiv 0$). Then Σ *locally minimizes area.*

proof: Every surface is locally a graph, so consider $U = \Gamma_f$ for some open subset $U \subset \Sigma$ and $f : D \subset \mathbb{R}^2 \to \mathbb{R}$, with D simply connected and compact. Let U' be some other surface with the same boundary: $\partial U' = \partial U$. Consider the unit normal $\nu = \frac{(-f_x, -f_y, 1)}{\sqrt{1+|\nabla f|^2}}$ to U as a z-independent vector field defined over all of $\mathbb{R}^3$. Note that $div(\nu) = -tr(S) = -2H \equiv 0$. Now, the regions U and U' are the boundary of some region Ω, so that by divergence theorem:

$$0 = \int_\Omega div(\nu)\ dV = \int_U dA - \int_{U'} \nu \cdot d\vec{S}.$$

In words, because the divergence of ν is zero, the flux of ν across U is the same as that across U'. Moreover the flux of ν across U is the area of U, so that, with ν' unit normal to U' and dA' area element of U':

$$Area(U) = \int_{U'} \nu \cdot d\vec{S} \leq \int_{U'} |\nu \cdot \nu'|\ dA' \leq \int_{U'} dA' = Area(U')$$

as ν, ν' are unit vectors. $\square$

Surfaces of constant mean curvature may also be characterized variationally and are physically interesting as models of equilibrium configurations of ‘membranes’ or ‘soap films’ subject to a constant pressure difference.

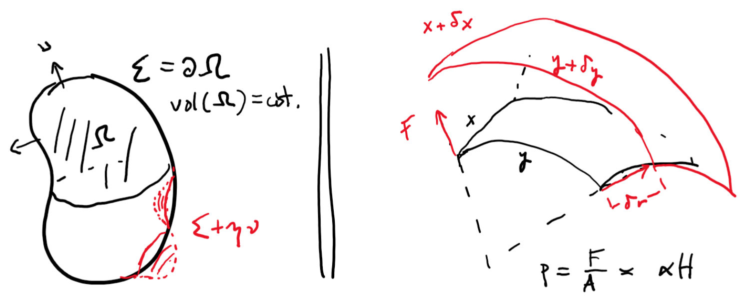

Figure 50. Closed surfaces with constant mean curvature may be described variationally as the extrema of total area among surfaces bounding regions of fixed total volume. Viewing the surface as an elastic membrane or soap film in equilibrium, the pressure at a given point is proportional to the mean curvature (see for example ch. 4 of J. Oprea’s *Differential geometry and its applications*).

Variationally, one considers the following problem: over compact surfaces bounding a region of fixed volume,[1] one seeks the extremals of the area functional, or the minimizers of the area functional,

$$\Sigma \mapsto Area(\Sigma), \quad \text{for } \ \Sigma \text{ s.t. } \Sigma = \partial\Omega, \quad vol(\Omega) = cst.$$

The extremals are characterized by the condition[2] that

$$0 = \int_\Sigma H\eta\ dA, \quad \forall \eta : \Sigma \to \mathbb{R} \text{ s.t. } \int_\Sigma \eta\ dA = 0$$

[1] Such surfaces are often called *closed* surfaces, meaning compact and with empty boundary.

[2] Let X be a vector field on $\mathbb{R}^3$ with $X|_\Sigma = \eta\nu$ for some $\eta : \Sigma \to \mathbb{R}$ and ν unit normal to Σ. For φ_ε the flow of X, we have $\frac{d}{d\varepsilon}|_{\varepsilon=0} Area(\varphi_\varepsilon(\Sigma)) = -\int_\Sigma 2\eta H\ dA$, while the condition that $\varphi_\varepsilon(\Sigma)$ contain a fixed volume imposes $0 = \int_\Omega div(X)\ dV = \int_\Sigma \eta\ dA$, by divergence theorem and $\Sigma = \partial\Omega$.

which is equivalent to the mean curvature, $H \equiv H_o = cst.$, being constant. To see this, take $c := \frac{\int_\Sigma H\ dA}{Area(\Sigma)}$ as the average value of H over Σ. Then:

$$0 = \int_\Sigma H - c\ dA \Rightarrow \int_\Sigma (H-c)^2\ dA = \int_\Sigma H(H-c)\ dA = 0$$

since by assumption $\int_\Sigma H\eta\ dA = 0$ for any $\int_\Sigma \eta\ dA = 0$, in particular with $\eta = H - c$. Now since $\int_\Sigma (H-c)^2\ dA = 0$, we must have $H - c \equiv 0$, ie $H \equiv c$ is constant.

We mention as well another well studied functional on surfaces, the *Willmore functional*:

$$\Sigma \mapsto \int_\Sigma H^2\ dA =: \mathscr{W}(\Sigma)$$

which may be thought of as an analogue of the elastic energy functional $\gamma \mapsto \int_\gamma \kappa^2\ ds$ on plane curves.

There are a number of interesting global results on these surfaces. For example:

Bernstein's theorem: If a global graph $z = f(x, y), (x, y) \in \mathbb{R}^2$ is a minimal surface, then it is a plane $(f(x, y) = ax + by + c)$.

Alexandrov's theorem: If a compact embedded surface (without boundary) has constant mean curvature then it is a sphere.

Wente's torus: There exist *immersed*[1] tori in $\mathbb{E}^3$ with constant mean curvature.

Catalan's theorem: If a minimal surface is *ruled* –every point contains a line– then it is part of a plane or helicoid[2].

Willmore estimates:[3] For Σ a compact immersed surface (without boundary), one has $\mathscr{W}(\Sigma) \geq 4\pi$ with equality iff Σ is a sphere. If Σ is an immersed torus, then $\mathscr{W}(\Sigma) \geq 2\pi^2$.

We mention as well that by our computations on surfaces of revolution (ex. # 1 pg. 47) and description of Delaunay roulettes we have that the only surfaces of revolution with constant mean curvature are obtained by revolving Delaunay roulettes (called the *catenoid, unduloid, or nodoid* by revolving the trace of a focus of a parabola, ellipse, or hyperbola resp.) as well as the limiting cases of cylinders, planes or spheres.

Finally the *Plateau problem* considers the existence of area minimizing surfaces spanning a given boundary curve and has led to important and interesting developements in analysis.

[1] In this course we have considered what are called embedded surfaces. Immersed surfaces admit local parametrizations but, for example, globally may admit self intersections.

[2] A helicoid surface may be parametrized by $(u \cos av, u \sin av, v), (u, v) \in \mathbb{R}^2$, with a constant.

[3] The estimate on tori (proposed in 1965 as the Willmore conjecture) was proved recently: F. Marques, A. Neves. *Min-max theory and the Willmore conjecture.* Annals of mathematics (2014): 683-782.

§11 examples

§12 remarks

Here we expand on some concepts used above.

Differentials: The *differential* or *Jacobian* of a differentiable[1] function $f : \mathbb{R}^n \to \mathbb{R}^m$ at a point $x_o \in \mathbb{R}^n$ is the linear transformation:

$$d_{x_o}f : \mathbb{R}^n \to \mathbb{R}^m, \quad v \mapsto \frac{d}{dt}|_{t=0} f(c(t))$$

where $c(t)$ is a curve in $\mathbb{R}^n$ with $c(0) = x_o,\ \dot{c}(0) = v$.

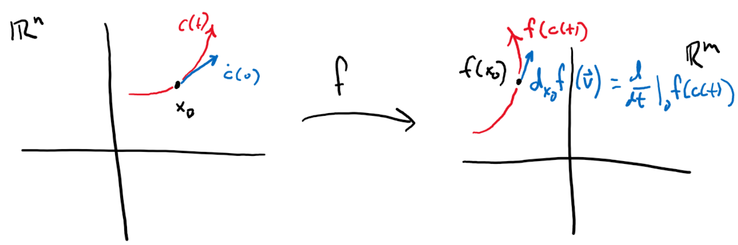

Figure 51. The differential, $d_{x_o}f$, of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ at a point x_o may be defined by acting on velocity vectors of curves through x_o.

In the standard bases, $e_1 = (1, 0, ..., 0), ..., e_n = (0, ..., 0, 1)$ and $E_1 = (1, 0, ..., 0), ..., E_m = (0, ..., 0, 1)$ of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, the Jacobian of $f(x_1, ..., x_n) = (f_1(x), ..., f_m(x))$ at $x_o \in \mathbb{R}^n$ is represented by the $m \times n$ matrix:

$$\left.\begin{pmatrix} \partial_{x_1} f_1 & \partial_{x_2} f_1 & \dots & \partial_{x_n} f_1 \\ \partial_{x_1} f_2 & \partial_{x_2} f_2 & \dots & \partial_{x_n} f_m \\ \vdots & \vdots & \dots & \vdots \\ \partial_{x_1} f_m & \partial_{x_2} f_m & \dots & \partial_{x_n} f_m \end{pmatrix}\right|_{x_o}$$

since $d_{x_o}f(e_j) = \frac{d}{dt}|_{t=0} f(x + te_j) = \partial_{x_j} f|_{x_o} = (\partial_{x_j} f_1, ..., \partial_{x_j} f_m)|_{x_o} = \sum_k \partial_{x_j} f_k|_{x_o} E_k$.

For parametrized surfaces, we consider differentiable maps $\sigma : \mathbb{R}^2 \to \mathbb{R}^3$ parametrizing patches of the surface. Writing $\sigma(u, v) = (\sigma_1(u, v), \sigma_2(u, v), \sigma_3(u, v))$ we have the matrix representation:

$$\left.\begin{pmatrix} \partial_u \sigma_1 & \partial_v \sigma_1 \\ \partial_u \sigma_2 & \partial_v \sigma_2 \\ \partial_u \sigma_3 & \partial_v \sigma_3 \end{pmatrix}\right|_{(u,v)}$$

for $d_{(u,v)}\sigma : \mathbb{R}^2 \to \mathbb{R}^3$. Or, in vector notation, that $\vec{v} = (v_1, v_2) \in \mathbb{R}^2$ is sent to $v_1 \partial_u \sigma + v_2 \partial_v \sigma$. For shorter notation – provided there is no risk of confusion about the base point where the differential is taken– one often uses the notation $\sigma_* = d_{(u,v)}\sigma$.

A fundamental property of differentials is their behaviour under composition, expressed via the *chain rule*, for $g : \mathbb{R}^\ell \to \mathbb{R}^m, f : \mathbb{R}^m \to \mathbb{R}^n$ we have:

$$d_{x_o}(f \circ g) = d_{g(x_o)}f \ d_{x_o}g.$$

We also make use at times above[2] –mostly without mention– of the *inverse function theorem* and some of its variants. These theorems relate properties of the linear map $d_{x_o}f : \mathbb{R}^n \to \mathbb{R}^m$ to local properties of the differentiable function $f : \mathbb{R}^n \to \mathbb{R}^m$ near x_o.

The inverse function theorem states that for a differentiable $f : \mathbb{R}^n \to \mathbb{R}^n$, if $d_{x_o}f : \mathbb{R}^n \to \mathbb{R}^n$ is an invertible linear map, then there exist open sets $U, V \subset \mathbb{R}^n$ containing x_o and $f(x_o)$ respectively such that $f|U : U \to V$ is invertible, and moreover the inverse $f^{-1} : V \to U$ is differentiable.

[1] The definition of differentiable guarantees the existence of this linear function: $\lim_{|v| \to 0} \frac{f(x_o + tv) - f(x_o) - d_{x_o}f(v)}{|v|} = 0$.

[2] For example to justify that Fermi coordinates or normal coordinates really give rise to local parametrizations of a surface.

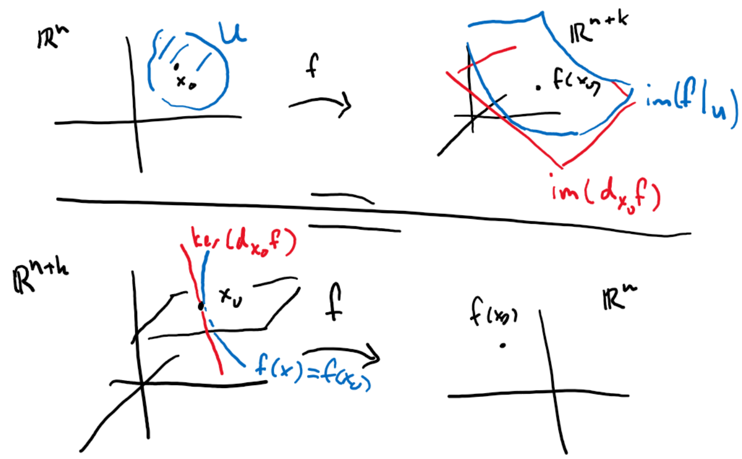

Figure 52. Locally, differentiable functions share certain properties with their differentials.

It follows from the inverse function several useful variants. For example, for a differentiable $f : \mathbb{R}^n \to \mathbb{R}^{n+k}$, if $d_{x_o}f : \mathbb{R}^n \to \mathbb{R}^{n+k}$ is an injective linear map, then there exists a neighborhood $U \subset \mathbb{R}^n$ of x_o such that $f|U$ is injective. Moreover there is on open set of $V \subset \mathbb{R}^{n+k}$ containing $f(x_o)$ and coordinates– $\varphi : V \to \mathbb{R}^n \times \mathbb{R}^k$ bijective and differentiable– such that $\varphi \circ f(x) = (x, 0)$. In particular it follows that for a differentiable map $\mathbb{R}^2 \to \mathbb{R}^3$, $(u, v) \mapsto \sigma(u, v)$ with $\partial_u\sigma, \partial_v\sigma$ independent, one always has that σ is locally 1-1 and $im(\sigma)$ may be locally realized as a graph over its tangent plane at each point.

We also mention that for a differentiable $f : \mathbb{R}^{n+k} \to \mathbb{R}^n$ with $d_{x_o}f$ an onto linear map then there exists neighborhoods $U \ni x_o$ and $V \ni f(x_o)$ and coordinates $\varphi : U \to \mathbb{R}^n \times \mathbb{R}^k, \psi : V \to \mathbb{R}^k$ such that $\psi \circ f \circ \varphi^{-1}(x, y) = y$.

Bilinear forms: Let V be an n-dimensional (real) vector space. A bilinear form on V is a map:

$$\beta : V \times V \to \mathbb{R}$$

such that for any fixed $v \in V$, the maps $V \to V$ by $w \mapsto \beta(w, v)$ or $w \mapsto \beta(v, w)$ are linear. A bilinear form is *symmetric* if $\beta(v, w) = \beta(w, v)$ for all $w, v \in V$.

A bilinear form may be represented in a basis as an $n \times n$ matrix. Let $v_1, ..., v_n$ be a basis for V. Then for $x = x_1v_1 + ... + x_nv_n, y = y_1v_1 + ... + y_nv_n$ we have:

$$\beta(x, y) = \sum_{i,j=1}^{n} \beta_{ij}x_iy_j, \quad \beta_{ij} := \beta(v_i, v_j)$$

or in matrix form:

$$\beta(x, y) = \begin{pmatrix} x_1 & ... & x_n \end{pmatrix} \begin{pmatrix} \beta_{11} & ... & \beta_{1n} \\ \vdots & ... & \vdots \\ \beta_{n1} & ... & \beta_{nn} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Bilinear forms are also often expressed using the notation of *tensors*. Note that the space of bilinear forms on V is itself a vector space (of dimension n^2), since for bilinear forms β_1, β_2 on V and $\lambda \in \mathbb{R}$, the map $(v, w) \mapsto \beta_1(v, w) + \lambda\beta_2(v, w)$ is again a bilinear form. Given a basis $v_1, ..., v_n$ for V, a basis for the space of bilinear forms is given by the bilinear forms $(x, y) \mapsto x_iy_j$ where $x = x_1v_1 + ... + x_nv_n, y = y_1v_1 + ... + y_nv_n$. We write these basis elements for the space of bilinear forms on V as:

$$v^i \otimes v^j : V \times V \to \mathbb{R}, \quad (x, y) \mapsto x_iy_j.$$

Then the matrix expression for β in the basis $v_1, .., v_n$ may be written as:

$$\beta = \sum_{i,j=1}^{n} \beta_{ij} v^i \otimes v^j.$$

The space of bilinear forms on V may also be called the 'tensor product of the dual space of V' and written $V^* \otimes V^*$. Recall that the *dual space* to V is the n-dimensional vector space consisting of linear maps:

$$v^* : V \to \mathbb{R}.$$

Two such linear maps v^*, u^* may be added and multiplied by scalars 'pointwise': $v^* + \lambda u^*$ is the linear map $V \to \mathbb{R},\ v \mapsto v^*(v) + \lambda u^*(v)$. Given a basis $v_1, ..., v_n$ of V there is an associated *dual basis*, $v^1, ..., v^n$ of V^* defined by:

$$v^j(x_1 v_1 + ... + x_n v_n) = x_j.$$

Our basis for bilinear forms, $v^i \otimes v^j$, induced by the basis $v_1, ..., v_n$ of V is then just given by 'pointwise multiplication' of the induced dual basis of V^*:

$$v^i \otimes v^j(x, y) = v^i(x) v^j(y) = x_i y_j.$$

On the plane $\mathbb{R}^2$ with standard basis $(1,0), (0,1)$, the associated dual basis is often written as dx, dy, meaning eg $dx : \mathbb{R}^2 \to \mathbb{R}, (x, y) \mapsto x$. A general bilinear form on $\mathbb{R}^2$ may then be written in tensor notation as:

$$\beta = \beta_{11} dx \otimes dx + \beta_{12} dx \otimes dy + \beta_{21} dy \otimes dx + \beta_{22} dy \otimes dy.$$

When β is symmetric, we have $\beta_{12} = \beta_{21}$, and one writes $dxdy := \frac{dx \otimes dy + dy \otimes dx}{2} = dydx, dx^2 = dx \otimes dx, dy^2 = dy \otimes dy$ so that:

$$\beta = \beta_{11}\ dx^2 + 2\beta_{12}\ dxdy + \beta_{22}\ dy^2.$$

Similarly, an anti-symmetric or *skew symmetric* bilinear form on $\mathbb{R}^2$, $\beta(v, w) = -\beta(w, v)$, may be written in tensor notation as:

$$\beta = \beta_{12}\ dx \wedge dy$$

where $dx \wedge dy := dx \otimes dy - dy \otimes dx$ (and $dx \wedge dx = dy \wedge dy = 0$).

Finally a bilinear form, β, induces linear maps $\beta_1, \beta_2 : V \to V^*$:

$$\beta_1(v) : V \to \mathbb{R}, \quad w \mapsto \beta(v, w) \qquad \beta_2(v) : V \to \mathbb{R}, \quad w \mapsto \beta(w, v).$$

When β is symmetric then $\beta_1 = \beta_2$.

An *inner product* on V is a symmetric bilinear form:

$$V \times V \to \mathbb{R}, \quad (v, w) \mapsto \langle v, w \rangle$$

such that $\langle v, v \rangle > 0$ whenever $v \neq 0$. An inner product induces an isomorphism (the 'musical isomorphism'):

$$V \xrightarrow{\flat} V^*, \quad v^\flat(w) := \langle v, w \rangle$$

since V and V^* are both of dimension n and $\ker(\flat) = \{0\}$. The inverse is written $V^* \xrightarrow{\sharp} V$.

Given an inner product on V, bilinear forms β on V may be identified with linear maps $B_1, B_2 : V \to V$, as $B_j := \sharp \circ \beta_j$ determined by: $\beta(v, w) = \langle B_1 v, w \rangle = \langle v, B_2 w \rangle$ for all $v, w \in V$. When β is symmetric, then one just has an associated $B = B_1 = B_2 : V \to V$ by

$$\beta(v, w) = \langle Bv, w \rangle.$$

Integrals: A *multiple integral*, eg $\iint_D f(x,y)\ dxdy$ of a function $f:\mathbb{R}^2\to\mathbb{R}$ over a planar region $D\subset\mathbb{R}^2$ geometrically represents the volume under the graph of f over D or a 'mass' of the region D with density f. One meets in multi-variable calculus a variety of similar integrals, eg 'line integrals', 'surface or flux integrals' etc.. Here we recall these integrals and describe the notation of differential forms.[1]

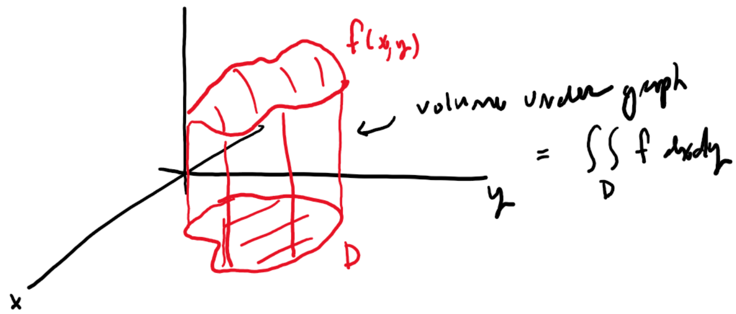

Figure 53. A multiple integral of f over a region D may be thought of as representing the mass of D with density f, or geometrically as a volume under the graph of f over D. Rigorously, this *Riemann integral* or *Darboux integral* is defined first over rectangular domains, $R=[x_o,x_1]\times[y_o,y_1]$. Let $\underline{\iint_R} f\ dxdy := \sup\{\sum m_{jk}(\xi_{j+1}-\xi_j)(\eta_{k+1}-\eta_k)\}$ taken over all partitions $\xi_o=x_o<\xi_1<\ldots<x_1=\xi_n,\ \eta_o=y_o<\eta_1<\ldots<y_1=\eta_n$ with $m_{jk}:=\inf_{(x,y)\in[\xi_j,\xi_{j+1}]\times[\eta_k,\eta_{k+1}]} f(x,y)$, and $\overline{\iint_R} f\ dxdy := \inf\{\sum M_{jk}(\xi_{j+1}-\xi_j)(\eta_{k+1}-\eta_k)\}$ with $M_{jk}:=\sup_{(x,y)\in[\xi_j,\xi_{j+1}]\times[\eta_k,\eta_{k+1}]} f(x,y)$. The function is said to be *Riemann integrable* over R when $\underline{\iint_R} f\ dxdy = \overline{\iint_R} f\ dxdy =: \iint_R f\ dxdy$. Every continuous function is Riemann integrable, and certain integrals may be evaluated by *Fubini's theorem* via finding anti-derivatives. Integrals over general domains may be defined via approximations by integrals over rectangular domains (the analytic aparatus behind the Riemann integral over general domains is that of *Jordan content* of planar regions). One obtains a more powerful and general integration theory by basing integration on the analytical apparatus of *Lebesgue measure*.

First, *line integrals* or 'work integrals', are certain integrals taken over curves. Namely, let $v:\mathbb{R}^3\to\mathbb{R}^3, p\mapsto (v_1(p),v_2(p),v_3(p))$ be a vector field on $\mathbb{R}^3$ and $\gamma\subset\mathbb{R}^3$ an (oriented) curve. Then:

$$\int_\gamma v_1\ dx + v_2\ dy + v_3\ dy = \int_\gamma v\cdot d\vec{s} := \int_{t_o}^{t_1} v(\gamma(t))\cdot\dot\gamma(t)\ dt$$

where $t\mapsto\gamma(t),\quad t\in[t_o,t_1]$ is a parametrization of γ. The value of the integral does not depend on the parametrization[2], provided the parametrization agrees with the orientation of the curve.

Next, *surface integrals* or 'flux integrals', are certain integrals taken over (oriented) surfaces. Namely, let $v:\mathbb{R}^3\to\mathbb{R}^3, p\mapsto(v_1(p),v_2(p),v_3(p))$ be a vector field on $\mathbb{R}^3$ and $\Sigma\subset\mathbb{R}^3$ an (oriented) surface. Then:

$$\int_\Sigma v_1\ dy\wedge dz + v_2\ dz\wedge dx + v_3\ dx\wedge dy = \int_\Sigma v\cdot d\vec{S} := \iint_D v(\sigma(u,v))\cdot(\sigma_u\times\sigma_v)\ dudv$$

where $D\ni(u,v)\mapsto\sigma(u,v)$ is a parametrization of Σ agreeing with the orientation.

The expressions we have written above in the 'integrands' are examples of differential forms, eg a '1-form' on $\mathbb{R}^3$ is the expression:

$$\omega := v_1\ dx + v_2\ dy + v_3\ dz$$

with $v_j:\mathbb{R}^3\to\mathbb{R}$. The notation may be related to our discussion of bilinear forms above. Recall that for the standard basis, $(1,0,0),(0,1,0),(0,0,1)$ of $\mathbb{R}^3$, it is customary to denote the resulting dual basis of $\mathbb{R}^{3*}$

[1]See for example ch. 4 §18 and ch. 7 of Arnold's book *Mathematical methods of classical mechanics*, or Bachman's *A geometric approach to differential forms*. Also, see these notes.

[2]This follows from the *change of variables formula*. In dimension 1 (line integrals): for $s\mapsto t(s)$ the change of variable, $\int_{t_o(s_o)}^{t_1(s_1)} f(t)\ dt = \int_{s_o}^{s_1} f(t(s))\ t'(s)\ ds$, while in general dimensions: for $y\mapsto\varphi(y)=x$ the change of variable, $\int_{\varphi(D)} f(x)\ dx = \int_D f(\varphi(y))\ \det(d_y\varphi)\ dy$.

as dx, dy, dz, eg $dx : \mathbb{R}^3 \to \mathbb{R}, (x, y, z) \mapsto x$. Then, the 1-form ω may be interpreted for each *fixed* $p \in \mathbb{R}^3$ as an element of $\mathbb{R}^{3*}$:

$$\omega_p := v_1(p)\ dx + v_2(p)\ dy + v_3(p)\ dy$$

Hence, line integrals are given in this notation by:

$$\int_\gamma \omega = \int_{t_o}^{t_1} \omega_{\gamma(t)}(\dot\gamma(t))\ dt.$$

Similarly, the '2-form', $\omega = v_1\ dy \wedge dz + v_2\ dz \wedge dx + v_3\ dx \wedge dy$ may be interpreted for each *fixed* $p \in \mathbb{R}^3$ as the skew-symmetric bilinear form:

$$\omega_p : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}, \quad (\vec a, \vec b) \mapsto v(p) \cdot (\vec a \times \vec b)$$

written in the basis $dy \wedge dz, dz \wedge dx, dx \wedge dy$ for skew-symmetric bilinear forms, eg $dx \wedge dy(\vec a, \vec b) = a_1 b_2 - a_2 b_1 = \hat k \cdot (\vec a \times \vec b)$. Surface integrals in this notation are then given by:

$$\int_\Sigma \omega = \iint\limits_D \omega_{\sigma(u,v)}(\sigma_u, \sigma_v)\ dudv.$$

Similarly, a '3-form', $\omega = \rho\ dx \wedge dy \wedge dz$ with $\rho : \mathbb{R}^3 \to \mathbb{R}$ may be interpreted for each *fixed* $p \in \mathbb{R}^3$ as the skew-symmetric map:

$$\omega_p : \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}, \quad (\vec a, \vec b, \vec c) \mapsto \rho(p)\ \vec c \cdot (\vec a \times \vec b) = \rho(p) \det(\vec a, \vec b, \vec c).$$

Triple integrals or 'oriented mass integrals' in this notation are then given by:

$$\int_\Omega \omega = \iiint\limits_\Phi \omega_{\varphi(u,v,w)}(\varphi_u, \varphi_v, \varphi_w)\ dudvdw$$

where $\varphi : \Phi \subset \mathbb{R}^3 \to \Omega \subset \mathbb{R}^3$ parametrizes the 3-dimensional region $\Omega \subset \mathbb{R}^3$.

In the integral calculus, there are number of integral formulas generalizing the fundamental theorem of calculus (eg Stoke's theorem, divergence theorem, etc.). In the classical notation these may be written:

$$\int_\gamma \nabla f \cdot d\vec s = f(\gamma(t_1)) - f(\gamma(t_o))$$

$$\int_{\partial\Sigma} v \cdot d\vec s = \int_\Sigma (\nabla \times v) \cdot d\vec S = \int_\Sigma (\nabla \times v) \cdot n\ dA$$

$$\int_{\partial\Omega} v \cdot d\vec S = \int_\Omega \nabla \cdot v\ dV$$

where $dA = |\sigma_u \times \sigma_v|\ dudv$ is called the area element of the surface (sometimes denoted dS) and $n = \frac{\sigma_u \times \sigma_v}{|\sigma_u \times \sigma_v|}$ the unit normal to the surface and $dV = dxdydz$ is called the volume element of $\mathbb{R}^3$. The operators $\nabla f = grad(f)$, $\nabla \times v = curl(v)$, and $\nabla \cdot v = div(v)$ are called the gradient of the function f, and the curl and divergence of the vector field v.

In the notation of differential forms, these formulas all may be considered as special cases of the *generalized Stoke's theorem*:

$$\int_{\partial V} \omega = \int_V d\omega$$

where the *exterior derivative*, $\omega \mapsto d\omega$, takes a k-form to a $(k+1)$-form. It may be defined by:

$$d\omega(v_1, ..., v_{k+1}) := \lim_{\varepsilon \to 0} \frac{\int_{\partial P_\varepsilon} \omega}{\varepsilon^{k+1}}$$

where P_ε is the parallelogram with sides $\varepsilon v_1, ..., \varepsilon v_{k+1}$. In coordinates, one has for 1-forms:

$$d(v_1\ dx + v_2\ dy + v_3\ dz) = dv_1 \wedge dx + dv_2 \wedge dy + dv_3 \wedge dz$$

where $dv_j = \partial_x v_j\ dx + \partial_y v_j\ dy + \partial_z v_j\ dz$. And for 2-forms:

$$d(v_1 dy \wedge dz + v_2 dz \wedge dx + v_3 dx \wedge dy) = dv_1 \wedge dy \wedge dz + dv_2 \wedge dz \wedge dx + dv_3 \wedge dx \wedge dy.$$

Under the correspondence of vector fields and differential forms:

$$\omega_v^1(\vec{a}) := v \cdot \vec{a}, \quad \omega_v^2(\vec{a}, \vec{b}) := v \cdot (\vec{a} \times \vec{b})$$

we have $\omega_{\nabla f}^1 = df$, $d(\omega_v^1) = \omega_{\nabla \times v}^2$, $d\omega_v^2 = (\nabla \cdot v)\ dx \wedge dy \wedge dz$ and may write the integral formulas above as:

$$\int_\gamma \omega_{\nabla f}^1 = f|_{\partial\gamma}$$

$$\int_{\partial\Sigma} \omega_v^1 = \int_\Sigma \omega_{\nabla \times v}^2$$

$$\int_{\partial\Omega} \omega_v^2 = \int_\Omega (\nabla \cdot v)\ dx \wedge dy \wedge dz.$$

Tensors, sum convention: We will restate some of our fundamental formulas for surfaces in the more efficient language of tensors and give coordinate expressions using the sum convention.[1]

We have already met several types of 'tensor fields' along a surface $\Sigma \subset \mathbb{E}^3$. First, a vector field, X, along Σ is a collection of tangent vectors $X_p \in T_p\Sigma$ for each $p \in \Sigma$. Similarly a 1-form, α, is a collection of dual vectors $\alpha_p : T_p\Sigma \to \mathbb{R}$, $\alpha_p \in T_p^*\Sigma$ for each $p \in \Sigma$. For example a function $f : \Sigma \to \mathbb{R}$ leads to a 1-form df by $d_p f : T_p\Sigma \to \mathbb{R}$, $v \mapsto \frac{d}{dt}|_{t=0} f(c(t))$ with $c(0) = p, \dot{c}(0) = v$.

The fundamental forms of Σ are also examples of tensor fields: for each $p \in \Sigma$ we have bilinear (symmetric) maps $T_p\Sigma \times T_p\Sigma \to \mathbb{R}$. As well the shape operator is for each $p \in \Sigma$ a linear transformation $T_p\Sigma \to T_p\Sigma$.

In general, a *tensor field* on Σ of type (r, s) is a collection of multilinear maps:[2]

$$\overbrace{T_p^*\Sigma \times ... \times T_p^*\Sigma}^{r-times} \times \overbrace{T_p\Sigma \times ... \times T_p\Sigma}^{s-times} \to \mathbb{R}$$

for each $p \in \Sigma$. For example, the fundamental forms are (symmetric) tensors of type (0,2). A 1-form is a tensor field of type (0,1) and a vector field may be viewed as a tensor field of type (1,0): we take $X_p : T_p^*\Sigma \to \mathbb{R}$, $\alpha_p \mapsto \alpha_p(X_p)$. The shape operator, $S_p : T_p\Sigma \to T_p\Sigma$, may be viewed as a type (1,1) tensor, by $T_p^*\Sigma \times T_p\Sigma \to \mathbb{R}$, $(\alpha_p, X_p) \mapsto \alpha_p(SX_p)$.

We denote the set of (r, s) tensor fields over Σ by $\mathfrak{T}_s^r$. The vector fields are commonly denoted by $\mathfrak{X}(\Sigma) := \mathfrak{T}_0^1$, and the k-forms –anti-symmetric $(0, k)$ tensors– are commonly denoted[3] $\Omega_k(\Sigma) \subset \mathfrak{T}_k^0$. It is also natural to regard type $(0, 0)$ tensor fields as functions on Σ, written by $\mathfrak{F}(\Sigma) = C^\infty(\Sigma) := \mathfrak{T}_0^0$.

Now, let us outline some fundamental types of 'derivatives' and operations that may be applied to tensors. We have seen in the previous section the *exterior derivative* which acts on differential forms:

$$d : \Omega_k(\Sigma) \to \Omega_{k+1}(\Sigma).$$

Vector fields and 1-forms are at each point dual to eachother, and so have a natural pairing to produce functions, denoted $(i_X\alpha)(p) := \alpha_p(X_p)$. This pairing may be generalized to give the *interior product*, acting on differential forms:

$$i_X : \Omega_k(\Sigma) \to \Omega_{k-1}(\Sigma), \quad (i_X\omega)(X_1, ..., X_{k-1}) := \omega(X, X_1, ..., X_{k-1}).$$

[1] A nice concise reference for this apparatus is ch. 2 of O'Neill's *Semi-Riemannian geometry*.

[2] In the same way we treated bilinear forms above, such multilinear maps of type (r, s) form a vector space denoted $\overbrace{T_p\Sigma \otimes ... \otimes T_p\Sigma}^{r-times} \otimes \overbrace{T_p^*\Sigma \otimes ... \otimes T_p^*\Sigma}^{s-times}$. In a basis v_j for $T_p\Sigma$ with dual basis v^j of $T_p^*\Sigma$, the elements $v_{j_1} \otimes ... \otimes v_{j_r} \otimes v^{i_1} \otimes ... \otimes v^{i_s}$ form a basis for these multilinear maps.

[3] For surfaces, one has $\Omega_k(\Sigma) = \{0\}$ when $k > 2$. For $k = 1$, there is no sense in anti-symmetric, and $\Omega_1(\Sigma) = \mathfrak{T}_1^0$, while for $k = 0$ one takes $\Omega_0(\Sigma) = \mathfrak{F}(\Sigma)$.

The interior derivative generalizes as the *contraction* operation:

$$C_i^j : \mathfrak{T}_s^r \to \mathfrak{T}_{s-1}^{r-1}, \quad (C_i^j\tau)(\alpha^1, ..., \alpha^{r-1}, X_1, ..., X_{s-1}) := tr(\tau|_{\alpha^l, X_m})$$

where $\tau|_{\alpha^l, X_m}$ is the (1,1) tensor $(\alpha, X) \mapsto \tau(\alpha^1, ..., \overset{j'th slot}{\alpha}, ..., \alpha^{r-1}, X_1, ..., \overset{i'th slot}{X}, ..., X_{s-1})$ (as we saw with the shape operator, (1,1) tensors may be viewed as linear maps $T_p\Sigma \to T_p\Sigma$, so their trace has sense). There is also the fundamental *tensor product* operation:

$$\mathfrak{T}_s^r \times \mathfrak{T}_{s'}^{r'} \to \mathfrak{T}_{s+s'}^{r+r'}, \quad (\tau, \tau') \mapsto \tau \otimes \tau'$$

given by $\tau \otimes \tau'(\alpha^1, ..., \alpha^{r+r'}, X_1, ..., X_{s+s'}) := \tau(\alpha^1, ..., \alpha^r, X_1, ..., X_s)\tau'(\alpha^{r+1}, ..., \alpha^{r+r'}, X_{s+1}, ..., X_{s+s'})$. Then for instance, $i_X\alpha = C_1^1(X \otimes \alpha)$.

The *Lie derivative* leads to a way to differentiate tensors along vector fields. First note that a vector field X on Σ has an associated *flow*, $\varphi_t : \Sigma \to \Sigma$, by taking $t \mapsto \varphi_t(p)$ to be an integral curve of X with initial condition p at $t = 0$. The Lie derivative of a function f along X is a function measuring the 'change of f along the flow of X':

$$\mathscr{L}_X f(p) := \frac{d}{dt}|_{t=0} f(\varphi_t(p)) = d_p f(X_p).$$

Likewise, the Lie derivative of another vector field Y along X is a vector field measuring the 'change of Y along the flow of X':

$$(\mathscr{L}_X Y)_p := \frac{d}{dt}|_{t=0} \varphi_{-t,*} Y_{\varphi_t(p)}.$$

It is customary to denote the Lie derivatives of vector fields with brackets:

$$[X, Y] := \mathscr{L}_X Y.$$

In general, the Lie derivative along a given vector field X operates on tensors by measuring 'change along the flow of X':

$$\mathscr{L}_X : \mathfrak{T}_s^r \to \mathfrak{T}_s^r$$

by $(\mathscr{L}_X\tau)_p(\alpha^1, ..., \alpha^r, X_1, ..., X_r) := \frac{d}{dt}|_{t=0}\tau_{\varphi_t(p)}(\alpha^1 \circ \varphi_{t,*}, ..., \alpha^r \circ \varphi_{t,*}, \varphi_{-t,*}X_1, ..., \varphi_{-t,*}X_s)$.[1]

The final type of derivative operation on tensors we will consider is the *covariant derivative*. Given *some* well-defined parallel transport on the surface, we may define a derivation of tensors analogous to Lie derivative measuring change of tensors under parallel transport. Namely, let X be a vector field on the surface with flow φ_t, and $\pi_t : T_p\Sigma \to T_{\varphi_t(p)}\Sigma$ the parallel transport along the curve $\varphi_t(p)$. Then we have:

$$\nabla_X : \mathfrak{T}_s^r \to \mathfrak{T}_s^r$$

$(\nabla_X\tau)_p(\alpha^1, ..., \alpha^r, X_1, ..., X_s) := \frac{d}{dt}|_{t=0}\tau_{\varphi_t(p)}(\alpha^1 \circ \pi_t, ..., \alpha^r \circ \pi_t, \pi_t X_1, ..., \pi_t X_s)$.[2] The parallel transports and associated covariant derivatives of interest allow one to differentiate vector fields along curves. Namely, for a given $v \in T_p\Sigma$ and vector field X, we have a well-defined vector $\nabla_v X_p \in T_p\Sigma$ determined by $\frac{d}{dt}|_{t=0}\pi_t^{-1}X_{c(t)}$ where $c(t)$ is *any* curve with $c(0) = p, \dot{c}(0) = v$. This is summarized in the general definition of an (affine) *connection* as a map:

$$\mathfrak{X}(\Sigma) \times \mathfrak{X}(\Sigma) \to \mathfrak{X}(\Sigma), \quad (X, Y) \mapsto \nabla_X Y$$

satisfying:

(i) $\nabla_{fX+Y}Z = f\nabla_X Z + \nabla_Y Z$,
(ii) $\nabla_X(fY + Z) = df(X)\, Y + f\nabla_X Y + \nabla_X Z$

[1] One may show that: $(\mathscr{L}_X\tau)(\alpha^1, ..., \alpha^r, X_1, ..., X_s) = \mathscr{L}_X(\tau(\alpha^1, ..., \alpha^r, X_1, ..., X_s)) - \sum \tau(\alpha^1, ..., \mathscr{L}_X\alpha^j, ..., \alpha^r, X_1, ..., X_s) - \sum \tau(\alpha^1, ..., \alpha^r, X_1, ..., \mathscr{L}_X X_k, ..., X_s)$. In particular, for a 1-form one has: $(\mathscr{L}_X\alpha)(Y) = \mathscr{L}_X(i_Y\alpha) - \alpha(\mathscr{L}_X Y)$.

[2] One may show that: $(\nabla_X\tau)(\alpha^1, ..., \alpha^r, X_1, ..., X_s) = \nabla_X(\tau(\alpha^1, ..., \alpha^r, X_1, ..., X_s)) - \sum \tau(\alpha^1, ..., \nabla_X\alpha^j, ..., \alpha^r, X_1, ..., X_s) - \sum \tau(\alpha^1, ..., \alpha^r, X_1, ..., \nabla_X X_k, ..., X_s)$. Where for a 1-form α, one has: $(\nabla_X\alpha)(Y) = \nabla_X(i_Y\alpha) - \alpha(\nabla_X Y)$ and for functions one has $\nabla_X f = \mathscr{L}_X f = df(X)$.

for any $X, Y, Z \in \mathfrak{X}(\Sigma)$ and $f \in \mathfrak{F}(\Sigma)$. Given such an affine connection, the associated parallel transport along curves is defined by requiring $\nabla_{\dot{c}} X = 0$.

Now given a parallel transport with associated affine connection we have for each $X \in \mathfrak{X}(\Sigma)$ the derivatives ∇_X. In fact –unlike for Lie derivatives– we may obtain here by property (i), a further derivation of tensors:

$$\nabla : \mathfrak{T}^r_s \to \mathfrak{T}^r_{s+1}$$

by $\nabla\tau(\alpha^1, ..., \alpha^r, X_1, ..., X_{s+1}) := (\nabla_{X_1}\tau)(\alpha^1, ..., \alpha^r, X_2, ..., X_{s+1})$.

For surfaces, the *Levi-Cevita connection* is determined by the first fundamental form, I, of the surface (through for example a rolling without slipping or twisting). In addition to the general properties (i), (ii) above, it satisfies:

metric compatibility: $\nabla I = 0$,
torsion free: $\nabla_X Y - \nabla_Y X = [X, Y]$.

In fact, the properties (i), (ii) along with the additional conditions of metric compatibility and torsion free determine the Levi-Cevita connection ∇, uniquely. The terminology of 'torsion' is that it is equivalent to the no twisting condition of the rolling, while the metric compatibility is equivalent to the no slipping condition. Torsion free'ness of a connection may also be seen to be equivalent to requiring that the induced Hessians of functions (see below) are symmetric bilinear forms.

Finally, the first fundamental form on a surface induces *musical isomorphisms* as we saw in the section on bilinear forms. Namely, we have:

$$\flat : \mathfrak{X}(\Sigma) \to \Omega_1(\Sigma), \quad X^\flat(Y) := I(X, Y)$$

with inverse $\sharp : \Omega_1(\Sigma) \to \mathfrak{X}(\Sigma)$. The musical isomorphisms may be extended to general tensor fields –the 'raising' or 'lowering' of indeces:[1]

$$\flat : \mathfrak{T}^r_s \to \mathfrak{T}^{r-1}_{s+1}, \quad \sharp : \mathfrak{T}^r_s \to \mathfrak{T}^{r+1}_{s-1}$$

by eg $\tau^\flat(\alpha^1, ..., \alpha^{r-1}, X_1, ..., X_{s+1}) := \tau(\alpha^1, ..., \alpha^{r-1}, X_1^\flat, X_2, ..., X_{s+1})$ and likewise $\tau^\sharp(\alpha^1, ..., \alpha^{r+1}, X_1, ..., X_{s-1}) := \tau(\alpha^1, ..., \alpha^r, (\alpha^{r+1})^\sharp, X_1, ..., X_{s-1})$.

These operators are the fundamental 'players' in developing a coordinate free way to write –and derive– equations describing the geometric structure on the surface. To work with these somewhat abstract definitions, one develops relations between them and coordinate expressions for their computation (for example in a course on manifolds or Riemannian geometry).

To finish and illustrate this notation, we will rewrite our main formulas derived above along with some new ones. We will write the corresponding coordinate expressions with the *sum convention*[2].

[1] As with the contraction operation, there are in general a number of choices for the musical isomorphims depending on which 'slot' is raised or lowered.

[2] This convention is that an index appearing as a superscript and a subscript is to be summed over (for surfaces from 1 to 2). It is most effective when one takes the coordinates as $(u^1, u^2) \mapsto \sigma(u^1, u^2)$. There is always danger of confusion with superscripts and exponents, however from context it is usually clear (for this reason some authors prefer to write $({}^1u, {}^2u)$ for the coordinates, although this is not standard). One also writes for short $\partial_i = \partial_i\sigma$, so that a vector field in coordinates is written as $X = X^j\partial_j$. Another convention commonly used is to denote partial derivatives of functions by commas, eg $\partial_j f = f_{,j}$, as well a matrix with components g_{ij} has the components of its inverse denoted by g^{ij}.

Fundamental forms, shape operator:

$$I \quad \longleftrightarrow \quad g_{ij}du^idu^j, \quad II \quad \longleftrightarrow \quad h_{ij}du^idu^j, \quad g_{ij} := I(\partial_i, \partial_j), \; h_{ij} := II(\partial_i, \partial_j)$$

$$I(Su, v) = II(u, v) \quad \longleftrightarrow \quad S_i^k g_{kj} = h_{ij}, \quad S_i^j = g^{jk} h_{ik}$$ [1]

Curvatures:

$$\det(S - \kappa_j id) = 0 \quad \longleftrightarrow \quad (h_{11} - \kappa_j g_{11})(h_{22} - \kappa_j g_{22}) = (h_{12} - \kappa_j g_{12})^2$$

$$K := \det(S), \quad H := \frac{1}{2}tr(S) \quad \longleftrightarrow \quad K = \frac{h_{11}h_{22} - h_{12}^2}{g_{11}g_{22} - g_{12}^2}, \quad H = \frac{g_{11}h_{22} - 2g_{12}h_{12} + g_{22}h_{11}}{2(g_{11}g_{22} - g_{12}^2)}$$

Levi-Cevita connection, parallel transport:

$$\nabla I = 0, \quad \nabla_X Y - \nabla_Y X = [X, Y] \quad \longleftrightarrow \quad \nabla_{\partial_i}\partial_j =: \Gamma_{ij}^k \partial_k, \quad \Gamma_{ij}^k = \frac{1}{2} g^{k\ell} \left(g_{j\ell,i} + g_{\ell i,j} - g_{ij,\ell} \right)$$

$$\nabla_{\dot\gamma} X = 0 \quad \longleftrightarrow \quad \dot X^k + X^i \dot u^j \Gamma_{ij}^k = 0$$

Geodesics:

$$\nabla_{\dot\gamma}\dot\gamma = 0 \quad \longleftrightarrow \quad \ddot u^k + \dot u^i \dot u^j \Gamma_{ij}^k = 0$$

Codazzi equations:[2]

$$\nabla S(X, Y) = \nabla S(Y, X) \quad \longleftrightarrow \quad h_{jk,i} - \Gamma_{ik}^\ell h_{j\ell} = h_{ik,j} - \Gamma_{jk}^\ell h_{i\ell}$$

Gauss equations:

$$\mathrm{Rm}(X, Y)Z := \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z \in \mathfrak{T}_3^1 \quad \longleftrightarrow \quad \mathrm{Rm}(\partial_i, \partial_j)\partial_k =: R_{ijk}^\ell \partial_\ell,$$ [3]

$$R_{ijk}^\ell = \Gamma_{jk,i}^\ell - \Gamma_{ik,j}^\ell + \Gamma_{jk}^m \Gamma_{im}^\ell - \Gamma_{ik}^m \Gamma_{jm}^\ell$$

$$I(\mathrm{Rm}(X, Y)Z, W) = II(Y, Z)II(X, W) - II(X, Z)II(Y, W) \quad \longleftrightarrow \quad R_{ijk}^m g_{m\ell} = h_{jk}h_{i\ell} - h_{ik}h_{j\ell}$$

Differential operators:[4]

$$grad(f) = \nabla f := (df)^\sharp \quad \longleftrightarrow \quad \nabla f = (g^{kj}\partial_j f)\partial_k$$

$$Hess(f) = \nabla(df) \in \mathfrak{T}_2^0 \quad \longleftrightarrow \quad Hess(f)(\partial_i, \partial_j) = \partial_i\partial_j f - \Gamma_{ij}^k \partial_k f$$ [5]

$$Hess(f)^\sharp = \nabla(grad(f)) \in \mathfrak{T}_1^1 \quad \longleftrightarrow \quad \partial_i \mapsto g^{jk}(\partial_i\partial_j f - \Gamma_{ij}^\ell \partial_\ell f)\partial_k$$

$$\nabla X \in \mathfrak{T}_1^1 \quad \longleftrightarrow \quad \partial_i \mapsto (\partial_i X^j + X^k \Gamma_{ik}^j)\partial_j$$

$$div(X) = tr(\nabla X) = C_1^1(\nabla X) \quad \longleftrightarrow \quad \partial_i X^i + X^k \Gamma_{ik}^i$$

$$\Delta f = tr(Hess(f)^\sharp) = div(\nabla f) \quad \longleftrightarrow \quad g^{ij}(\partial_i\partial_j f - \Gamma_{ij}^k \partial_k f)$$

---

[1] This coordinate expression for S contains the Weingarten equations: $\partial_j n = -S(\partial_j) = -S_j^k \partial_k = -g^{k\ell} h_{\ell j} \partial_k$. Note as well that we may rewrite the definition of S by $S^\flat = II$ or $S = II^\sharp$.

[2] These may also be written $\nabla_X(SY) - \nabla_Y(SX) = S([X, Y])$, or as $\nabla II(X, Y, Z) = \nabla II(Y, X, Z)$.

[3] This is the *Riemann curvature tensor*. It may be motivated geometrically by considering holonomy of parallel transport of Z around a parallelogram loop with sides spanned by $\varepsilon X, \varepsilon Y$ and taking a limit as $\varepsilon \to 0$.

[4] These differential operators have similar geometric interpretations on Σ as they do in multivariable calculus. For example, with φ_t the flow of X: $\frac{d}{dt}|_{t=0} \int_{\varphi_t(D)} \omega_{area} = \int_D div(X)\ \omega_{area}$. Which follows from the identity $d(i_X \omega_{area}) = div(X)\omega_{area}$, for ω_{area} the 2-form on Σ defined by taking the value 1 on I orthonormal pairs (in coordinates $\omega_{area} = \sqrt{\det(g)}du^1 \wedge du^2$).

[5] Geometrically, this bilinear form may be defined using the parallel transport by $Hess(f)_p(u, v) = \frac{d^2}{dsdt}|_{t=s=0} f(\gamma_t(s))$, where $\gamma(t)$ is a curve through p with initial velocity u and $s \mapsto \gamma_t(s)$ is a curve through $\gamma(t)$ with initial velocity the parallel transport of v along γ from p to $\gamma(t)$. It has the expression $Hess(f)(X, Y) = \mathscr{L}_X \mathscr{L}_Y f - \mathscr{L}_{\nabla_X Y} f$.

Exercises:

1. Let V be an n-dimensional (real) vector space with inner product $\langle\cdot,\cdot\rangle$, and β a symmetric bilinear form on V with $B : V \to V$ defined by $\langle Bv, w\rangle = \beta(v, w), \quad \forall v, w \in V$.

   (a) show that the critical values[1] of the function $V\backslash 0 \to \mathbb{R}, \quad v \mapsto \frac{\beta(v,v)}{\langle v,v\rangle}$ are the eigenvalues of B.

   (b) show the eigenvalues of B are real.

   (c) show the eigenvectors of B are orthogonal with respect to the inner product $\langle\cdot,\cdot\rangle$.

2. ...

[1] A critical value of a function, f, is its value, $f(v_o)$, at a critical point v_o. The critical points being those points, v_o, in the domain of the function at which $d_{v_o} f = 0$.

References

There are an abundance of good references on differential geometry. Here are a few main ones that you may want to check out to see more details or topics not covered in these notes.

- The book of M. do Carmo[1] is, for good reason, a standard.
- The lecture notes of T. Shifrin[2] (available online here). Good for examples and exercises.
- I also like the book of W. Klingenberg[3], for its conciseness.

Some other interesting references are the book of Breiskorn and Knörrer[4] which contains an interesting list of plane curves, the notes of Petrunin and Barrera[5] contain interesting exercises and 'metric' ideas, as well as the book of Oprea[6] for more examples and computer programs. The survey of Osserman[7] is a standard on minimal surfaces.

[1] M. Do Carmo, *Differential geometry of curves and surfaces.* Courier Dover Publications, 2016.
[2] T. Shifrin, *Differential geometry: a first course in curves and surfaces.* University of Georgia, 2015.
[3] W. Klingenberg, *A course in differential geometry.* Vol. 51. Springer Science & Business Media, 2013.
[4] E. Brieskorn, H. Knörrer. *Plane Algebraic Curves*, trans. J. Stillwell. Springer Science & Business Media, (2012).
[5] A. Petrunin, S. Barrera. *What is differential geometry: curves and surfaces.* arXiv:2012.11814 (2020). here.
[6] J. Oprea, *Differential geometry and its applications.* Vol. 59. American Mathematical Soc., (2019).
[7] R. Osserman, *A survey of minimal surfaces.* Courier Corporation, (2013).