This course is an introduction to the geometry of smooth curves and surfaces in Euclidean space $\mathbb{R}^n$ (in particular for $n = 2, 3$). The local shape of a curve or surface is described in terms of its *curvatures*. Many of the big theorems in the subject – such as the Gauss–Bonnet theorem, a highlight at the end of the semester – deal with integrals of curvature. Some of these integrals are topological constants, unchanged under deformation of the original curve or surface.

We will usually describe particular curves and surfaces locally via parametrizations, rather than, say, as level sets. Whereas in algebraic geometry, the unit circle is typically be described as the level set $x^2 + y^2 = 1$, we might instead parametrize it as $(\cos t, \sin t)$.

Of course, by *Euclidean space* [DE: *euklidischer Raum*] we mean the vector space $\mathbb{R}^n \ni x = (x_1, \ldots, x_n)$, equipped with with the standard *inner product* or *scalar product* [DE: *Skalarprodukt*] $\langle a, b\rangle = a \cdot b := \sum a_i b_i$ and its associated norm $|a| := \sqrt{\langle a, a\rangle}$.

A. CURVES

Given any interval $I \subset \mathbb{R}$, a continuous map $\alpha\colon I \to \mathbb{R}^n$ is called a *(parametrized) curve* [DE: *parametrisierte Kurve*] in $\mathbb{R}^n$. We write $\alpha(t) =: (\alpha_1(t), \ldots, \alpha_n(t))$.

We say α is C^k if it has continuous derivatives of order up to k. Here of course C^0 means nothing more than continuous, while C^1 is a minimal degree of smoothness, which is insufficient for many of our purposes. Indeed, for this course, rather than tracking which results require, say, C^2 or C^3 smoothness, we will use *smooth* [DE: *glatt*] to mean C^∞ and will typically assume that all of our curves are smooth.

Examples (parametrized on $I = \mathbb{R}$):

- $\alpha(t) := (a\cos t, a\sin t, bt)$ is a helix in $\mathbb{R}^3$ (for $a, b \neq 0$);
- $\beta(t) := (t^2, t^3)$ is a smooth parametrization of a plane curve with a cusp;
- $\gamma(t) := (\sin t, \sin 2t)$ is a figure-8 curve in $\mathbb{R}^2$;
- $\mu(t) := (t, t^2, \ldots, t^n)$ is called the *moment curve* in $\mathbb{R}^n$.

A *simple curve* [DE: *einfache Kurve*] is one where the map $\alpha\colon I \to \mathbb{R}^n$ is injective. A *closed curve* [DE: *geschlossene Kurve*] is one where $\alpha\colon \mathbb{R} \to \mathbb{R}^n$ is T-periodic for some $T > 0$, meaning $\alpha(t + T) = \alpha(T)$ for all $t \in \mathbb{R}$. Of course no closed curve is simple in the above sense; instead we define a *simple closed curve* [DE: *einfach geschlossene Kurve*] as a closed curve where α is injective on the half-open interval $[0, T)$.

A smooth (or even just C^1) curve α has a *velocity vector* [DE: *Geschwindigkeitsvektor*] $\dot\alpha(t) \in \mathbb{R}^n$ at each point. The fundamental theorem of calculus says this velocity can be integrated to give the displacement vector

$$\int_a^b \dot\alpha(t)\, dt = \alpha(b) - \alpha(a).$$

The *speed* [DE: *Bahngeschwindigkeit*] of α is $|\dot\alpha(t)| \geq 0$. We say α is *regular* [DE: *regulär*] if the speed is positive (that is, if the velocity never vanishes). Then the speed is a (smooth) positive function of t. (The cusped curve β above is not regular at $t = 0$; the other examples given are regular.)

The *length* [DE: *Länge*] of a smooth curve α is defined as $\operatorname{len}(\alpha) = \int_I |\dot\alpha(t)|\, dt$. (For a closed curve, of course, we should integrate from 0 to T instead of over the whole real line.) For any subinterval $[a, b] \subset I$, we see that

$$\int_a^b |\dot\alpha(t)|\, dt \geq \left|\int_a^b \dot\alpha(t)\, dt\right| = \big|\alpha(b) - \alpha(a)\big|.$$

This simply means that the length of any curve is at least the straight-line distance between its endpoints.

The length of an arbitrary curve can be defined (following Jordan) as its total variation:

$$\operatorname{len}(\alpha) := \operatorname{TV}(\alpha) := \sup_{t_0 < \cdots < t_n \in I} \sum_{i=1}^n \big|\alpha(t_i) - \alpha(t_{i-1})\big|.$$

This is the supremal length of inscribed polygons. (One can show this Jordan length is finite over finite intervals if and only if α has a Lipschitz reparametrization, e.g., by arclength. For a Lipschitz curve, the velocity is defined almost everywhere, so the integrals we used above – giving displacement as the integral of velocity and length as the integral of speed – exist in the sense of Lebesgue.)

If J is another interval and $\varphi\colon J \to I$ is an orientation-preserving homeomorphism, i.e., a strictly increasing surjection, then $\alpha\circ\varphi\colon J \to \mathbb{R}^n$ is a parametrized curve with the same image (or *trace*) as α, called a *reparametrization* of α. (Note that the *reverse* curve $\bar\alpha\colon -I \to \mathbb{R}^n$, defined by $\bar\alpha(t) := \alpha(-t)$, traces the same image in reverse order; this could be called an orientation-reversing reparametrization.)

When studying arbitrary continuous curves, it's sometimes helpful to allow more general reparametrizations via $\varphi\colon J \to I$ which is monotonic but not strictly monotonic. That is, we allow a reparametrization that stops at one point for a while – or that removes such a constant interval.

We instead focus on regular smooth curves α. Then if $\varphi\colon J \to I$ is a *diffeomorphism* [DE: *Diffeomorphismus*] (a smooth map with nonvanishing derivative, so that φ^{-1} is also smooth) then $\alpha \circ \varphi$ is again smooth and regular. We are interested in properties invariant under such *smooth reparametrizations*. Declaring a (regular smooth) curve to be equivalent to any smooth reparametrization, this gives an equivalence relation on the space of all parametrized curves. Formally, we could define an *unparametrized (smooth) curve* as an equivalence class. These are really the objects we want to study, but we do so implicitly, using parametrized curves and focusing on properties that are independent of parametrization, switching to a different parametrization when convenient.

For a fixed $t_0 \in I$ we define the arclength function $s(t) := \int_{t_0}^t |\dot\alpha(t)|\, dt$. Here s maps I to an interval J of length $\operatorname{len}(\alpha)$. If α is a regular smooth curve, then $s(t)$ is smooth, with positive derivative $\dot s = |\dot\alpha| > 0$ equal to the speed. Thus it has a smooth inverse function $\varphi\colon J \to I$. We say $\beta = \alpha \circ \varphi$ is the *arclength parametrization* [DE: *Parametrisierung nach Bogenlänge*] (or *unit-speed parametrization*) of α. We have $\beta(s) = \alpha(\varphi(s))$, so $\beta(s(t)) = \alpha(\varphi(s(t))) = \alpha(t)$. It follows that β has constant speed 1, and thus that the arclength of $\beta|_{[a,b]}$ is $b - a$.

The arclength parametetrization is hard to write down explicitly for most examples – we have to integrate a square root, then invert the resulting function. (There has been some work in computer-aided design on so-called "pythagorean hodograph curves", curves with rational parametrizations whose speed is also a rational function, with no square root. But this still doesn't get us all the way to a unit-speed parametrization.)

The fact that the arclength parametrization always exists, however, means that we can use it when proving theorems, and this is usually easiest. (Even when considering curves with less smoothness, e.g., C^k, there is a general principle that no regular parametrization is smoother than the arclength parametrization.)

Although for an arbitrary parameter we have used the name t (thinking of time) and written d/dt with a dot, when we use the arclength parametrization, we'll call the parameter s and write d/ds with a prime. Of course, for any function f along the curve, the chain rule says

$$\frac{df}{ds}\frac{ds}{dt} = \frac{df}{dt}, \qquad \text{i.e., } f' = \dot f/\dot s = \dot f/|\dot\alpha|.$$

Suppose now that α is a regular smooth unit-speed curve. Then its velocity α' is everywhere a unit vector, the (unit) *tangent vector* [DE: *Tangenten(einheits)vektor*] $T(s) := \alpha'(s)$ to the curve. (In terms of an arbitrary regular parametrization, we have of course $T = \dot\alpha/|\dot\alpha|$.)

`End of Lecture 8 Apr 2019`

We should best think of $T(s)$ as a vector based at $p = \alpha(s)$, perhaps as an arrow from p to $p + T(s)$, rather than as a point in $\mathbb{R}^n$. The *tangent line* to α at the point $p = \alpha(s)$ is the line $\{p + tT(s) : t \in \mathbb{R}\}$. (Of course, a nonsimple curve might pass through a point $p \in \mathbb{R}^n$ more than once, with different tangent vectors, so this language is not technically precise.)

Although we have agreed to consider mainly smooth (C^∞) curves, it is interesting to note that the tangent line is the the limit (as $h \to 0$) of secant lines through p and $\alpha(s + h)$, as long as α has a first derivative at s. If α is C^1 near s, then the tangent line is even the arbitrary limit of secant lines through $\alpha(s + h)$ and $\alpha(s + k)$.

While the velocity vectors of a curve depends on the parametrization, the tangent line and unit tangent vector do not; they are properties of an unparametrized curve. We are really most interested in properties that are also independent of rigid motion. It is not hard to show that a Euclidean motion of $\mathbb{R}^n$ is a rotation $A \in \mathrm{SO}(n)$ followed by a translation by some vector $v \in \mathbb{R}^n$: $x \mapsto Ax + v$. Thus α could be considered equivalent to $A\alpha + v\colon I \to \mathbb{R}^n$, $t \mapsto A\alpha(t) + v$. Of course, given any two lines in space, there is a rigid motion carrying one to the other. To find Euclidean invariants of curves, we need to take higher derivatives. We define the *curvature vector* [DE: *Krümmungsvektor*] $\vec\kappa := T' = \alpha''$; its length is the *curvature* [DE: *Krümmung*] $\kappa := |\vec\kappa|$.

Recall the Leibniz product rule for the scalar product: if v and w are vector-valued functions, then $(v\cdot w)' = v'\cdot w + v\cdot w'$. In particular, if $v \perp w$ (i.e., $v \cdot w \equiv 0$) then $v' \cdot w = -w' \cdot v$. And if $|v|$ is constant then $v' \perp v$. (Geometrically, this is just saying that the tangent plane to a sphere is perpendicular to the radius vector.) In particular, we have $\vec\kappa \perp T$.

Example: the circle $\alpha(t) = (r\cos t, r\sin t)$ of radius r (parametrized here with constant speed r) has

$$T = (-\sin t, \cos t), \qquad \vec\kappa = \frac{-1}{r}(\cos t, \sin t), \qquad \kappa \equiv 1/r.$$

Given regular smooth parametrization α with speed $\sigma := \dot s = |\dot\alpha|$, the velocity is σT, so the acceleration vector is

$$\ddot\alpha = (\sigma T)^{\cdot} = \dot\sigma T + \sigma\dot T = \dot\sigma T + \sigma^2 T' = \dot\sigma T + \sigma^2\vec\kappa.$$

Solving for $\vec\kappa$ we get the formula

$$|\dot\alpha|^2\,\vec\kappa = \ddot\alpha - \langle\ddot\alpha, T\rangle\, T = \ddot\alpha - \frac{\langle\ddot\alpha,\dot\alpha\rangle\,\dot\alpha}{|\dot\alpha|^2}$$

for the curvature of a curve not necessarily parametrized at unit speed.

Any three distinct points in $\mathbb{R}^n$ lie on a unique circle (or line). The osculating circle to α at p is the limit of such circles through three points along α approaching p. It is also the limit of circles tangent to α at p and passing through another point of the curve approaching p. Again, one can investigate the exact degree of smoothness required to have such limits exist.

Note that the second-order Taylor series for a unit-speed curve around the point $p = \alpha(0)$ (we assume without further comment that $0 \in I$) is:

$$\alpha(s) = p + sT(0) + \frac{s^2}{2}\vec\kappa(0) + O(s^3).$$

These first terms parametrize a parabola agreeing with α to second order (i.e., with the same tangent and curvature vector). Geometrically, it is nicer to use the *osculating circle* [DE: *Schmiegekreis*], the unique circle agreeing with α to second order at p (degenerating to a line if $\vec\kappa = 0$). It has radius $1/\kappa$ and center $p + \vec\kappa/\kappa^2$. Thus we can also write

$$\alpha(s) = p + \sin(\kappa s)T/\kappa + \cos(\kappa s)\vec\kappa/\kappa^2 + O(s^3).$$

(Here T and κ are constants, the unit tangent and curvature at $s = 0$.)

Comparing these two expansions, we note that a curve with constant acceleration or second derivative is a parabola; its points are equivalent by shearing. Geometrically more natural is a circle – a curve of constant curvature; its points are equivalent by a rotation, a rigid motion.)

Considering $T\colon I \to \mathbb{S}^{n-1} \subset \mathbb{R}^n$, we can think of this as another curve in $\mathbb{R}^n$ – called the *tantrix* (short for tangent indicatrix) of α – which happens to lie on the unit sphere. Assuming α was parametrized by arclength, the curve $s \mapsto T(s)$ has speed κ. Thus it is regular if and only if the curvature of α never vanishes. (Note on curves with nonvanishing curvature – in $\mathbb{R}^2$ versus $\mathbb{R}^3$.)

A1. Plane Curves

Now let's consider in particular plane curves ($n = 2$). We equip $\mathbb{R}^2$ with the standard orientation and let J denote the counterclockwise rotation by 90° so that $J(e_1) = e_2$ and for any vector v, $J(v)$ is the perpendicular vector of equal length such that $\{v, Jv\}$ is an oriented basis.

Given a (regular smooth) plane curve α, its (unit) *normal vector* [DE: *Normaleneinheitsvektor*] N is defined as $N(s) := J(T(s))$. Since $\vec{\kappa} = T'$ is perpendicular to T, it is a scalar multiple of N. Thus we can define the (signed) *curvature* κ_g of α by $\kappa_g N := \vec{\kappa}$ (so that $\kappa_g = \pm|\vec{\kappa}| = \pm\kappa$). For an arbitrary regular parametrization of α, we find

$$\kappa_g = \frac{\det(\dot\alpha, \ddot\alpha)}{|\dot\alpha|^3}.$$

From $N \perp T$ and $T' = \kappa_g N$, we see immediately that $N' = -\kappa_g T$. We can combine these equations as

$$\begin{pmatrix} T \\ N \end{pmatrix}' = \begin{pmatrix} 0 & \kappa_g \\ -\kappa_g & 0 \end{pmatrix} \begin{pmatrix} T \\ N \end{pmatrix}.$$

Rotating orthonormal frame, infinitesimal rotation (speed κ_g) given by skew-symmetric matrix. The curvature tells us how fast the tangent vector T turns as we move along the curve at unit speed.

Since $T(s)$ is a unit vector in the plane, it can be expressed as $(\cos\theta, \sin\theta)$ for some $\theta = \theta(s)$. Although θ is not uniquely determined (but only up to a multiple of 2π) we claim that we can make a smooth choice of θ along the whole curve. Indeed, if there is such a θ, its derivative is $\theta' = \kappa_g$. Picking any θ_0 such that $T(0) = (\cos\theta_0, \sin\theta_0)$ define $\theta(s) := \theta_0 + \int_0^s \kappa_g(s)\, ds$.

This lets us prove what is often called the *fundamental theorem of plane curves* [DE: *Hauptsatz der lokalen Kurventheorie*] (although it really doesn't seem quite that important): Given a smooth function $\kappa_g\colon I \to \mathbb{R}$ there exists a smooth unit-speed curve $\alpha\colon I \to \mathbb{R}$ with signed curvature κ_g; this curve is unique up to rigid motion. First note that integrating κ_g gives the angle function $\theta\colon I \to \mathbb{R}$ (uniquely up to a constant of integration), or equivalently gives the tangent vector $T = (\cos\theta, \sin\theta)$ (uniquely up to a rotation). Integrating T then gives α (uniquely up to a vector constant of integration, that is, up to a translation).

`End of Lecture 12 Apr 2019`

Now suppose α is a closed plane curve, that is, an L-periodic map $\mathbb{R} \to \mathbb{R}^2$. As above, by integrating the signed curvature, we get an angle function $\theta\colon \mathbb{R} \to \mathbb{R}$. This, however, is not necessarily periodic. Instead, $\theta(L) = \theta(0) + 2\pi n$ for some integer n called the *turning number* [DE: *Umlaufzahl*] (or rotation index or ...) of α. (It follows that $\theta(s+kL) - \theta(s) = kn$ for any integer k and any s.) Note that the *total signed curvature* of α is

$$\int_0^L \kappa_g\, ds = \theta(L) - \theta(0) = 2\pi n.$$

(If we reverse the orientation of α, we negate the signed curvature and the turning number.)

A famous result in topology is the *Jordan curve theorem* [DE: *Jordan'scher Kurvensatz*], saying that a simple closed plane curve divides the plane into two regions, one of which (called the *interior* [DE: *Innere*]) is bounded. Assuming the curve is oriented so that its interior is on the left, then the "theorem on turning tangents", more often known even in English by the German name *Umlaufsatz*, says that its turning number is always $+1$. Equivalently, the total signed curvature is 2π. (This is a special case of the Gauss–Bonnet theorem, needed as a lemma for the general case, so we will give a proof later.)

We will also later prove Fenchel's theorem that for any closed curve in $\mathbb{R}^n$, the total (unsigned) curvature satisfies $\int \kappa\, ds \geq 2\pi$ (with equality only for convex plane curves, where $\kappa = \kappa_g$).

A2. The Four-Vertex Theorem

A *vertex* [DE: *Scheitelpunkt*] of a smooth plane curve is an extremal point of κ_g, that is a point where κ_g achieves a local minimum or maximum, so that $\kappa_g' = 0$. Since any real-valued function on a compact set achieves a global minimum and maximum, any closed curve has at least two vertices.

The *Four-Vertex Theorem* [DE: *Vierscheitelsatz*] says that any simple closed plane curve α has at least four vertices. (Note counterexample $r = 1 + 2\sin\theta$ in polar coords if curve not embedded.) We give a proof due to Bob Osserman (1985).

Lemma A2.1. *Lemma: Given a nonempty compact set K in the plane (which might be the trace of a curve α) there is a unique smallest circle c enclosing K, called the* circumscibed circle *[DE:* Umkreis*]. (If K consists of a single point then c degenerates to that point.)*

Sketch of proof. By compactness, K is contained in the closed ball of radius R around the origin, for some $R > 0$. We are looking for the smallest closed ball containing K; it suffices to consider the compact family of balls of radius at most R centered at points within distance $2R$ of the origin. The balls containing K form a closed subfamily, thus also compact. Therefore the continuous radius function achieves its minimum and we have existence. Uniqueness follows almost immediately: If there were two minimal ballss containing K, then K would be contained in their intersection, which is contained in a strictly smaller ball. □

We can immediately derive several properties of the circumscribed circle:

1. c must touch K (for otherwise we could shrink c);

2. $c \cap K$ cannot lie in an open semicircle of c (for otherwise we could translate c to contradict (1);

3. $c \cap K$ thus contains at least two points, and if there are only two they are antipodal on c.

We next need to consider the relation between the curvatures of two tangent curves. So suppose α and β are two regular curves with the same tangent vector at p. If $\kappa_g^\alpha > \kappa_g^\beta$ at p

then α stays to the left of β in some neighborhood of p. A weak converse statement follows immediately by exchanging the curves: if α stays to the left of β, then at least the weak inequality $\kappa_g^\alpha \ge \kappa_g^\beta$ holds at p.

(Suppose p is a point along α where $\kappa_g' \ne 0$ – so p is not a vertex. Then that the osculating circle to α at p *crosses* α at p. Most people's sketches of osculating cicles are wrong!)

Now consider the circumcircle c of a closed cuve α. Since α stays inside (that is, to the left of) c, at any point $p \in c \cap \alpha$ the curvature of α is at least that of c.

Theorem A2.2 (Four-Vertex Theorem)**.** *Any simple closed plane curve α has at least four vertices.*

Proof. Let the curvature of the circumcircle c be k. If $c \cap \alpha$ includes an arc, there is nothing to prove. Otherwise suppose $c \cap \alpha$ includes at least $n \ge 2$ points p_i. (At these points $\kappa_g \ge i$.) We claim each arc α_i between consecutive p_i and p_{i+1} contains a point with $\kappa_g < k$. Along this arc, the minimum of κ_g is achieved at an interior point q_i, a vertex of α where $\kappa_g < k$. Now the arc from q_{i-1} to q_i includes the point p_i with $\kappa_g \ge k$, so the maximum of κ_g is achieved at an interior point p_i', a vertex of α with $\kappa_g \ge k$. Thus we have found $2n \ge 4$ vertices as desired.

To prove the claim, consider the one-parameter family of circular arcs from p_i to p_{i+1}, with signed curvatures decreasing from k. At the beginning, the arc α_i lies (weakly) to the left of the arc of c. As we decrease the curvature, there is a last circular arc c' that still touches the interior of α_i; it is tangent to α_i at at least one interior point. Since α_i stays to the right of c', its signed curvature is at most that of c', which is strictly less than k.

Where did we use the fact that the curve α is simple? (Recall that the theorem fails without this assumption!)

When two curves are tangent at p and don't cross locally, we got an inequality between their signed curvatures. But this assumes their orientations agree at p. By the Jordan curve theorem, a simple curve α bounds a compact region K. Clearly, α and K have the same circumcircle c. If both curves are oriented to have the compact regions to the left, then these orientations agree. Similarly, further application of the Jordan curve theorem ensure that the oriented circular arc from p_i to p_{i+1} used above agrees in orientation with α_i. □

`End of Lecture 15 Apr 2019`

A3. Evolutes and the Nesting Theorem

Given a curve $\alpha\colon I \to \mathbb{R}^n$ with nonvanishing curvature, its *evolute* [DE: *Evolute*] $\beta\colon I \to \mathbb{R}^n$ is the curve of centers of osculating circles: $\beta(t) := \alpha(t) + \vec\kappa(t)/\kappa(t)^2$. Let us consider in particular a unit-speed plane curve α with $\kappa = \kappa_g > 0$ and write $r = 1/\kappa$ for the radius of curvature. Then the evolute is $\beta(s) = \alpha(s) + r(s)N(s)$. Its velocity is $\beta' = T + r'N + rN' = r'N$, so its speed is $|r'(s)|$. (The evolute is singular where α has a vertex.) The acceleration of the evolute is $r''N + r'N' = r''N - r'T/r$, so its curvature is $\frac{1}{r|r'|}$. (Note that this approaches infinity as we approach a vertex of α – the evolute has a cusp.)

Now consider a planar arc α with strictly monotonic, nonvanishing curvature. By the formula above, its evolute also has nonvanishing curvature, so in particular, the distance $|\beta(s_2) - \beta(s_1)|$ is stricly less than the arclength $\int_{s_1}^{s_2} |r'(s)|\,ds = \left|\int r'\,ds\right| = \left|r(s_2) - r(s_1)\right|$. This simply says the distance between the centers of two osculating circles to α is less than the difference of their radii, that is, the circles are strictly nested. This is the *nesting theorem* of Tait (1896) and Kneser (1914): the osculating circles along a planar arc with stricly monotonic, nonvanishing curvature are strictly nested.

A4. The Isoperimetric Inequality

Another global result about plane curves is the *isoperimetric inequality* [DE: *isoperimetrische Ungleichung*]. If a simple closed curve of length L bounds a region of area A, then $4\pi A \le L^2$. (Equality holds only for a circle.)

If $R \subset \mathbb{R}^2$ is the region enclosed by the simple closed (C^1) curve $\alpha\colon [a,b] \to \mathbb{R}$, $\alpha(t) = (x(t), y(t))$, then we have by Green's theorem

$$A = \int_R dx\,dy = \int_\alpha x\,dy = \int_a^b x\dot y\,dt = -\int_a^b y\dot x\,dt.$$

(Actually, the formula gives an appropriately defined algebraic area even if the curve is not simple; no change if parametrization backtracks a bit.)

The trick suggested by Erhard Schmidt (1939) to prove the isoperimetric inequality is to consider an appropriate comparison circle. We deal with a smooth curve α. First find two parallel lines tangent to α such that α lies in the strip between them. Choose coordinates to make these the vertical lines $x = \pm r$. (Here $2r$ is the *width* of α in the given direction.) Parametrize α by arclength over $[0, L]$ by $(x(s), y(s))$ and parametrize the circle of radius r over $[0, L]$ by $\beta(s) = (x(s), \bar y(s))$: with the same function $x(s)$ as for α, and thus $\bar y(s) = \pm\sqrt{r^2 - x(s)^2}$. (Note about non-convex curves, etc.)

Note that the unit normal vector to α is $N = (-y', x')$, so $\langle N(s), \beta(s)\rangle = -xy' + \bar y x'$. We have $A = \int_0^L xy'\,ds$ and $\pi r^2 = -\int_0^L \bar y x'\,ds$. Thus

$$A + \pi r^2 = \int_0^L xy' - \bar y x'\,ds = \int_0^L \langle -N, \beta\rangle\,ds \le \int_0^L |N|\,|\beta|\,ds = Lr.$$

Thus by the arithmetic-geometric mean inequality,

$$\sqrt{A\pi r^2} \le (A + \pi r^2)/2 \le Lr/2.$$

Squaring and dividing by r^2 gives the isoperimetric inequality.

It is not hard to check that if all these inequalities hold with equality, then α must be a circle.

A5. The Cauchy–Crofton Formula

Given a unit vector $u = u(\theta) = (\cos\theta, \sin\theta) \in \mathbb{S}^1 \subset \mathbb{R}^2$, the orthogonal projection to the line in direction u is $\pi_u\colon \mathbb{R}^2 \to$

$\mathbb{R}^2$, $x \mapsto \langle x, u\rangle\, u$. If $\alpha\colon I \to \mathbb{R}^2$ is a smooth plane curve, then $\pi_u\alpha = \pi_u \circ \alpha$ is its projection (usually not regular!).

The Cauchy–Crofton formula says the length of α is $\pi/2$ time the average length of these projections. By average length we mean

$$\fint_{\mathbb{S}^1} \operatorname{len}(\pi_u\alpha)\, du = \fint_0^{2\pi} \operatorname{len}(\pi_{u(\theta)}\alpha)\, d\theta := \frac{1}{2\pi}\int_0^{2\pi} \operatorname{len}(\pi_{u(\theta)}\alpha)\, d\theta$$

`End of Lecture 26 Apr 2019`

To prove this, first note that if α is a line segment, the average projected length is independent of its position and orientation and proportional to its length. That is, the theorem holds for line segments with some constant c in place of $\pi/2$. (We could easily compute $c = \pi/2$ by integrating a trig function, but wait!) Next, by summing, it holds for all polygons (with the same c). Finally, it holds for smooth curves (or indeed for all rectifiable curves) by taking a limit of inscribed polygons. (To know we can switch the averaging integral with the limit of ever finer polygons, we can appeal for instance to Lebesgue's monotone convergence theorem.) To compute $c = \pi/2$ it is easiest to consider the unit circle α with length 2π and constant projection length 4.

Note that everything we have said also works for curves in $\mathbb{R}^n$ (projected to lines in different directions) – only the value of c will be different. Similarly, for an appropriate $c = c_{n,k}$ we get that the length of a curve in R^n is c times the average length of projections to all different k-dimensional subspaces.

For any *closed* plane curve α, the length of $\pi_u\alpha$ is at least twice the *width* of α in the direction u. If α is a convex plane curve, we have equality, so Cauchy–Crofton says the length is π times the average width. For instance any curve of constant width 1 (like the Reuleaux triangle on an equilateral triangle of side length 1, named after Franz Reuleaux, Rector at TU Berlin in the 1890s) has length π. A unit square has minimum width 1 and maximum width $\sqrt{2}$; since its length is 4, the average width is $4/\pi$.

Writing the various different lines perpendicular to u as $\ell_{u,h} := \{x : \langle x, u\rangle = h\}$ for $h \in \mathbb{R}$, we see that $\operatorname{len} \pi_u\alpha = \int_{\mathbb{R}} \#(\alpha \cap \ell_{u,h})\, dh$. Thus Cauchy–Crofton can be formulated as

$$\operatorname{len}\alpha = \frac{1}{4}\int_0^{2\pi}\int_{\mathbb{R}} \#(\alpha \cap \ell_{u(\theta),h})\, dh\, d\theta.$$

A6. Fenchel's theorem

Fenchel's theorem says the total curvature of any closed curve in $\mathbb{R}^n$ is at least 2π. (Equality holds only for convex plane curves.) To prove this, recall that the tantrix $T(s)$ has speed $\kappa(s)$ and thus its length is the total curvature of α. On the other hand, the tantrix lies in no open hemisphere of $\mathbb{S}^{n-1}$, for if we had $\langle T(s), u\rangle > 0$ for all s then we would get

$$\begin{aligned} 0 < \int_0^L \langle T(s), u\rangle\, ds &= \Big\langle u, \int_0^L T(s)\, ds\Big\rangle \\ &= \langle u, \alpha(L) - \alpha(0)\rangle = \langle u, 0\rangle = 0, \end{aligned}$$

a contradiction. Fenchel's theorem is thus an immediate corollary of the theorem below saying that a short spherical curve is contained in a spherical cap.

We will state all results for general n, but on first reading one should probably think of the case $n = 3$ where α lies on the usual unit sphere $\mathbb{S}^2 \subset \mathbb{R}^3$.

To investigate spherical curves in more detail note first that for points $A, A' \in \mathbb{S}^{n-1} \subset \mathbb{R}^n$ the spherical distance (the length of the shortest spherical path, a great circle arc) between them is

$$\rho(A, A') = \arccos\langle A, A'\rangle = 2\arcsin(|A - A'|/2) \le \pi.$$

The points are *antipodal* [DE: *antipodisch*] if $A = -A'$ (i.e., $\rho = \pi, \ldots$). A nonantipodal pair is connected by a unique shortest arc, with midpoint $M = (A + A')/|A + A'|$.

Lemma A6.1. *Suppose A, A' nonantipodal with midpoint M; suppose $\rho(X, M) < \pi/2$. Then $2\rho(X, M) \le \rho(X, A) + \rho(X, A')$.*

Note that this can be used to show that the distance from X to points along a great circle is a convex function, when resticted to the semicircle where the distance is at most $\pi 2$.

Proof. First note that A, A', X all lie in some three dimensional subspace of $\mathbb{R}^n$, so we work there, and in particular on $\mathbb{S}^2$. Consider a 2-fold rotation around M, taking A to A' and X to some point X'. Using the triangle inequality and the symmetry, we get

$$2\rho(X, M) = \rho(X, X') \le \rho(X, A) + \rho(A, X') = \rho(X, A) + \rho(X, A')$$

as desired. □

Theorem A6.2. *Suppose α is a closed curve on $\mathbb{S}^{n-1}$ of length $L < 2\pi$. Then α is contained in some spherical cap $\{x \in \mathbb{S}^{n-1} : \rho(X, M) \le L/4\}$ of (angular) radius $L/4 < \pi/2$, and in particular in some open hemisphere.*

Note that, as promised, Fenchel's theorem is an immediate corollary of this result.

Proof. Pick two points A, A' on α dividing the arclength in half. Then $\rho(A, A') \le L/2 < \pi$. Let M be the midpoint and let X be any point on α. If $\rho(X, M) < \pi/2$, then by the lemma,

$$\rho(X, M) \le (\rho(X, A) + \rho(X, A'))/2 \le \operatorname{len}(\alpha_{AXA'})/2 = L/4.$$

Thus the distance from M to any point on α is either at most $L/4$ or at least $\pi/2$. By continuity, the same possibility holds for all X; picking $X = A$ we see it is the first possibility. □

There are of course other approaches to proving Fenchel's theorem. One goes through an integral geometry formula analogous to our last version of Cauchy–Crofton. (We'll state it just for curves in $\mathbb{S}^2$ but it holds – with the same constant π – in any dimension.) For $u \in \mathbb{S}^2$, the great circle $u^\perp$ is the set of points orthogonal to u. Then the formula says the length of α equals π times the average number of intersections of α with these great circles. (When α itself is a great circle, this is clear, since there are always 2 intersections.)

First note that the length of a spherical curve is the limit of the lengths of spherical inscribed polygons (made of great circle arcs). (Indeed the spherical inscribed polygon always has length larger than the euclidean polygon with the same vertices, which is already approaching the length of the curve from below.) Then just as for Cauchy–Crofton, we check this formula first for great circle arcs, then for polygons and then by a (trickier) limiting argument for smooth curves.

With this formula, one can prove Fenchel's theorem for smooth curves by considering height functions $\langle\alpha(s), u\rangle$. Each has at least two critical points (min, max), but critical points satisfy $T(s) \in u^\perp$. That is, the tantrix intersects every great circle at least twice, and thus has length at least 2π.

Without giving precise definitions about knots, we can understand the Fáry–Milnor theorem: a nontrivally knotted curve in $\mathbb{R}^3$ has total curvature at least 4π. For suppose for some height function $\langle\alpha(s), u\rangle$ there was only one min and one max. At each intermediate height, there are exactly two points of α. Joining these pairs by horizontal segments gives an embedded disk spanning α, showing it is unknotted. For a knotted curve, every height function must have at least four critical points, meaning four intersections of the tantrix with every great circle.

`End of Lecture 29 Apr 2019`

A7. Schur's comparison theorem and Chakerian's packing theorem

Schur's theorem is a precise formulation of the intuitive idea that bending an arc more brings its endpoints closer together.

Suppose α is an arc in $\mathbb{R}^n$ of length L, and consider a comparison arc $\tilde\alpha$ in $\mathbb{R}^2 \subset \mathbb{R}^n$ of the same length, such that with respect to a common arclength parameter s, the curvature of $\tilde\alpha$ is positive and everywhere at least that of of $\tilde\alpha$: $\tilde\kappa(s) \ge \kappa(s)$. Assuming that $\tilde\alpha$ with its endpoints joined by a straight segment gives a convex (simple closed) curve, we conclude that its endpoints are closer:

$$\left|\alpha(L) - \alpha(0)\right| \ge \left|\tilde\alpha(L) - \tilde\alpha(0)\right|.$$

Proof: by convexity, we can find s_0 such that the tangent $T_0 := \tilde T(s_0$ to $\tilde\alpha$ is parallel to $\tilde\alpha(L) - \tilde\alpha(0)$. Move α by a rigid motion so that $\alpha(s_0) = \tilde\alpha(s_0)$ and they share the tangent vector T_0 there. We have

$$\left|\alpha(L) - \alpha(0)\right| \ge \langle\alpha(L) - \alpha(0), T_0\rangle = \int_0^L \langle T(s), T_0\rangle\, ds,$$

while for $\tilde\alpha$, our choice of T_0 gives equality:

$$\left|\tilde\alpha(L) - \tilde\alpha(0)\right| = \langle\tilde\alpha(L) - \tilde\alpha(0), T_0\rangle = \int_0^L \langle \tilde T(s), T_0\rangle\, ds,$$

Thus it suffices to show $\langle T(s), T_0\rangle \ge \langle\tilde T(s), T_0\rangle$ (for all s).

We start from s_0 (where both sides equal 1) and move out in either direction. While $\tilde T$ moves straight along a great circle with speed $\tilde\kappa$, a total distance less than π, we see that T moves at slower speed κ and perhaps not straight. Thus is geometrically clear that T is always closer to the starting direction. In formulas,

$$\langle\tilde T(s), T_0\rangle = \cos\int_{s_0}^s \tilde\kappa\, ds \le \cos\int_{s_0}^s \kappa\, ds \le \langle T(s), T_0\rangle.$$

(The last inequality follows since $\int\kappa\, ds$ is the length of the tantrix, while $\arccos\langle T(s), T_0\rangle$ is the distance between its endpoints.)

Note that this same proof can be made to work for arbitrary curves of finite total curvature. The case of polygonal curves is known as Cauchy's arm lemma and was used in his proof (1813) of the rigidity of convex polyhedra, although his proof of the lemma was not quite correct.

Chakerian proved the following packing result (which again can be generalized to all curves although we consider only smooth curves): A closed curve of length L in the unit ball in $\mathbb{R}^n$ has total curvature at least L. To check this, simply integrate by parts:

$$\operatorname{len}\alpha = \int\langle T, T\rangle\; ds = \int\langle -\alpha, \vec\kappa\rangle\; ds \le \int|\alpha|\,\kappa\, ds \le \int\kappa\, ds.$$

What about nonclosed curves? We just pick up a boundary term in the integration by parts, and find that length is at most total curvature plus 2.

A8. Framed space curves

We now specialize to consider curves in three-dimensional space $\mathbb{R}^3$. Just as for plane curves we used the 4-fold rotation J, in 3-space we will use its analog, the vector cross product. Recall that $v \times w = -w \times v$ is a vector perpendicular to both v and w.

A *framing*Rahmen along a smooth space curve α is a (smooth) choice of a unit normal vector $U(s)$ at each point $\alpha(s)$. Defining $V(s) := T(s) \times U(s)$ we have an (oriented) orthonormal frame $\{T, U, V\}$ for $\mathbb{R}^3$ at each point of the curve, and the idea is to follow how this frame rotates. As before, expressing the derivatives in the frame itself gives a skew-symmetric matrix:

$$\begin{pmatrix} T \\ U \\ V \end{pmatrix}' = \begin{pmatrix} 0 & \kappa_U & \kappa_V \\ -\kappa_U & 0 & \tau_U \\ -\kappa_V & -\tau_U & 0 \end{pmatrix} \begin{pmatrix} T \\ U \\ V \end{pmatrix}.$$

Here κ_U, κ_V and τ_U are functions along the curve which depend on the choice of framing. We see that $T' = \vec\kappa = \kappa_U U + \kappa_V V$, so these are just the components of the curvature vector in the chosen basis for the normal plane. (And $\kappa^2 = \kappa_U^2 + \kappa_V^2$.) The third function τ_U measures the twisting or torsion of the framing U.

Sometimes in physical problems a framing is given to us by material properties of a bent rod. Mathematically, the curve α might lie on a smooth surface in space; then we often choose U to be the surface normal so that the *conormal* V is (like T) tangent to the surface. (We will explore such *Darboux frames* [DE: *Darboux-Rahmen*] in detail when we study surfaces.)

But when no external framing is given to us, there are two ways to choose a nice framing such that one of the entries in the matrix above vanishes. The first has no twisting ($\tau_U = 0$), and such a $\{T, U, V\}$ is called a parallel frame or Bishop frame. Given any U_0 at $\alpha(s_0)$ we want U' to be purely tangential, indeed

$$U' = -\kappa_U T = -\langle\vec{\kappa}, U\rangle\, T.$$

But this ODE has a unique solution. Since it prescribes $U' \perp U$ the solution will have constant length, and since $\langle U', T\rangle = -\langle T', U\rangle$, the solution will stay normal to T. If we rotate a parallel framing by a constant angle φ in the normal plane (that is, replace U by $\cos\varphi U + \sin\varphi V$) then we get another parallel framing (corresponding to a different U_0). Indeed any two parallel framings differ by such a rotation. Parallel frames are very useful, for instance in computer graphics when drawing a tube around a curve. One disadvantage is that along a closed curve, a parallel framing will usually not close up.

The second special framing comes from prescribing $\kappa_V = 0$, i.e., $\vec{\kappa} = \kappa_U U$. That is, U should be the unit vector in the direction $\vec{\kappa}$. Here the disadvantage is that things only work nicely for curves of nonvanishing curvature $\kappa \neq 0$. Assuming this condition, we rename U as the *principal normal* [DE: *Hauptnormaleneinheitsvektor*] N and V as the *binormal* [DE: *Binormaleneinheitsvektor*] B and call $\{T, N, B\}$ the *Frenet frame* [DE: *Frenet-Rahmen*]. We have

$$\begin{pmatrix} T \\ N \\ B \end{pmatrix}' = \begin{pmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & -\tau & 0 \end{pmatrix}\begin{pmatrix} T \\ N \\ B \end{pmatrix},$$

where $\kappa(s)$ is the curvature and $\tau(s)$ is called the *torsion* [DE: *Torsion*] of α. In terms of a unit-speed parametrization, we have $\alpha' = T$, $\alpha'' = T' = \vec{\kappa} = \kappa N$, so $N = \vec{\kappa}/\kappa$. Finally, $N' = -\kappa T + \tau B$ so $\tau = \langle N', B\rangle = |N' + \kappa T|$. The expansion of the third derivative in the Frenet frame is

$$\alpha''' = (\kappa N)' = \kappa' N + \kappa N' = -\kappa^2 T + \kappa' N + \kappa\tau B.$$

Expressions in terms of an arbitrary parametrization of α with speed $\sigma(t)$ are left as an exercise. Here the nonvanishing curvature condition just says that $\dot\alpha$ and $\ddot\alpha$ are linearly independent, so that $\{\dot\alpha, \ddot\alpha, \dot\alpha\times\ddot\alpha\}$ is an oriented basis. The orthonormal frame $\{T, N, B\}$ is the result of applying the Gram–Schmidt process to this basis.

`End of Lecture 3 May 2019`

Of course N and B span the *normal plane* [DE: *Normalebene*] to α at $p = \alpha(s)$. The curve stays to second order in the *osculating plane* [DE: *Schmiegeebene*] spanned by T and N, which contains the osculating circle. The plane spanned by T and B is called the *rectifying plane* [DE: *Streckebene*] (since the projection of α to that plane has curvature vanishing at p).

The Taylor expansion of α to third order around $p = \alpha(0)$ is

$$\alpha(s) \approx p + \Big(s - \tfrac{s^3}{6}\kappa^2\Big)T + \Big(\tfrac{s^2}{2}\kappa + \tfrac{s^3}{6}\kappa'\Big)N + \Big(\tfrac{s^3}{6}\kappa\tau\Big)B$$

where of course T, N, B, κ, τ and κ' are all evaluated at $s = 0$. Exercise: look at the projections to the three planes above, and see which quadratic and cubic plane curves approximate them.

The "fundamental theorem of space curves" says that given functions $\kappa, \tau\colon I \to \mathbb{R}$ with $\kappa > 0$ determine a space curve (uniquely up to rigid motion) with that curvature and torsion. This is basically a standard theorem about existence and uniqueness of solutions to an ODE. For any given $\{T_0, N_0, B_0\}$ the matrix ODE above has a solution, which stays orthonormal and thus gives a framing. (Changing the initial condition just rotates the frames by a constant rotation.) As in the case of plane curves, integrating $T(s)$ recovers the curve α (uniquely up to translation).

Example: a curve with constant curvature and torsion is a helix. Its tantrix traces out a circle on $\mathbb{S}^2$ at constant speed κ. Any curve whose tantrix lies is a circle on $\mathbb{S}^2$ (i.e., makes constant angle with some fixed vector u) is called a *generalized helix*. Exercise: this condition is equivalent to τ/κ being constant.

Suppose now α is a unit-speed curve with $\kappa > 0$. If $\{T, N, B\}$ is the Frenet frame and $\{T, U, V\}$ is a parallel frame, then how are these related? We have of course

$$\begin{pmatrix} N \\ B \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} U \\ V \end{pmatrix}$$

for some $\theta = \theta(s)$. Then $\vec{\kappa} = \kappa N = \kappa\cos\theta\, U + \kappa\sin\theta\, V$ meaning that $\kappa_U = \kappa\cos\theta$ and $\kappa_V = \kappa\sin\theta$. Differentiating $B = -\sin\theta\, U + \cos\theta\, V$ gives

$$-\tau N = B' = -\theta'(\cos\theta\, U + \sin\theta\, V) + 0T = -\theta' N$$

so that $\theta' = \tau$ or $\theta = \int \tau\, ds$. (The constant of integration corresponds to the freedom to rotate the parallel frame.) We see that the twisting or torsion τ of the Frenet frame really does give the rate θ' at which it rotates relative to the twist-free Bishop frame. Sometimes it is useful to use a *complex curvature* $\kappa(s)e^{i\theta(s)} = \kappa_U(s) + i\kappa_V(s)$. Well defined up to global rotation by $e^{i\theta_0}$ in the complex plane (corresponding again to the freedom to rotate the parallel frame).

It is clear that a space curve lies in a plane if and only if $\tau \equiv 0$, if and only if θ is constant, if and only if the complex curvature stays on some fixed line through 0.

As another example, the complex curvature of a helix traces out the circle $|z| = \kappa$ at constant speed.

Bishop (1975) demonstrated the usefulness of the parallel frame by characterizing (C^2 regular) space curves that lie on some sphere. Indeed, α lies on a sphere of radius $1/d$ if and only if its complex curvature lies on a line at distance d from $0 \in \mathbb{C}$. In an appropriately rotated parallel frame, this line will be the line $\kappa_U \equiv d$. (The characterization in terms of the Frenet frame is more awkward, needing special treatment for points where τ and κ' vanish.)

To prove this, note that by translating and rescaling we can treat the case of $\alpha \subset \mathbb{S}^2$, i.e., $\langle\alpha, \alpha\rangle \equiv 1$. It follows that $\alpha \perp T$ so $U := -\alpha$ is a framing of α. Because $U' = -T$ is purely tangential, we see that this framing is parallel. That is, $U = -\alpha$, $V = \alpha \times T$ is a Bishop frame. The equation $U' = -T$ means $\kappa_U \equiv 1$, as desired. (Note that since the position vector on $\mathbb{S}^2$ is also the normal vector to the spherical surface, $\{T, U, V\}$ is also a Darboux frame for α as a curve in $\mathbb{S}^2$.) Conversely, suppose α has a parallel frame $\{T, U, V\}$ with $\kappa_U \equiv 1$, i.e., $U' = T$. Then $\alpha - U$ is a constant point P, meaning α lies on the unit sphere around P.

A9. Framings for curves in higher dimensions

A *framing* along a smooth curve α in $\mathbb{R}^n$ is a choice of oriented orthonormal basis $\{E_1, E_2, \dots, E_n\}$ at each point of α, where $E_1(s) = T(s)$ is the unit tangent vector, and each $E_i(s)$ is a smooth function. Of course the other E_i (for $i \geq 2$) are normal vectors. The infinitesimal rotation of any framing is given, as in the three-dimensional case, by a skew-symmetric matrix, here determined by the $\binom{n}{2}$ entries above the diagonal. Again it is helpful to choose special framings where only $n-1$ of these entries are nonzero.

In a parallel framing, these are the entries of the top row. That is, the curvature vector T' is an arbitrary combination $\sum \kappa_i E_i$ of the normal vectors E_i, but each of them is parallel with derivative $-\kappa_i T$ only in the tangent direction. Given any framing at an initial point, solving an ODE gives us a parallel frame along the curve.

The generalized Frenet frame exists only under the (somewhat restrictive) assumption that the first $n-1$ derivatives $\dot{\alpha}$, $\ddot{\alpha}$, …, $\alpha^{(n-1)}$ are linearly independent, and $\{T, E_2, \dots, E_n\}$ is then the Gram–Schmidt orthonormalization of these vectors. For this frame, it is only the matrix entries just above the diagonal that are nonzero. Thus

$$E_i' := \tau_i E_{i+1} - \tau_{i-1} E_{i-1}$$

In particular $T' = \tau_1 E_2$ so $\tau_1 = \kappa$ is the usual curvature and E_2 is the *principal normal* (the unit vector in the direction of $\vec{\kappa}$). The τ_i are called Frenet curvatures. A "fundamental theorem" says that for any functions $\tau_i(s)$ with $\tau_i > 0$ for $i < n-1$, there is a curve with these Frenet curvatures; it is unique up to rigid motion.

`End of Lecture 6 May 2019`

B. SURFACES

Given an open set $U \subset \mathbb{R}^m$ and a map $f\colon U \to \mathbb{R}^n$ we write $D_p f$ for the derivative of f at $p \in U$, the linear map $\mathbb{R}^k \to \mathbb{R}^n$ such that $f(p+v) \approx f(p) + D_p f(v)$. We say $p \in U$ is a *critical point* of f if $D_p f$ is not surjective (which is automatic if $m < n$); and we say $q = f(p) \in \mathbb{R}^n$ is a *critical value*. Any other $q \in \mathbb{R}^n$ is called a *regular value* of f. We say f is an *immersion* if $D_p f$ is injective at every p (which requires $m \le n$).

Intuitively, a subset $M \subset \mathbb{R}^n$ is a *smooth embedded k-dimensional submanifold* if every point $p \in M$ has an open neighborhood $U \subset \mathbb{R}^n$ in which M looks like an open set in $\mathbb{R}^k$. From analysis we recall several equivalent precise formulations:

1. Diffeomorphism: There is an open $V \subset \mathbb{R}^n$ and a diffeomorphism $\varphi\colon U \to V \subset \mathbb{R}^n$ taking $U \cap M$ to $V \cap (\mathbb{R}^k \times \{0\})$.
2. Level set: There is a smooth map $h\colon U \to \mathbb{R}^{n-k}$ such that 0 is a regular value of h and $U \cap M = h^{-1}(0)$.
3. Parametrization: There is an open set $V \subset \mathbb{R}^k$ and a smooth immersion $f\colon V \to \mathbb{R}^n$ that is a homeomorphism from V onto $U \cap M$.
4. Graph: There is an open set $V \subset \mathbb{R}^k$ and and a smooth map $h\colon V \to \mathbb{R}^{n-k}$ such that $U \cap M$ is the graph of h – up to permutation of coordinates in $\mathbb{R}^n$.

Here of course, a diffeomorphism gives a level set representation, and a graph is a special kind of parametrization.

For the rest of this semester, we will consider surfaces ($k = 2$) in $\mathbb{R}^3$. An example is the graph of a smooth function $f\colon U \to \mathbb{R}$, parametrized by $(u, v) \mapsto (u, v, f(u, v))$ or given as the zero-set of $F(x, y, z) := f(x, y) - z$.

Another example would be the unit sphere $\mathbb{S}^2$, the level set $x^2 + y^2 + z^2 = 1$. It can be covered by six open hemispheres on which it is a graph in one of the coordinate directions. Using stereographic projection we can parametrize all but a single point of the sphere by an immersion from $\mathbb{R}^2$. The usual geographic coordinates (lattitude and longitude) $(\varphi, \theta) \mapsto (\cos\theta\cos\varphi, \sin\theta\cos\varphi, \sin\varphi)$ give an immersion $(-\pi/2, \pi/2) \times \mathbb{R} \to \mathbb{R}^3$, which is injective if restricted to $\theta \in (-\pi, \pi)$.

We will typically use parametrizations to describe our surfaces. Let $\mathbf{x}\colon U \to \mathbb{R}^3$ be a smooth map defined on an open subset $U \subset \mathbb{R}^2$. At a point $(u, v) \in U$ we write

$$\mathbf{x}_u(u,v) = \frac{\partial \mathbf{x}}{\partial u} = D_{(u,v)}\mathbf{x}(\partial_u), \quad \mathbf{x}_v(u,v) = \frac{\partial \mathbf{x}}{\partial v} = D_{(u,v)}\mathbf{x}(\partial_v)$$

for the partial derivatives. (Here we are thinking of $D_{(u,v)}\mathbf{x}$ as a linear map on $T_{(u,v)}\mathbb{R}^2$; of course this is naturally isomorphic to $\mathbb{R}^2$, but it is helpful to think this way, since we will put a scalar product on it that depends on (u, v). Partly because the standard basis vectors for $T_{(u,v)}\mathbb{R}^2 = \mathbb{R}^2$ will not be orthonormal, we call them ∂_u, ∂_v rather than e_1, e_2.) The derivative $D_{(u,v)}\mathbf{x}$ is injective if and only if $\mathbf{x}_u$ and $\mathbf{x}_v$ are linearly independent (if and only if $\mathbf{x}_u \times \mathbf{x}_v \neq 0$). If this is true at all points of U, then $\mathbf{x}$ is an immersion.

Note that an immersion need not be one-to-one; even if it is, it need not be a homeomorphism onto its image (figure 8, spiral examples). If it is, then of course its image is a smooth submanifold. Given an immersion $\mathbf{x}$, any point in U has a neighborhood V such that $\mathbf{x}|_V$ is a homeomorphism onto its image. (We prove this as part of Lemma B2.1 below.)

Given an immersion $\mathbf{x}$ parametrizing a surface M, the span of $\mathbf{x}_u$ and $\mathbf{x}_v$ (the image of $D\mathbf{x}$) is two-dimensional, and is called the *tangent plane* T_pM at $p = \mathbf{x}(u, v)$. Orthogonal to this is the *normal line* N_pM, spanned by $\mathbf{x}_u \times \mathbf{x}_v$.

Note that we typically blur the distinction between a point $(u, v) \in U$ and its image $p = \mathbf{x}(u, v) \in M = \mathbf{x}(U)$. We write, for instance, $\mathbf{x}_u(p) = \mathbf{x}_u(u, v)$ interchangeably.

Example (surfaces of revolution): Suppose we have a regular curve $\alpha(t) = (r(t), 0, z(t))$ in the $x > 0$ half of the xz-plane. Consider the map

$$\mathbf{x}(t, \theta) = (r(t)\cos\theta, r(t)\sin\theta, z(t)).$$

This is an immersion (domain $I \times \mathbb{R}$, but injective only on smaller pieces), parametrizing a *surface of revolution*. Consider injectivity issues, tangent, normal, etc. – see homework.

Example (ruled surface): A surface swept out by straight lines (a *ruled surface*) can be parametrized by a base curve $\beta(t)$ and a director field $\delta(t)$ by setting $\mathbf{x}(t, u) = \beta(t) + u\delta(t)$.

`End of Lecture 10 May 2019`

B1. Curves, length and area

How do we describe a curve in a surface M? If $\mathbf{x}\colon U \to \mathbb{R}^3$ is a surface patch and $\alpha\colon I \to U \subset \mathbb{R}^2$ is a (regular smooth) curve in U then $\beta = \mathbf{x} \circ \alpha$ is a (regular smooth) curve in $\mathbb{R}^3$ lying on the surface M. Conversely, any curve on M can be described this way. We postpone a discussion of the details and of the effects of changing coordinates (to an overlapping surface patch).

Writing $\alpha(t) = (u(t), v(t))$ we have $\dot\alpha = (\dot u, \dot v)$ and by the chain rule the velocity vector $\dot\beta$ of $\beta = \mathbf{x} \circ \alpha$ is thus $\dot u\mathbf{x}_u + \dot v\mathbf{x}_v$. We see that the tangent plane T_pM spanned by $\{\mathbf{x}_u, \mathbf{x}_v\}$ is exactly the set of all velocity vectors to smooth curves in M through $p = \mathbf{x}(u, v)$.

The speed of β is of course given by the Euclidean norm of its velocity vector; the tangent space T_pM inherits an inner product $\langle\cdot,\cdot\rangle$ as a subspace of $T_p\mathbb{R}^3 = \mathbb{R}^3$. The basis $\{\mathbf{x}_u, \mathbf{x}_v\}$ is of course in most cases not orthonormal. The inner product is a symmetric bilinear form and is expressed with respect to this basis by the symmetric matrix

$$g_p := \begin{pmatrix} E & F \\ F & G \end{pmatrix} := \begin{pmatrix} \langle\mathbf{x}_u, \mathbf{x}_u\rangle & \langle\mathbf{x}_u, \mathbf{x}_v\rangle \\ \langle\mathbf{x}_v, \mathbf{x}_u\rangle & \langle\mathbf{x}_v, \mathbf{x}_v\rangle \end{pmatrix}.$$

(Here, E, F and G are traditional names for the entries of this matrix.) The matrix representation means that if $a = a_u\mathbf{x}_u + a_v\mathbf{x}_v$ and $b = b_u\mathbf{x}_u + b_v\mathbf{x}_v$ are two tangent vectors, then their inner product is

$$\langle a,b\rangle = \begin{pmatrix} a_u & a_v \end{pmatrix}\begin{pmatrix} \langle \mathbf{x}_u,\mathbf{x}_u\rangle & \langle \mathbf{x}_u,\mathbf{x}_v\rangle \\ \langle \mathbf{x}_v,\mathbf{x}_u\rangle & \langle \mathbf{x}_v,\mathbf{x}_v\rangle \end{pmatrix}\begin{pmatrix} b_u \\ b_v \end{pmatrix}.$$

Of course the associated quadratic form $a \mapsto \langle a,a\rangle$ is given by the same matrix. This is called the first fundamental form of the surface, and we use $g = g_p$ as a name for the matrix and for the bilinear/quadratic form. (Often $I = I_p$, from the roman numeral one, is used instead.)

Returning to the curve β, we get

$$|\dot\beta|^2 = g_p(\dot\beta) = \begin{pmatrix} \dot u & \dot v \end{pmatrix} g_p \begin{pmatrix} \dot u \\ \dot v \end{pmatrix} = E\dot u^2 + 2F\dot u\dot v + G\dot v^2.$$

The length of the curve β is of course then the integral of speed: $\operatorname{len}\beta = \int \sqrt{g(\dot\beta)}\,dt$.

Note that the velocity $(\dot u,\dot v)$ of α has the same expression (in the standard basis of $\mathbb{R}^2$) as the velocity of β (in our basis $\{\mathbf{x}_u,\mathbf{x}_v\}$). We often blur the distinction between $T_{(u,v)}U = T_{(u,v)}\mathbb{R}^2$ and T_pM, and that between α and β, etc. We can think, for instance, of g_p as defining a new inner product on $T_{(u,v)}\mathbb{R}^2$, whose matrix is $\left(\begin{smallmatrix} E & F \\ F & G \end{smallmatrix}\right)$ with respect to the standard basis $\{\partial_u,\partial_v\}$. (Technically, this inner product is the *pullback* of the inner product on T_pM under the linear map $D_{(u,v)}\mathbf{x}$.)

We can use the first fundamental form to measure not only length but also area. The parallelogram spanned by $\mathbf{x}_u$ and $\mathbf{x}_v$ has area $|\mathbf{x}_u\times\mathbf{x}_v|$ and we note

$$|\mathbf{x}_u\times\mathbf{x}_v|^2 = |\mathbf{x}_u|^2|\mathbf{x}_v|^2 - \langle \mathbf{x}_u,\mathbf{x}_v\rangle^2 = EG - F^2 = \det g.$$

The area of the surface patch is then

$$\int_U |\mathbf{x}_u\times\mathbf{x}_v|\,du\,dv = \int_U \sqrt{\det g}\,du\,dv.$$

Note also that a surface patch $\mathbf{x}(u,v)$ is regular (an immersion) if and only if $EG - F^2 = \det g$ is nonvanishing; this is often the easiest way to test the linear independence of $\mathbf{x}_u$ and $\mathbf{x}_v$.

Although it is easy to arrange that $\{\mathbf{x}_u,\mathbf{x}_v\}$ is an orthonormal basis – so that g is the identity matrix – at one given point of interest (say, $(0,0)\in U$), it is too much to hope that a general surface have a parametrization in which $\{\mathbf{x}_u,\mathbf{x}_v\}$ is an orthonormal basis everywhere. (We will later classify the "intrinsically flat" surfaces for which this is possible. As an example think of generalized cylinders – ruled surfaces with constant director δ.)

There are, however, various special classes of parametrizations which have some of the same advantages. We say a surface patch $\mathbf{x}$ is *orthogonal* if $\mathbf{x}_u \perp \mathbf{x}_v$, that is if $F = \langle \mathbf{x}_u,\mathbf{x}_v\rangle = 0$ or equivalently if $g = \left(\begin{smallmatrix} E & 0 \\ 0 & G \end{smallmatrix}\right)$ is a diagonal matrix. An orthogonal parametrization is *conformal* if $|\mathbf{x}_u| = |\mathbf{x}_v|$, that is, if $E = G$ or equivalently if g is a scalar multiple of the identity matrix. This means exactly that the map $\mathbf{x}\colon U \to \mathbb{R}^3$ preserves angles between tangent vectors (or equivalently between curves). It is known that any surface admits a conformal parametrization locally. (This is a version of the uniformization theorem from complex analysis.) Conformal coordinates are also called isothermal coordinates.

We have already mentioned the normal line N_pM spanned by $\mathbf{x}_u\times\mathbf{x}_v$. The parametrization $\mathbf{x}$ has an implicit orientation which allows us to pick out a *unit normal vector*

$$\boldsymbol{\nu} = \boldsymbol{\nu}_p := \frac{\mathbf{x}_u\times\mathbf{x}_v}{\left|\mathbf{x}_u\times\mathbf{x}_v\right|}.$$

Note that a different parametrization (like $y(u,v) := x(v,u)$) may give the opposite normal vector $\boldsymbol{\nu}$. Some surfaces are globally nonorientable, meaning that no continuous choice of $\boldsymbol{\nu}$ across the whole surface is possible.

B2. Smooth maps, change of parametrization, differentials

We usually talk about smoothness of maps defined on an open subset of $\mathbb{R}^m$. If $A \subset \mathbb{R}^m$ is an arbitrary subset, then a map $f\colon A \to \mathbb{R}^n$ is said to be *smooth* if it has a smooth extension $\bar f$ to some open $U \supset A$. (It suffices to check this locally in a neighborhood of each point. Standard properties – like the fact that the composition of two smooth maps is smooth – follow immediately.) In the case when A is a surface, we'd like to check that is the same as requiring smoothness in coordinates.

Lemma B2.1. *If* $\mathbf{x}\colon U \to M$ *is a regular parametrization then* $\mathbf{x}^{-1}\colon \mathbf{x}(U) \to U$ *is smooth. (Thus we say* $\mathbf{x}$ *is a diffeomorphism onto its image.)*

Proof. Assume (without loss of generality) that $(0,0)\in U$. We will check smoothness near $p = \mathbf{x}(0,0)$. Consider the function $\mathbf{y}\colon (t,u,v) \mapsto \mathbf{x}(u,v) + t\boldsymbol{\nu}_p$, $\mathbf{y}\colon \mathbb{R}\times U \to \mathbb{R}^3$. At the origin, its partial derivatives $(\boldsymbol{\nu}_p,\mathbf{x}_u,\mathbf{x}_v)$ are linearly independent. That is, $D_0\mathbf{y}$ is bijective. By the inverse function theorem, $\mathbf{y}$ is injective and has a smooth inverse on some neighborhood of $p = \mathbf{y}(0)$. But of course this inverse is locally the desired extension of $\mathbf{x}^{-1}$, showing that $\mathbf{x}^{-1}$ is smooth. □

Suppose $M \subset \mathbb{R}^3$ is a surface parametrized by $\mathbf{x}\colon U \to M = \mathbf{x}(U)$. The fact that $\mathbf{x}$ and $\mathbf{x}^{-1}$ are both smooth immediately shows:

1. $f\colon M \to \mathbb{R}^n$ is smooth $\iff$ $f\circ\mathbf{x}$ is smooth.
2. $f\colon \mathbb{R}^n \to M$ is smooth $\iff$ $\mathbf{x}^{-1}\circ f\colon \mathbb{R}^n \to \mathbb{R}^2$ is smooth.

Combining these facts, if $N \subset \mathbb{R}^3$ is a second surface parametrized by $\mathbf{y}\colon V \to N = \mathbf{y}(V)$ then we can also consider a map $f\colon M \to N$. It is smooth if and only if $\mathbf{y}^{-1}\circ f\circ\mathbf{x}$ is a smooth map $U \to V$. (Note also that in this case, the smooth extension $\bar f$ to a neighborhood of M in $\mathbb{R}^3$ can be chosen to take values in N.)

Now suppose we have two parametrizations $\mathbf{x}\colon U \to M$ and $\mathbf{y}\colon V \to M$ with overlapping images. That means on the open subset $W = \mathbf{x}(U)\cap\mathbf{y}(V)$ of M we have two different systems of coordinates. Then the map $\varphi := \mathbf{y}^{-1}\circ\mathbf{x}\colon x^{-1}(W) \to y^{-1}(W)$ is a diffeomorphism between these open subsets of U and V (with inverse $\varphi^{-1} = \mathbf{x}^{-1}\circ\mathbf{y}$): Being a composition of homeomorphisms, φ is a homeomorphism, but we also see that φ (and symmetrically φ^{-1}) is smooth.

`End of Lecture 13 May 2019`

If $f\colon \mathbb{R}^n \to M \subset \mathbb{R}^3$ with $f(a) = p$ then of course $D_a f$ is a linear map from $\mathbb{R}^n$ to $T_pM \subset \mathbb{R}^3$. Similarly, given a smooth map $f\colon M \to \mathbb{R}^n$ we get a differential $D_pf\colon T_pM \to \mathbb{R}^n$, the restriction of $D_p\bar f$ for any extension $\bar f$. (Different extensions will have different derivatives in the normal direction $\boldsymbol{\nu}_p$ but not in tangent directions, since the derivative in any tangent direction can be computed as the derivative along a curve in M, where $\bar f = f$ is determined.) Finally, combining these two observations, the derivative of a map $f\colon M \to N$ between surfaces in $\mathbb{R}^3$ is of course a linear map $D_pf\colon T_pM \to T_{f(p)}N \subset \mathbb{R}^3$.

B3. The Gauss map and the shape operator

A key tool for studying curves was the unit tangent vector and its derivatives. A similar role for surfaces is played by the unit normal $\boldsymbol{\nu}$. Given a smooth surface M, the map $\boldsymbol{\nu}\colon p \mapsto \boldsymbol{\nu}_p \in \mathbb{S}^2 \subset \mathbb{R}^3$ is called the *Gauss map* of the surface M, and is a smooth map $\boldsymbol{\nu}\colon M \to \mathbb{S}^2$. Note that the Gauss map of $\mathbb{S}^2$ itself is the identity map. (Or the antipodal map, if we oriented $\mathbb{S}^2$ with the inward normal.)

Going back to the general case, the differential of the Gauss map at $p \in M$ is a linear map $D_p\boldsymbol{\nu}\colon T_pM \to T_{\boldsymbol{\nu}_p}\mathbb{S}^2$. But these are the same plane – the plane in $\mathbb{R}^3$ with normal $\boldsymbol{\nu}$. (Of course any two planes are isomorphic vector spaces, but these are *naturally* isomorphic.) Thus we can view $D_p\boldsymbol{\nu}$ as a linear operator on T_pM. Its negative $S_p := -D_p\boldsymbol{\nu}\colon T_pM \to T_pM$ is called the *shape operator* (or Weingarten operator).

Recall that an operator $A\colon V \to V$ on an inner product space V is called *self-adjoint* if $\langle Av, w\rangle = \langle v, Aw\rangle$ for all $v, w \in V$. This is equivalent to saying that the bilinear form $(v, w) \mapsto \langle Av, w\rangle$ is symmetric (and thus induces a quadratic form $v \mapsto \langle Av, v\rangle$). Note that if A is 2-dimensional with basis $\{e, f\}$ then it suffices to check $\langle Ae, f\rangle = \langle e, Af\rangle$.

Proposition B3.1. *The shape operator S_p on T_pM is self-adjoint.*

Proof. Consider a parametrization $x\colon U \to M$ of a neighborhood of p and use $\{\mathbf{x}_u, \mathbf{x}_v\}$ as a basis for T_pM. The claim is that $\big\langle D_p\boldsymbol{\nu}(\mathbf{x}_u), \mathbf{x}_v\big\rangle = \big\langle D_p\boldsymbol{\nu}(\mathbf{x}_v), \mathbf{x}_u\big\rangle$. Write $\boldsymbol{\nu}_u := D_p\boldsymbol{\nu}(\mathbf{x}_u)$ and $\boldsymbol{\nu}_v := D_p\boldsymbol{\nu}(\mathbf{x}_v)$ for these partial derivatives of $\boldsymbol{\nu} \circ \mathbf{x}$ (and write $\mathbf{x}_{uv}$ for the mixed second partial of $\mathbf{x}$). Differentiating $\langle \boldsymbol{\nu}, \mathbf{x}_u\rangle \equiv 0$ in the $\mathbf{x}_v$ direction gives $\langle \boldsymbol{\nu}_v, \mathbf{x}_u\rangle = -\langle \boldsymbol{\nu}, \mathbf{x}_{uv}\rangle$, while differentiating $\langle \boldsymbol{\nu}, \mathbf{x}_v\rangle \equiv 0$ in the $\mathbf{x}_u$ direction gives $\langle \boldsymbol{\nu}_u, \mathbf{x}_v\rangle = -\langle \boldsymbol{\nu}, \mathbf{x}_{uv}\rangle$. Thus $\langle \boldsymbol{\nu}_u, \mathbf{x}_v\rangle = \langle \boldsymbol{\nu}_v, \mathbf{x}_u\rangle$, proving the claim. $\square$

Given this proposition, the shape operator S_p defines a quadratic form $v \mapsto \langle S_pv, v\rangle$ on T_pM, called the *second fundamental form h_p* of M, often written using the Roman numeral as $I\!I_p(v) := h_p(v)$. Note that arguments as in the proof show

$$\langle \boldsymbol{\nu}_u, \mathbf{x}_u\rangle = -\langle \boldsymbol{\nu}, \mathbf{x}_{uu}\rangle, \qquad \langle \boldsymbol{\nu}_v, \mathbf{x}_v\rangle = -\langle \boldsymbol{\nu}, \mathbf{x}_{vv}\rangle.$$

Thus the matrix of the second fundamental form w.r.t. the basis $\{\mathbf{x}_u, \mathbf{x}_v\}$ is

$$h_p := \begin{pmatrix} L & M \\ M & N \end{pmatrix} := \begin{pmatrix} -\langle \boldsymbol{\nu}_u, \mathbf{x}_u\rangle & -\langle \boldsymbol{\nu}_u, \mathbf{x}_v\rangle \\ -\langle \boldsymbol{\nu}_v, \mathbf{x}_u\rangle & -\langle \boldsymbol{\nu}_v, \mathbf{x}_v\rangle \end{pmatrix} = \begin{pmatrix} \langle \boldsymbol{\nu}, \mathbf{x}_{uu}\rangle & \langle \boldsymbol{\nu}, \mathbf{x}_{uv}\rangle \\ \langle \boldsymbol{\nu}, \mathbf{x}_{vu}\rangle & \langle \boldsymbol{\nu}, \mathbf{x}_{vv}\rangle \end{pmatrix}.$$

We see that to compute the first and second fundamental forms of a parametrized surface, we start by computing the first and second partial derivatives ($\mathbf{x}_u$, $\mathbf{x}_v$; $\mathbf{x}_{uu}$, $\mathbf{x}_{uv}$, $\mathbf{x}_{vv}$), then compute the cross product $\mathbf{x}_u \times \mathbf{x}_v$ and its length $|\mathbf{x}_u \times \mathbf{x}_v|$. The scalar products among the first derivatives give the matrix g_p. The scalar products of the second derivatives with $\mathbf{x}_u \times \mathbf{x}_v$, divided by the length of this normal vector, give the matrix h_p.

As usual, it is easier to find the matrix for the bilinear/quadratic form h_p than to find the matrix for the associated operator, the shape operator S_p. (Since $\{\mathbf{x}_u, \mathbf{x}_v\}$ is not generally orthonormal, it is easier to find the scalar products of $\boldsymbol{\nu}_u$ with the basis elements than to find its expression in the basis.) But by linear algebra we know $h_p = g_pS_p$, or equivalently $S_p = g_p^{-1}h_p$. Of course the inverse of a 2×2 matrix is easy to compute:

$$g_p^{-1} = \begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1} = \frac{1}{EG - F^2}\begin{pmatrix} G & -F \\ -F & E \end{pmatrix}.$$

The first and second fundamental forms are emphasized in many textbooks because they are easiest to compute in coordinates. But the shape operator S_p at a point $p \in M$ is more directly meaningful. It encodes all the different notions of curvature of the surface M at the point p, capturing the second-order behavior of the surface, or more precisely, exactly those parts which are independent of parametrization and invariant under rigid motion.

B4. Curvatures of a surface

Recall a few facts about a self-adjoint linear operator A on an inner product space V. Its eigenvalues are all real; its eigenvectors are perpendicular (since $\langle \lambda v, w\rangle = \langle Av, w\rangle = \langle v, Aw\rangle = \langle v, \mu w\rangle$ implies $\langle v, w\rangle = 0$ for $\lambda \neq \mu$). That is, we can choose an orthonormal basis of eigenvectors, and then A is of course represented by a diagonal matrix. The largest and smallest eigenvalues are the minimum and maximum of the quadratic form $\langle Av, v\rangle$ over the unit sphere in V. (Of course, if v is a unit eigenvector with eigenvalue λ then $\langle Av, v\rangle = \lambda$.)

Especially important are the symmetric functions of the eigenvalues. (These are the coefficients of the characteristic polynomial $\det(\lambda I - A)$, whose roots are the eigenvalues.) In particular, the product of the eigenvalues is the *determinant* $\det A$ and their sum is the *trace* $\operatorname{tr} A$. The average eigenvalue $\operatorname{tr} A/\dim V$ is also the average of $\langle Av, v\rangle$ over the whole unit sphere.

Now let's consider the shape operator S_p on T_pM. Its eigenvalues k_1 and k_2 are called the *principal curvatures* (of M at p); the eigenvectors are the *principal curvature directions*, forming two orthogonal lines in T_pM. We can choose unit eigenvectors e_1 and e_2 such that $\{e_1, e_2, \boldsymbol{\nu}\}$ is an oriented orthonormal basis. We define the *Gauss curvature*

$$K := k_1k_2 = \det S_p$$

and the *mean curvature*.

$$H := \frac{k_1 + k_2}{2} = \frac{1}{2}\operatorname{tr} S_p.$$

Note that K is independent of orientation, while H changes sign if we switch the sign of $\boldsymbol{\nu}$; more intrinsic is the mean curvature vector $\vec{H} = H\boldsymbol{\nu}$. (Note also that some authors define the mean curvature with the opposite sign and/or without the factor $1/2$ – despite the name "mean".)

`End of Lecture 17 May 2019`

As we will see, the intrinsic local shape of the surface is determined by the Gauss curvature K, in particular, qualitatively by its sign. We say $p \in M$ is an *elliptic* point if $K(p) > 0$ (that is, the principal curvatures have the same sign) or a *hyperbolic* point if $K(p) < 0$ (... opposite signs). A point where $K(p) = 0$ is called a *parabolic* point.

A point where $k_1 = k_2$ is called an *umbilic* point; in particular a *planar* point has $k_1 = k_2 = 0$. (Nonplanar umbilic points are of course elliptic.) At umbilic points, the principal directions are not uniquely defined and the normal curvature defined below is constant. Note that some authors use "parabolic" to mean what we call "parabolic but not planar"; this no longer depends just on K, so it is not an intrinsic notion.

The Gauss curvature K and the mean curvature H are smooth functions on any smooth surface. The principal curvatures k_1 and k_2 are the two roots $H \pm \sqrt{H^2 - K}$ of the characteristic polynomial $k^2 - 2Hk + K$ of the shape operator. These are smooth functions only away from umbilic points (where they coincide because the square root vanishes).

For a unit tangent vector $w \in T_pM$, the *normal curvature* of M in direction w is $h_p(w) = \langle S_p w, w\rangle$. Note that for an arbitrary nonzero vector $w \in T_pM$,

$$h_p\left(\frac{w}{|w|}\right) = \frac{\langle S_p w, w\rangle}{\langle w, w\rangle} = \frac{h_p(w)}{g_p(w)}.$$

We can write any unit normal vector as $w = \cos\theta\, e_1 + \sin\theta\, e_2$ and we find the normal curvature of M in this direction is $\cos^2\theta\, k_1 + \sin^2\theta\, k_2$, a weighted average of the principal curvatures. Of course, the mean curvature is the average normal curvature over the whole circle of directions; the principal curvatures are the minimum and maximum of the normal curvature.

The intersection of M with a normal plane at p – a plane spanned by $\boldsymbol{\nu}_p$ and some unit tangent vector $w \in T_pM$ – is a curve α with $T = w$ and $\vec{\kappa} = h_p(w)\boldsymbol{\nu}$. Orienting this normal plane such that $\{w, \boldsymbol{\nu}\}$ is an oriented basis, we can check that $h_p(w) = \kappa_g$: the normal curvature is the (signed) curvature of the normal slice α. We save the detailed calculation for later, when we will show that the normal curvature $h_p(w)$ is the component normal to M of the curvature of an arbitrary curve through p in direction w.

Let us now consider a surface given as a graph: $\mathbf{x}(u, v) = (u, v, f(u, v))$, in particular at a point p where $\operatorname{grad} f = 0$ so that the surface is horizontal there. We have $\mathbf{x}_u = (1, 0, f_u)$, $\mathbf{x}_v = (0, 1, f_v)$ so that at p this is an orthonormal basis for T_pM, meaning $g_p = I$. Of course $\boldsymbol{\nu}_p = (0, 0, 1)$. Since the normal (vertical) components of the second derivatives of $\mathbf{x}$ are the second derivatives of f, we see that

$$h_p = \operatorname{hess} f = \begin{pmatrix} f_{uu} & f_{uv} \\ f_{vu} & f_{vv} \end{pmatrix}.$$

Of course, since we have an orthonormal basis, this is also the matrix of the shape operator. Thus we have $K = f_{uu}f_{vv} - f_{uv}^2$ and $2H = \operatorname{tr}(\operatorname{hess} f) = f_{uu} + f_{vv} = \Delta f$. (In general, one should think of mean curvature as a geometric version of the Laplacian; in terms of the intrinsic Laplace–Beltrami operator Δ_M, we have for instance $\Delta_M \mathbf{x} = 2\vec{H}$.)

If we start with any point p on an arbitrary surface M, we can apply a rigid motion (or equivalently, choose new Euclidean coordinates) to put it into a standard position as follows. First translate so that $p = 0$ is the origin, then rotate so that $\boldsymbol{\nu}_p = (0, 0, 1)$ is vertical. Note that the surface is then locally a graph $z = f(x, y)$ with $\operatorname{grad} f = 0$ at p as above. Finally, rotate around the vertical axis until the x- and y-axes are principal directions. As above, with respect to the standard basis for $T_pM = \mathbb{R}^2$, the matrix for S_p is $\operatorname{hess} f$; this is now the diagonal matrix $\left(\begin{smallmatrix} k_1 & 0 \\ 0 & k_2 \end{smallmatrix}\right)$. This means the second-order Taylor expansion of f around 0 is $f(x, y) = 0 + k_1x^2 + k_2y^2 + \cdots$, where the remainder terms are third-order.

Thus we see that at any point $p \in M$ there is a uniquely determined paraboloid ($z = k_1x^2 + k_2y^2$ in the rotated coordinates) that has second-order contact with M at p. Two surfaces tangent at p have second-order contact if and only if they have the same principal curvatures and directions there (i.e., have the same ocsulating paraboloid). Up to rigid motion, two surfaces agree to second order (at given points) if and only if they have the same Gauss and mean curvatures there.

Note: for curves we preferred to talk about osculating circles (with constant curvature κ) rather than osculating parabolas (with constant acceleration vector). For surfaces, we might want to use a surface with constant principal curvatures (or equivalently, constant K and H). But we will see later this happens only for spheres and planes and cylinders. (Similarly, you will show in homework that spheres and planes are the only totally umbilic surfaces.) Although surfaces are in some sense determined by their curvatures, this is much more complicated than saying space curves are determined by $\kappa(s)$ and $\tau(s)$. First there are compatibility conditions (PDE not ODE: compatibility basically says $\boldsymbol{\nu}_{uv} = \boldsymbol{\nu}_{vu}$) and second there's no standard parametrization (like arclength).

From the Taylor series or osculating paraboloid, we do see for instance that near an elliptic point p, the surface is locally convex – it stays to one side of its tangent plane. This is not true at a hyperbolic point; instead T_pM cuts M locally in two curves crossing at p; their tangent vectors at p are exactly the directions with vanishing normal curvature, called *asymptotic* directions.

We could also consider the intersections with nearby parallel planes (say at distances $\pm\varepsilon^2$ to either side of the tangent plane). Unless p is a planar point, these planes will intersect M approximately in the curves $k_1x^2 + k_2y^2 = \pm\varepsilon^2$, which are scaled (by ε) versions of the *Dupin indicatrix*, defined to be the set of $w \in T_pM$ such that $\langle S_p w, w\rangle = \pm 1$. This is an ellipse at an elliptic point, a pair of hyperbolas (with common asymptotes in the asymptotic directions) at a hyperbolic point, and a pair of parallel lines at a nonplanar parabolic point.

`End of Lecture 20 May 2019`

B5. Curves on surfaces and the Darboux frame

Let's now consider an arbitrary curve α on a surface M. The *Darboux frame* along the curve is $\{T, \eta, \boldsymbol{\nu}\}$ where T is of course the unit tangent to α, $\boldsymbol{\nu}$ is the surface normal, and $\eta := \boldsymbol{\nu} \times T$ is called the *conormal*. As for our other frames, the derivative (with respect to an arclength parameter for α) gives a skew-symmetric matrix:

$$\begin{pmatrix} T \\ \eta \\ \boldsymbol{\nu} \end{pmatrix}' = \begin{pmatrix} 0 & \kappa_g & \kappa_n \\ -\kappa_g & 0 & \tau_g \\ -\kappa_n & -\tau_g & 0 \end{pmatrix} \begin{pmatrix} T \\ \eta \\ \boldsymbol{\nu} \end{pmatrix}.$$

Of course $\boldsymbol{\nu}$ depends only on the surface M, and so $\boldsymbol{\nu}' = D_p\boldsymbol{\nu}(T) = -S_p(T)$. Comparing with the above, we find $\kappa_n = \langle S_p(T), T\rangle$ and $\tau_g = \langle S_p(T), \eta\rangle$. Thus $\kappa_n = h_p(T)$ is the normal curvature of M in the direction T. It is the normal part of the curvature $\vec{\kappa} = \kappa_g\eta + \kappa_n\boldsymbol{\nu}$ of α. The tangential part κ_g is called the *geodesic curvature*. The remaining derivative τ_g, the twisting of the Darboux frame, is called the *geodesic torsion*. Like the normal curvature κ_n, the geodesic torsion $\tau_g = h_p(T, \eta) = h_p(\eta, T)$ depends only on M (and T). Curves in M having tangent T at p differ (to second-order) only in having different geodesic curvatures. (Walking in the mountains, we have a choice of turning left or right; whether we curve up or down is fixed by staying on the earth's surface.)

Curves α on M for which one of these quantities vanishes special. Curvature lines are curves for which $\tau_g = 0$, meaning that T is always a principal direction. (This is no condition at an umbilic point.) We have $\boldsymbol{\nu}' = -\kappa_n T$, where κ_n is one of the principal curvatures. The Darboux frame is a parallel (Bishop) frame along a curvature line. Curvature-line coordinates are (necessarily orthogonal) coordinates for which the coordinate lines are curvature lines. Locally away from umbilic points, this happens exactly when both the first and second fundamental forms have diagonal matrices in the basis $\{\mathbf{x}_u, \mathbf{x}_v\}$. (The coordinates we used for surfaces of revolution had this form.) This makes many computations much easier.

Asymptotic curves are those for which $\kappa_n \equiv 0$, that is, T is always an asymptotic direction. (This of course requires $K \le 0$.) The conormal η is the principal normal N of an asymptotic line; the Darboux frame is the Frenet frame. Locally near a hyperbolic point, asymptotic coordinates always exist, and are characterized by the second fundamental form having an off-diagonal matrix $h = \left(\begin{smallmatrix} 0 & M \\ M & 0 \end{smallmatrix}\right)$.

Geodesics are curves for which $\kappa_g = 0$. Given a starting point p and a starting direction T, there is always a unique geodesic, the solution to the ODE $T' = \kappa_n\boldsymbol{\nu}$. We will consider this in more detail later. (The surface normal is the principal normal $\boldsymbol{\nu} = N$ and the curvature equals the normal curvature.)

Any straight line contained in a surface (for instance, the rulings on a ruled surface) has constant T and is both an asymptotic line and a geodesic. (The geodesic torsion τ_g gives the speed at which the surface normal and conormal rotate around the line.)

A curvature line which is also a geodesic has constant conormal η; equivalently it is the intersection of M with a plane meeting M perpendicularly (like the generating curves on a surface of revolution).

The surface normal $\boldsymbol{\nu}$ is constant along a curvature line which is also asymptotic; such a curve is the intersection of M with a plane always tangent to M and consists of course of parabolic points. (Example: top or bottom of round torus – or of tube around any plane curve.)

B6. Vector fields and line fields

A (smooth) *vector field* X on $U \subset \mathbb{R}^2$ is a smooth map $U \to \mathbb{R}^2$ interpreted as $p \mapsto X_p \in T_p\mathbb{R}^2 = \mathbb{R}^2$. That is, we think of X_p as an arrow based at p. A *flow line* (or integral curve or trajectory) of X is a curve α in U whose velocity is given by X, that is, $\dot\alpha(t) = X_{\alpha(t)}$ for all t. This is a system of ODEs in the two variables u and v. The standard theorems on ODEs say give not only existence and uniqueness of flow lines, but also smooth dependence on the initial point. That is:

Theorem B6.1. *Let X be a smooth vector field on $U \subset \mathbb{R}^2$. For any $p \in U$ there exists a neighborhood $V \ni p$, a time interval $I = (-\varepsilon, \varepsilon)$, and a unique smooth map $\alpha\colon V \times I \to \mathbb{R}^2$ (called the* local flow *of X) such that for each $q \in V$ the curve $t \mapsto \alpha(q, t)$ is an integral curve of X through q, meaning that $\alpha(q, 0) = q$ and $\partial\alpha(q, t)/\partial t = X_{\alpha(q,t)}$.* □

For fixed $t \in I$ the map $q \mapsto \alpha(q, t)$ is called the flow of X by time t. Note that the uniqueness of integral curves implies that flowing by time s and then by time t is the same as flowing by time $s + t$, i.e., $\alpha(\alpha(q, s), t) = \alpha(q, s + t)$ (whenever both sides are defined). Taking $s = -t$ we see that the flow by time t is invertible: it is a diffeormorphism from V to its image in U.

Corollary B6.2. *If X is a vector field on U and $X_p \neq 0$ for some $p \in U$, then there exists a neighborhood $W \ni p$ and a smooth function $f\colon W \to \mathbb{R}$ which is constant along each flow line of X but has $Df \neq 0$ everywhere. (Such an f is called a local first integral for X.)*

Proof. Assume without loss of generality that $p = (0, 0)$ and $X_p = (1, 0)$. Let $\alpha\colon V \times I \to U$ be a local flow and consider its restriction $\bar\alpha$ to the two-dimensional cross-section $\{u = 0\}$ (transverse to X_p). This restriction $\bar\alpha(v, t)$ has nonsingular derivative at 0, so it locally has an inverse on some open neighborhood of $(0, 0)$, mapping this diffeomorphically to some $W \ni p$. We can take f to be the v-coordinate of this inverse. □

`End of Lecture 24 May 2019`

Definition B6.3. A (smooth) *vector field* X on a surface M is a function $M \to \mathbb{R}^3$ such that $X_p = X(p) \in T_pM$ for each $p \in M$.

With respect to a parametrization $\mathbf{x}\colon U \to M$, we can write $X_p = a(p)\mathbf{x}_u + b(p)\mathbf{x}_v = D_{(u,v)}\mathbf{x}(a, b)$ for smooth real-valued functions a, b on M. (Smoothness in these coordinates is equivalent to smoothness in $\mathbb{R}^3$.) We see that, just as $D_{(u,v)}\mathbf{x}$ gives a pointwise isomorphism between $T_{(u,v)}\mathbb{R}^2$ and $T_p(M)$, the derivative $D\mathbf{x}$ overall gives a one-to-one correspondance between vector fields on $U \subset \mathbb{R}^2$ and those on $\mathbf{x}(U) \subset M$.

Thus all local results about vector fields hold also on surfaces. (We can literally just replace U by M in the theorem or corollary above.) A stronger way to express the corollary is to say that around a point where $X_p \neq 0$ there are coordinates such that $X = \mathbf{x}_u$. (The restriction $\alpha(v,t)$ gives a new parametrization of W in terms of coordinates (v,t); the function f used above is the v coordinate and we rename t as u to get $X = \mathbf{x}_u$.) Summarizing, we can say: any nonzero vector field is locally constant in appropriate coordinates.

Theorem B6.4. *Suppose X and Y are two vector fields on M which are linearly indpendent at some point p. Then there exists a parametrization $\mathbf{x}\colon U \to M$ of some neighborhood $W \ni p$ such that $\mathbf{x}_u \| X$ and $\mathbf{x}_v \| Y$ on W.*

Note that it is too much to ask that $\mathbf{x}_u = X$ and $\mathbf{x}_v = Y$; coordinate vector fields always commute (in the sense that the time-t flow along X and the time-s flow along Y are commuting diffeomorphisms). But general vector fields X and Y do not have this property.

Proof. Let u and v be first integrals of Y and X (respectively) on some neighborhood $V \ni p$. The map $(u,v)\colon V \to \mathbb{R}^2$ has nonsingular differential at p. (Since $\dot u = 0$ only along curves tangent to X_p while $\dot v = 0$ only tangent to Y_p, these can never both vanish.) Thus it locally has an inverse, the desired parametrization. □

The theorem only depends on the equivalence classes of X and Y under $X \sim fX$ (where f is a nonvanishing scalar function). These equivalence classes are called *line fields* (or direction fields). Note that globally, a line field may be nonorientable. Locally, however, we can always pick a consistent orientation for the lines, so a line field always arises as above from a nonvanishing vector field. The theorem is really about a pair of (nowhere equal) line fields. (Line fields have unparametrized integral curves.)

We apply this theorem to derive local existence of various special kinds of coordinates.

Corollary B6.5. *Any $p \in M$ has a neighborhood with an orthogonal parametrization.*

Proof. We just need to find a pair of orthogonal vector fields. Start with an arbitrary parametrization $\mathbf{x}$ of some neighborhood of p. Set $X = \mathbf{x}_u$ and $Y = \boldsymbol{\nu} \times \mathbf{x}_u$ (that is, $Y_p = \boldsymbol{\nu}_p \times X_p$) and apply the theorem. The new parametrization has coordinate lines in the (orthogonal) X and Y directions. □

Corollary B6.6. *If $p \in M$ is a nonumbilic point, then some neighborhood of p can be parametrized by curvature-line coordinates.*

Proof. In a orientable neighborhood without umbilics, we can distinguish the principal curvatures (say $k_1 < k_2$). Then we get two line fields – along the eigenspaces for k_1 and k_2 respectively, and we can simply apply the theorem to get curvature-line coordinates. (We omit the details that show the principal directions are smooth functions.) □

Corollary B6.7. *If $p \in M$ is a hyperbolic point, then some neighborhood of p can be parametrized by asymptotic coordinates.*

Proof. In an orientable neighborhood where $K < 0$ we have two asymptotic lines at each point, and can distinguish them globally (one, say, is to the left of the negative principal curvature direction). Thus we get two line fields and can apply the theorem. □

B7. First variation of length

We want to understand the geometric meaning of the mean curvature H. In particular, if we consider variations of a surface, we will see how to express the derivative of area in terms of H. If a surface has $H \equiv 0$ then it is a critical point for area, called a *minimal surface*.

First we consider the simpler case of the length of a (compact) curve. Suppose $\alpha_t(s)$ is a smoothly varying family of smooth curves in $\mathbb{R}^n$. We assume that $\alpha(s) = \alpha_0(s)$ is parametrized by its arclength s. (But s is not an arclength parameter for the other curves α_t.) If α is not closed, we assume the variation is supported on a compact subinterval K away from the endpoints. (That is, $\alpha_t(s)$ is independent of t for s outside of K.)

We can take a Taylor series in t and get

$$\alpha_t(s) = \alpha(s) + t\xi(s) + O(t^2),$$

where ξ is a variation vector field along α. We will see that the derivative of length depends only on this infinitesimal variation, and not on the higher order terms we have omitted. (One could think of the vector field ξ as being a tangent vector to the infinite dimensional space of curves at the "point" α.)

We find the velocity of α_t is $\alpha_t' = T + t\xi' + O(t^2)$, and hence

$$|\alpha_t'|^2 = |T + t\xi'|^2 + O(t^2) = 1 + 2t\,\langle \xi', T\rangle + O(t^2).$$

Therefore $\frac{d}{dt}\big|_{t=0} |\alpha_t'| = \langle \xi', T\rangle$. Using $\operatorname{len}(\alpha_t) = \int_0^L |\alpha_t'|\,ds$ we find

$$\begin{aligned}\frac{d}{dt}\bigg|_{t=0} \operatorname{len}(\alpha_t) &= \frac{d}{dt}\bigg|_{t=0} \int |\alpha_t'|\,ds \\ &= \int \frac{d}{dt}\bigg|_{t=0} (|\alpha_t'|)\,ds = \int \langle \xi', T\rangle\,ds.\end{aligned}$$

Here smoothness justifies interchanging the derivative and integral. Next we integrate by parts, recalling that $T' = \vec\kappa = \kappa N$; our assumptions mean that the endpoint terms vanish. We find

$$\delta_\xi \operatorname{len}(\alpha) := \frac{d}{dt}\bigg|_{t=0} \operatorname{len}(\alpha_t) = -\int \langle \xi, \vec\kappa\rangle\,ds.$$

We can think of the right-hand side as the inner product on the function space $L^2(I, \mathbb{R}^3)$, the tangent space at α to the infinite dimensional space of curves, to which the vector fields ξ and $\vec{\kappa}$ along α belong. Thus the formula can be thought of as saying that $\vec{\kappa}$ is the negative gradient of the length functional.

`End of Lecture 27 May 2019`

A curve is length-critical if no variation changes its length to first order. That is, if we have $\delta_\xi \operatorname{len} = 0$ for all ξ, which is the case exactly when $\kappa \equiv 0$. Of course we know that straight lines minimize length.

We see that (as claimed above) the derivative $\delta_\xi \operatorname{len}$ only depends on ξ and not on the higher-order terms. Also, it is independent of the tangential part of ξ – if we keep the same family of curves α_t but change their parametrizations, that corresponds to changing ξ by a tangential field but clearly has no effect on (the derivative of) length. Any family of curves can be reparametrized so that the variation field ξ is normal.

Note furthermore that $\delta_\xi \operatorname{len}$ depends only the component of ξ in the principal normal direction. The so-called Hasimoto flow is the PDE $\xi = T \times \vec{\kappa} = \kappa B$ for a moving curve α_t in $\mathbb{R}^3$, which physically is an approximation to smoke-ring flow. Since we move only in the binormal direction, this flow preserves length; indeed is it a so-called integrable system which also preserves a whole hierarchy of other invariants.

Curve-shortening flow is the PDE $\xi = \vec{\kappa}$ for a moving curve α_t whose length decreases as fast as possible, since we follow the (negative) gradient direction. It is one of the earliest examples studied of a geometric flow, and has interesting properties like preserving embeddedness of plane curves.

Of course, if α lies on a surface $M \subset \mathbb{R}^3$, then our formula $\delta_\xi \operatorname{len}(\alpha) = -\int \langle \xi, \vec{\kappa} \rangle \, ds$ holds for all variations, including those that keep α on M. Their variation vector fields ξ are tangent to M. As before, up to reparametrization, we can assume ξ is normal to α. Thus $\xi = \varphi\eta$ is a varying multiple of the conormal vector η. We get $\delta_\xi \operatorname{len} \alpha = -\int \varphi\kappa_g \, ds$. Although only straight lines are length-critical with respect to all variations in space, we see that, considering only variations within M, a curve α is length-critical if and only if it has $\kappa_g \equiv 0$, that is, if and only if it is a geodesic.

Here are a few facts without proof. We will return to them next semester for more general manifolds. First, sufficiently short arcs of any geodesic are length-minimizing. On any surface, we can define a metric by setting $d(p, q)$ to be the infimal length of paths from p to q. (One shows that this infimum never vanishes for $p \neq q$ and that the metric topology coincides with the usual topology on M.) If the infimum is realized, that is, if a shortest path along M from p to q exists then it is a geodesic. If M is closed, then shortest paths always exist.

B8. Minimal surfaces

We now want to do the similar calculation to find the first variation of surface area. A more sophisticated approach would use the characterization $2\vec{H} = \Delta_M \mathbf{x}$ and an intrisic version of Stokes' theorem. We will take a more hands-on approach in coordinates.

Consider an initial surface with an orthogonal parametrization $\mathbf{x}\colon U \to M$. Let $\varphi\colon U \to \mathbb{R}$ have compact support in U and describe a normal variation of M. That is, we consider the family of surfaces $\mathbf{x}^t := \mathbf{x} + t\varphi\boldsymbol{\nu}$. (Guided by our experience with curves, we realize that nothing would change if we added tangential terms or higher-order terms.)

We find $\mathbf{x}^t_u = \mathbf{x}_u + t\varphi\boldsymbol{\nu}_u + t\varphi_u\boldsymbol{\nu}$ and $\mathbf{x}^t_v = \mathbf{x}_v + t\varphi\boldsymbol{\nu}_v + t\varphi_v\boldsymbol{\nu}$. Recalling that we have assumed $\langle \mathbf{x}_u, \mathbf{x}_v \rangle = 0$, this gives

$$\begin{aligned}
\langle \mathbf{x}^t_u, \mathbf{x}^t_u \rangle &= \langle \mathbf{x}_u, \mathbf{x}_u \rangle + 2t\varphi \langle \mathbf{x}_u, \boldsymbol{\nu}_u \rangle + O(t^2), \\
\langle \mathbf{x}^t_u, \mathbf{x}^t_v \rangle &= t\varphi \langle \mathbf{x}_u, \boldsymbol{\nu}_v \rangle + t\varphi \langle \mathbf{x}_v, \boldsymbol{\nu}_u \rangle + O(t^2), \\
\langle \mathbf{x}^t_v, \mathbf{x}^t_v \rangle &= \langle \mathbf{x}_v, \mathbf{x}_v \rangle + 2t\varphi \langle \mathbf{x}_v, \boldsymbol{\nu}_v \rangle + O(t^2).
\end{aligned}$$

This can be written as $g^t = g - 2t\varphi h + O(t^2)$.

Since in our orthogonal coordinates $g = \left(\begin{smallmatrix} E & 0 \\ 0 & G \end{smallmatrix}\right)$ is a diagonal matrix, the off-diagonal entries are irrelevant for the first-order calculation of $\det g^t$. Writing $h = \left(\begin{smallmatrix} L & M \\ M & N \end{smallmatrix}\right)$ we get in fact

$$\begin{aligned}
\det g^t &= (E - 2t\varphi L)(G - 2t\varphi N) + O(t^2) \\
&= EG - 2t\varphi(LG + NE) + O(t^2).
\end{aligned}$$

Taking the square root gives

$$\sqrt{\det g^t} = \sqrt{EG}\Big(1 - t\varphi\frac{LG + NE}{EG}\Big) + O(t^2),$$

but we recognize the fraction $\frac{LG+NE}{EG} = 2H$ as twice the mean curvature. Thus

$$\delta_\varphi \sqrt{\det g} = \frac{d}{dt}\Big|_{t=0} \sqrt{\det g^t} = -2\varphi H \sqrt{\det g}.$$

To find the first variation of area, we simply integrate this over U:

$$\begin{aligned}
\delta_\varphi \operatorname{area}(\mathbf{x}) &= \delta_\varphi \int_M dA = \delta_\varphi \int_U \sqrt{\det g} \, du \, dv \\
&= \int_U \big(\delta_\varphi \sqrt{\det g}\big) \, du \, dv = \int_U -2\varphi H \sqrt{\det g} \, du \, dv \\
&= -2 \int_M \varphi H \, dA
\end{aligned}$$

We see that the mean curvature is the negative gradient for area – to save area one should move the surface in the direction of the mean curvature vector. A surface is area-critical if and only if δ_φ area is zero for every variation φ, that is, if and only if $H \equiv 0$. Such a surface is called a minimal surface. Least-area surfaces spanning a given boundary are known to exist and be smooth; they are thus minimal surfaces. (It can also be shown that, given a minimal surface M, any sufficiently small piece of M – indeed any piece which is a graph in some direction – is the least-area way to span its boundary.)

Minimal surfaces have many interesting properties. For instance the Gauss map $\boldsymbol{\nu}\colon M \to \mathbb{S}^2$ is (anti)conformal, since its differential has matrix $\left(\begin{smallmatrix} k & 0 \\ 0 & -k \end{smallmatrix}\right)$ in an orthonormal basis of principal directions at a point with $K = -k^2$. One can check that in any conformal parametrization $\mathbf{x}\colon U \to M$ of a minimal surface, the coordinate functions – or more generally all height functions $\langle \mathbf{x}, u\rangle : U \to \mathbb{R}$ for constant $u \in \mathbb{S}^2$ – are harmonic functions. Thus they can be thought of as the real parts of complex holomorphic functions, leading to the so-called Weierstrass representation. (Thinking of $\mathbb{S}^2$ as the Riemann sphere $\hat{\mathbb{C}}$, the Gauss map itself is a meromorphic function.)

`End of Lecture 31 May 2019`

B9. Isometries

What do we mean when we talk about *intrinsic* properties of a surface, properties that only depend on the intrinsic geometry of the surface and not on how it sits in space? More precisely, these are properties that are the same for any two isometric surfaces.

Definition B9.1. An *isometry* is a diffeomorphism $\varphi\colon M \to N$ between two surfaces which preserves the scalar product on tangent spaces. That is, for any $p \in M$ and any $v, w \in T_pM$, we have $\langle D_p\varphi(v), D_p\varphi(w)\rangle = \langle v, w\rangle$. It follows that φ preserves the length of curves: $\mathrm{len}(\varphi \circ \alpha) = \mathrm{len}(\alpha)$ for any curve in M.

As a trivial example, if φ is a rigid motion of $\mathbb{R}^n$ then of course it restricts to any surface M to give an isometry $M \to \varphi M$. Less trivially, $\mathbb{R} \times (-\pi, \pi)$ is isometric to the unit cylinder in $\mathbb{R}^3$ with one vertical line removed. (We see from this simple example that the mean curvature H, for instance, is not an intrinsic notion; the surprising result later will be that the Gauss curvature K is intrinsic.)

Note that if $\mathbf{x}\colon U \to \mathbf{x}(U) = M$ is a parametrization and $\varphi\colon M \to N$ is a diffeomorphism then of course $\mathbf{y} := \varphi \circ \mathbf{x}\colon U \to N$ is a parametrization. We see that $\varphi = \mathbf{y} \circ \mathbf{x}^{-1}$ is an isometry if and only if, at corresponding points, the first fundamental forms for $\mathbf{x}$ and $\mathbf{y}$ have the same matrix with respect to the coordinate bases. That is, $\langle \mathbf{x}_u, \mathbf{x}_v\rangle = \langle \mathbf{y}_u, \mathbf{y}_v\rangle$, etc.

Definition B9.2. We say surfaces M and N are *locally isometric* if each point in either surface has a neighborhood isometric to an open subset of the other surface.

Note that we can assume the neighborhoods are small enough to be parametrized patches. Then we test local isometry by finding parametrizations with the same first fundamental form $g(u, v) = g = \left(\begin{smallmatrix} E & F \\ F & G \end{smallmatrix}\right)$.

The cylinder and plane form an example of locally isometric surfaces, as do the catenoid and helicoid. (In each case the second surface is topologically the universal covering of the first.)

Relaxing the condition of isometry, we can consider *conformal* (angle-preserving) maps $\varphi\colon M \to N$. Here the condition is that there is a positive function $\lambda\colon M \to \mathbb{R}$ (called the *conformal factor*) such that for any $v, w \in T_pM$ we have $\langle D_p\varphi(v), D_p\varphi(w)\rangle = \lambda(p)^2 \langle v, w\rangle$.

We have seen the sense in which two surfaces are locally isometric if and only if they have the same first fundamental form g. A somewhat surprising result is that all surfaces are locally conformal. (By transitivity, it suffices to prove that any surface has conformal parametrizations. This is a PDE result that we won't try to prove here.) The analog is not true for higher-dimensional manifolds – only certain special metrics are *conformally flat*.

B10. Covariant derivatives

If M is a surface, and Y is a (vector-valued) function on M, then we know what the directional derivative of Y at $p \in M$ in direction $w \in T_pM$ means: $\partial_w Y(p) = D_pY(w)$ is the derivative of Y along any curve (in M through p) with velocity vector w. Now suppose $Y\colon M \to \mathbb{R}^3$ is a tangent vector field (meaning $Y_p \in T_pM$ for each p). In general, its directional derivatives $\partial_w Y$ will not be tangent to M. Indeed, since the tangent spaces change as $p \in M$ moves, a tangent vector is forced to change in the normal direction just to stay tangent. We can make this precise using the second fundamental form – since $\langle Y, \boldsymbol{\nu}\rangle \equiv 0$, we get (omitting the subscript p, where all values are taken):

$$\langle \partial_w Y, \boldsymbol{\nu}\rangle = \langle DY(w), \boldsymbol{\nu}\rangle = -\langle D\boldsymbol{\nu}(w), Y\rangle = \langle S(w), Y\rangle = h(w, Y).$$

This normal change in Y is forced by the geometry of M.

The intrinsic change in Y is given by the tangential parts of its directional derivatives, called *covariant derivatives*. We write

$$\nabla_w Y = (\partial_w Y)^{\parallel} = \partial_w Y - \langle \partial_w Y, \boldsymbol{\nu}\rangle \boldsymbol{\nu} = D_pY(w) - h(w, Y_p) \in T_pM.$$

To define $\partial_w Y$ and $\nabla_w Y$ it of course not necessary that Y be defined on a whole neighborhood of p (in M) – it suffices if Y is a vector field (tangent to M) along some curve $\alpha\colon I \to M$ through p with velocity (parallel to) w. If Y is defined along α sometimes we write $\frac{\nabla}{dt} Y := \nabla_{\dot\alpha(t)} Y$ for the tangential part of the derivative $D_pY(\dot\alpha) = \dot Y = \frac{d}{dt} Y(\alpha(t))$ of Y along the curve α. (Whenever we talk about t-derivatives of Y, we are really differentiating the composition $Y \circ \alpha$.)

If α is parametrized at unit speed, then $\frac{\nabla}{ds} Y = \nabla_T Y$ is the tangential part of Y'. In particular, one example of a vector-field along α is $Y := T$; comparing the expression for the covariant derivative with the Darboux equation $T' = \kappa_g \eta + \kappa_n \boldsymbol{\nu}$, we see that $\nabla_T T = \kappa_g \eta$ gives the geodesic curvature. The equation $\kappa_g = 0$ for a geodesic can be written $\nabla_T T = 0$. (A curve satisfies $\nabla_{\dot\alpha} \dot\alpha = 0$ if and only if it is a geodesic in M parametrized at constant – not necessarily unit – speed.)

We say the vector field Y is *parallel* along α if $\frac{\nabla}{dt} Y \equiv 0$. Given Y_p at any initial point $p = \alpha(0)$, there is a unique way to extend it to a parallel field Y along α (solving the ODE $\nabla_T Y = 0$). Any curve α from p to q thus gives a map $T_pM \to T_qM$ called *parallel transport* – taking an initial vector at p to the value at q of the parallel field along α. It is important to note that this parallel transport from p to q does depend on the choice of the curve α from p to q – it is not a *natural* identification of the distinct tangent spaces.

Important properties are include the following: a parallel vector field has constant length; a pair of parallel fields make constant angle; thus, parallel transport is an orthogonal map $T_pM \to T_qM$, a map respecting the scalar products. This is easy to confirm: parallel fields have derivatives only in the normal direction $\boldsymbol{\nu}$, so $\langle X, Y\rangle' = \langle X', Y\rangle + \langle X, Y'\rangle = 0$.

Given two vector fields X and Y on M, the covariant derivative $\nabla_{X_p} Y$ is defined at every point $p \in M$, thus defining a new vector field $\nabla_X Y$. Note that $\nabla_X Y$ is $\mathbb{R}$-linear in each argument:

$$\nabla_{X+X'} Y = \nabla_X Y + \nabla_{X'} Y, \quad \nabla_X(Y + Y') = \nabla_X Y + \nabla_X Y',$$
$$\nabla_{aX} Y = a\nabla_X Y = \nabla_X(aY) \quad (for) \quad a \in \mathbb{R}.$$

If $f\colon M \to \mathbb{R}$ is a smooth function, then of course fX means the vector field whose value at $p \in M$ is $(fX)_p := f(p)X_p$. Since $\nabla_X Y$ at p depends only on X_p, we find that $\nabla_{fX} Y = f\nabla_X Y$. But on the other hand, the Leibniz product rule gives

$$\nabla_X(fY) = (\nabla_X f)Y + f\nabla_X Y,$$

where we adopt the convention that $\nabla_X f := \partial_X f$, the directional derivative of f in the direction X. Since any vector field is a combination $f\mathbf{x}_u + g\mathbf{x}_v$, these formulas will allow us to express covariant derivatives of arbitrary vector fields in terms of the covariant derivatives of the coordinate vector fields.

If X and Y are two vector fields on M, then in general $\nabla_X Y$ and $\nabla_Y X$ are unequal. The difference is called the *Lie bracket*:

$$[X, Y] := \nabla_X Y - \nabla_Y X.$$

A special property of coordinate vector fields is that they have vanishing Lie bracket: $[\mathbf{x}_u, \mathbf{x}_v] = 0$. To verify this, note that $\nabla_{\mathbf{x}_u}\mathbf{x}_v$ is by definition the tangential part of $\mathbf{x}_{uv}$. Thus $[\mathbf{x}_u, \mathbf{x}_v] = 0$ is a trivial consequence of $\mathbf{x}_{uv} = \mathbf{x}_{vu}$. Given two linearly independent vector fields X and Y, we discussed the fact that we cannot always find coordinates with $X = \mathbf{x}_u$ and $Y = \mathbf{x}_v$. Indeed the condition $[X, Y] = 0$ is exactly what is needed – this is part of the Frobenius Theorem (covered next semester).

B11. Christoffel symbols

For ease of writing equations in coordinates, we will change notation a bit: we write $(u^1, u^2) := (u, v)$ and use the subscript i for a partial derivative with respect to u^i, so for instance $\mathbf{x}_i := \mathbf{x}_{u^i} = \partial\mathbf{x}/\partial u^i$. Then we can write the entries of the matrices for the first and second fundamental forms as $g_{ij} = \langle \mathbf{x}_i, \mathbf{x}_j \rangle$ and $h_{ij} = \langle \boldsymbol{\nu}, \mathbf{x}_{ij} \rangle = -\langle \boldsymbol{\nu}_i, \mathbf{x}_j \rangle$.

We now want to explicitly calculate covariant derivatives in coordinates. We start with the covariant derivatives $\nabla_{\mathbf{x}_i}\mathbf{x}_j$ of the coordinate vector fields; since these are tangent vectors, they can be expressed in terms of the coordinate basis. We introduce the *Christoffel symbols* Γ^k_{ij} as their components. That is, the Γ^k_{ij} are defined by

$$\nabla_{\mathbf{x}_i}\mathbf{x}_j =: \sum_{k=1}^{2} \Gamma^k_{ij}\mathbf{x}_k = \Gamma^1_{ij}\mathbf{x}_1 + \Gamma^2_{ij}\mathbf{x}_2.$$

(The vanishing of the Lie bracket can now be expressed as the symmetry $\Gamma^k_{ij} = \Gamma^k_{ji}$.) Since we already know the normal component $h_{ij} = \langle \mathbf{x}_{ij}, \boldsymbol{\nu} \rangle$ of $\mathbf{x}_{ij}$, we could write the equations above as

$$\mathbf{x}_{ij} = \partial_{\mathbf{x}_i}\mathbf{x}_j = \Gamma^1_{ij}\mathbf{x}_1 + \Gamma^2_{ij}\mathbf{x}_2 + h_{ij}\boldsymbol{\nu}.$$

(This is called the Gauss formula.)

At every point in the surface, we have a (nonorthonormal) frame $\{\mathbf{x}_1, \mathbf{x}_2, \boldsymbol{\nu}\}$. The derivatives of $\boldsymbol{\nu}$ expressed in this frame – and the normal components of the derivatives of the $\mathbf{x}_i$ – are given by the second fundamental form and shape operator. The tangential parts of the derivatives $\mathbf{x}_{ij}$ are new – given by the Christoffel symbols.

Of course, as we saw with S and h, when dealing with nonorthonormal bases, it is easier to compute scalar products than components. Here we have

$$\langle \mathbf{x}_{ji}, \mathbf{x}_k \rangle = \langle \nabla_{\mathbf{x}_i}\mathbf{x}_j, \mathbf{x}_k \rangle = \Big\langle \sum_\ell \Gamma^\ell_{ij}\mathbf{x}_\ell, \mathbf{x}_k \Big\rangle = \sum_\ell \Gamma^\ell_{ij} g_{\ell k} =: \Gamma_{ijk}.$$

Here we multiply by the matrix $g = (g_{ij})$ to "lower an index". It is customary to write the inverse matrix as

$$(g^{ij}) := g^{-1} = \frac{1}{EG - F^2}\begin{pmatrix} G & -F \\ -F & E \end{pmatrix}.$$

Then multiplying by g^{-1} "raises the index" again: $\Gamma^k_{ij} = \sum_\ell g^{k\ell}\Gamma_{ij\ell}$. Of course we still have the symmetry $\Gamma_{ijk} = \Gamma_{jik}$.

Suppose we write X and Y in coordinates as $X = \sum \alpha^i \mathbf{x}_i$ and $Y = \sum \beta^i \mathbf{x}_i$ (for smooth functions α^i, β^i). Then by linearity and the product rule, we have

$$\begin{aligned}\nabla_X Y &= \sum_{i,j} \alpha^i \nabla_{\mathbf{x}_i}(\beta^j \mathbf{x}_j) = \sum_{i,k} \alpha^i \Big(\partial_{\mathbf{x}_i}\beta^k + \sum_j \beta^j \Gamma^k_{ij}\Big)\mathbf{x}_k \\ &= \sum_{i,k} \alpha^i \Big(\frac{\partial \beta^k}{\partial u^i} + \sum_j \beta^j \Gamma^k_{ij}\Big)\mathbf{x}_k.\end{aligned}$$

Now we consider derivatives of the coefficients of the first fundamental form:

$$\begin{aligned} g_{ij,k} := \partial_{\mathbf{x}_k} g_{ij} &= \partial_{\mathbf{x}_k}\langle \mathbf{x}_i, \mathbf{x}_j \rangle = \langle \partial_{\mathbf{x}_k}\mathbf{x}_i, \mathbf{x}_j \rangle + \langle \partial_{\mathbf{x}_k}\mathbf{x}_j, \mathbf{x}_i \rangle \\ &= \langle \mathbf{x}_{ik}, \mathbf{x}_j \rangle + \langle \mathbf{x}_{jk}, \mathbf{x}_i \rangle = \Gamma_{kij} + \Gamma_{kji}.\end{aligned}$$

This is a set of eight equations. We could write them out explicitly in the classical notation, getting $E_u = 2\Gamma_{111}$, etc. But let's just cyclically permute the equation above and use the symmetry of the Christoffel symbols:

$$\begin{aligned} g_{ij,k} &= \Gamma_{kij} + \Gamma_{kji} = \Gamma_{kij} + \Gamma_{jki}, \\ g_{jk,i} &= \Gamma_{ijk} + \Gamma_{ikj} = \Gamma_{ijk} + \Gamma_{kij}, \\ g_{ki,j} &= \Gamma_{jki} + \Gamma_{jik} = \Gamma_{jki} + \Gamma_{ijk}.\end{aligned}$$

Subtracting the top equation from the sum of the other two gives $g_{jk,i} + g_{ki,j} - g_{ij,k} = 2\Gamma_{ijk}$. More important than the exact form of this equation is the fact that it confirms the intrinsic nature of the covariant derivative: The Christoffel symbols, and thus all our formulas for covariant derivatives, can be expressed in terms of the first fundamental form (and its derivatives) alone.

After this excursion into very abstract notation, let's look concretely at these equations in more classical notation. We specialize to the case of an orthogonal parametrization (where

$g_{12} \equiv 0$); this makes all the equations a bit simpler. The equations $2\Gamma_{ijk} = g_{jk,i} + g_{ki,j} - g_{ij,k}$ become:

$$\begin{aligned} 2\Gamma_{111} &= g_{11,1} = E_u, & -2\Gamma_{112} &= g_{11,2} = E_v, \\ 2\Gamma_{121} &= g_{11,2} = E_v, & 2\Gamma_{122} &= g_{22,1} = G_u, \\ -2\Gamma_{221} &= g_{22,1} = G_u, & 2\Gamma_{222} &= g_{22,2} = G_v. \end{aligned}$$

Of course multiplying by the inverse of a diagonal matrix is easy, so $\Gamma^k_{ij} = \sum_\ell g^{k\ell}\Gamma_{ij\ell}$ becomes $\Gamma^1_{ij} = \Gamma_{ij1}/E$ and $\Gamma^2_{ij} = \Gamma_{ij2}/G$. That is, we get

$$\begin{aligned} \Gamma^1_{11} &= E_u/2E, & \Gamma^1_{12} &= E_v/2E & \Gamma^1_{22} &= -G_u/2E \\ \Gamma^2_{11} &= -E_v/2G, & \Gamma^2_{12} &= G_u/2G & \Gamma^2_{22} &= G_v/2G \end{aligned}$$

End of Lecture 10 June 2019

B12. Compatability conditions

Suppose we are given symmetric matrices g_{ij} and h_{ij} varying smoothly on a given domain $U \subset \mathbb{R}^2$. Is there some parametrization $\mathbf{x}\colon U \to \mathbb{R}^3$ with these as first and second fundamental form, respectively? Of course from g and h we know the Christoffel symbols and the matrix for the shape operator. Thus we try to solve the Gauss/Weingarten system

$$\mathbf{x}_{ij} = \Gamma^1_{ij}\mathbf{x}_1 + \Gamma^2_{ij}\mathbf{x}_2 + h_{ij}\boldsymbol{\nu}, \quad \boldsymbol{\nu}_i = -S(\mathbf{x}_i).$$

(Of course we also need that $\boldsymbol{\nu}$ is the unit normal vector to the surface given by $\mathbf{x}$. As long as that holds at some initial point, the Gauss/Weingarten system is set up to ensure it stays true, since the scalar products of $\boldsymbol{\nu}$ with itself and with the $\mathbf{x}_i$ will be constant.)

Unlike for ODEs, solutions to PDEs exist only if compatibility equations are satisfied. (The basic idea is that given functions g and h, the system $f_x = g, f_y = h$ can have a solution f only if $g_y = f_{xy} = f_{yx} = h_x$.) We will write down the compatibility equations for our system (equating $\mathbf{x}_{ijk} = \mathbf{x}_{ikj}$); these are clearly necessary.

Using the notation $\partial_k := \partial_{x_k} = \partial/\partial u^k$ for partial derivatives, differentiating the Gauss formula gives

$$\mathbf{x}_{ijk} = (\partial_k h_{ij})\boldsymbol{\nu} + h_{ij}\boldsymbol{\nu}_k + \sum_\ell \Big((\partial_k\Gamma^\ell_{ij})\mathbf{x}_\ell + \Gamma^\ell_{ij}\mathbf{x}_{\ell k}\Big).$$

The Gauss/Weingarten system shows us how to write the right-hand side in the basis $\{\mathbf{x}_1, \mathbf{x}_2, \boldsymbol{\nu}\}$. Each $\mathbf{x}_{ijk} = \mathbf{x}_{ikj}$ then gives us three scalar compatibility equations; since the equation is nontrivial only when $j \neq k$, the two cases of interest are $\mathbf{x}_{112} = \mathbf{x}_{121}$ and $\mathbf{x}_{221} = \mathbf{x}_{212}$. We first consider the normal components $\partial_k h_{ij} + \sum \Gamma^\ell_{ij} h_{\ell k}$ of $\mathbf{x}_{ijk}$. These normal components give the two *(Mainardi–)Codazzi* equations (evidently first discovered by Peterson), the first of our compatibility conditions:

$$\begin{aligned} \partial_2 h_{11} + \Gamma^1_{11}h_{12} + \Gamma^2_{11}h_{22} &= \partial_1 h_{12} + \Gamma^1_{12}h_{11} + \Gamma^2_{12}h_{21}, \\ \partial_1 h_{22} + \Gamma^1_{22}h_{11} + \Gamma^2_{22}h_{12} &= \partial_2 h_{12} + \Gamma^1_{12}h_{12} + \Gamma^2_{12}h_{22}. \end{aligned}$$

Here of course, the Christoffel symbols should be viewed as functions of the g_{ij}.

Recalling that the shape operator has matrix $g^{-1}h$, we write this in index notation as $h^j_i := \sum_k g^{jk}h_{ki}$, so that $\boldsymbol{\nu}_i = -S(\mathbf{x}_i) = -\sum_j h^j_i \mathbf{x}_j$. This lets us express, say, the $\mathbf{x}_2$ component of the equation $\mathbf{x}_{112} = \mathbf{x}_{121}$. We get

$$-h_{11}h^2_2 + \partial_2\Gamma^2_{11} + \sum_\ell \Gamma^\ell_{11}\Gamma^2_{\ell 2} = -h_{12}h^2_1 + \partial_1\Gamma^2_{12} + \sum_\ell \Gamma^\ell_{12}\Gamma^2_{\ell 1}.$$

This can be written as

$$\partial_2\Gamma^2_{11} - \partial_1\Gamma^2_{12} + \Gamma^1_{11}\Gamma^2_{12} + \Gamma^2_{11}\Gamma^2_{22} - \Gamma^1_{12}\Gamma^2_{11} - \Gamma^2_{12}\Gamma^2_{21} = h_{11}h^2_2 - h_{21}h^2_1.$$

Expanding $h_{11} = h^1_1 g_{11} + h^2_1 g_{21}$ and $h_{21} = h^1_2 g_{11} + h^2_2 g_{21}$, the right-hand side becomes

$$g_{11}(h^1_1 h^2_2 - h^2_1 h^1_2) = g_{11}\det S = g_{11}K.$$

Since the left-hand side is intrinsic (expressible in terms of the first fundamental form alone), so is the Gauss curvature K. That is, we have proved Gauss's *Theorema Egregium* ("remarkable theorem"): The Gauss curvature K is an intrinsic notion, remaining unchanged under local isometries, as when a surface is bent without stretching.

The equation above in the form

$$\partial_2\Gamma^2_{11} - \partial_1\Gamma^2_{12} + \Gamma^1_{11}\Gamma^2_{12} + \Gamma^2_{11}\Gamma^2_{22} - \Gamma^1_{12}\Gamma^2_{11} - \Gamma^2_{12}\Gamma^2_{21} = g_{11}\frac{\det h}{\det g},$$

again with the understanding that the Christoffel symbols should be expressed as functions of the g_{ij} and their derivatives, is the *Gauss equation*, the last compatibility condition.

A theorem of Bonnet (basically using standard results about first-order PDEs) now says these compatibility conditions are also sufficient. If symmetric matrix functions g_{ij} and h_{ij} (with g positive definite) satisfy the Gauss and Codazzi equations, then there is a surface with these fundamental forms, unique up to rigid motion. We will not go into the details of the proof.

Let us return to the special case of orthogonal coordinates, and write the intrinsic formula for Gauss curvature more explicitly. In classical notation, we get

$$2EK = -\partial_v\Big(\frac{E_v}{G}\Big) - \partial_u\Big(\frac{G_u}{G}\Big) + \frac{E_uG_u}{2EG} - \frac{E_vG_v}{2G^2} + \frac{E_v^2}{2EG} - \frac{G_u^2}{2G^2},$$

or equivalently

$$\begin{aligned} -2EGK &= E_{vv} - \frac{E_vG_v}{G} + G_{uu} - \frac{G_u^2}{G} \\ &\qquad - \frac{E_uG_u}{2E} + \frac{E_vG_v}{2G} - \frac{E_v^2}{2E} + \frac{G_u^2}{2G} \\ &= \sqrt{EG}\Big(\partial_v\Big(\frac{E_v}{\sqrt{EG}}\Big) + \partial_u\Big(\frac{G_u}{\sqrt{EG}}\Big)\Big). \end{aligned}$$

For conformal coordinates with conformal factor $\lambda = e^\varphi > 0$ we have $E = G = \lambda^2$ and the formula becomes

$$K = -\frac{1}{\lambda^2}\Big(\partial_v\Big(\frac{\lambda_v}{\lambda}\Big) + \partial_u\Big(\frac{\lambda_u}{\lambda}\Big)\Big) = -e^{-2\varphi}\Delta\varphi.$$

End of Lecture 14 June 2019

Bonnet's theorem leaves many related open questions. One is the following: to what extent the compatibility conditions determine the second fundamental form from the first fundamental form? This is a rigidity question: when does a surface admit a unique isometric embedding into $\mathbb{R}^3$? Unlike hypersurfaces in higher dimensions – see Kühnel's book – surfaces can always be locally embedded in many ways. Globally, however, there are often rigidity results (like for convex surfaces) and other (abstract) surfaces cannot be globally embedded at all. We may return to these results later.

For curves, of course, there is no intrinsic geometry – any two curves are locally isometric, for instance by choosing arclength parametrizations for both of them. Two plane curves are related by a rigid motion if and only if they have the same (extrinsic) curvature κ (as a function of arclength). Note, however, that specifying curvature as a given function $\kappa(t)$ of an unspecified parameter t gives hardly any information about the shape of the curve. Any arc of monotonic curvature can be parametrized by $\kappa = t$ for instance.

What is the situation for surfaces? Two surfaces are locally isometric (intrinsically equivalent, one might say) if and only if they have the same first fundamental form g in corresponding coordinates. This implies that they have the same Gauss curvature $K(u, v)$ in such coordinates. By Bonnet's theorem, two surfaces are related by a rigid motion (extrinsically equivalent, one might say) if and only if they have the same first and second fundamental forms g and h in corresponding coordinates. This implies that they have the same Gauss and mean curvatures K and H (or equivalently, the same principal curvatures) in such coordinates.

If we ask when surfaces with the same K are isometric, then we are faced with the same problem as for curves of not knowing what parametrization is being used to compare the curvature functions. For instance, as long as $p \in M$ is not a critical point of the function K, then a neighborhood of p is foliated by lines of constant K and we can use $v := K$ as one of the two coordinates in a regular parametrization of this neighborhood. One case where this problem doesn't arise is that of surfaces with constant Gauss curvature.

Suppose we have two isometric surfaces with the same mean curvature (or equivalently the same principal curvatures). This does not always imply that that they are related by a rigid motion. What happens is that, even though the eigenvalues of the shape operator are the same on both surfaces, the eigenvectors (the principal directions) can rotate. An important example is that of minimal surfaces: it turns out that any minimal surface has an isometric *conjugate minimal surface*. Here the curvature directions have become asymptotic directions and vice versa. (In fact, these sit in a one-parameter family of isometric minimal surfaces with all possible asymptotic directions.)

B13. Surfaces of constant curvature

There are many interesting facts about surfaces with constant Gauss curvature or constant mean curvature. Our first goal is a thereom of Minding saying any two surfaces with the same constant $K \equiv c$ are locally isometric. Then we turn to theorems of Liebmann that characterize the round sphere as the unique closed surface of constant K and the unique convex surface of constant H.

We will use one further special kind of parametrization:

Definition B13.1. An orthogonal parametrization $\mathbf{x}\colon U \to \mathbb{R}^3$ gives *geodesic parallel coordinates* if $|\mathbf{x}_u| \equiv 1$.

Using the notation $a = |\mathbf{x}_v|$, we have in geodesic parallel coordinates $E \equiv 1$ and $G = a^2$. Specializing our formula for K to this case gives $-2a^2K = a\partial_u(2aa_u/a)$, that is, $K = -a_{uu}/a$.

Clearly, in geodesic parallel coordinates, the u-curves are parametrized at unit speed: any pair of v-curves cut segements of equal length from all the u-curves. We claim the u-curves (for each constant v) are geodesics, so that the v-curves should really be considered as parallel curves at constant distance from each other. (The name "geodesic parallel coordinates" then comes from the fact that the u- and v-coordinate lines are geodesics and parallels, respectively.)

To check the claim, we must show $\nabla_{\mathbf{x}_u}\mathbf{x}_u = 0$. This is equivalent to $\Gamma^1_{11} = 0 = \Gamma^2_{11}$, or to $\Gamma_{111} = 0 = \Gamma_{112}$. But in orthogoal coordinates, we had $\Gamma_{111} = E_u/2$ and $\Gamma_{112} = -E_v/2$.

Lemma B13.2. *Any surface M locally admits geodesic parallel coordinates. Indeed we can choose any given curve $\alpha\colon v \mapsto \alpha(v) \in M$ as the v-curve $u \equiv 0$.*

Proof. Each u-curve $v \equiv c$ is determined as the geodesic starting at $\alpha(c)$ in the conormal direction. The only thing that one needs to check is that the coordinates stay orthogonal. But since the u-curves are geodesics, $\mathbf{x}_{uu}$ is normal to M, so in particular $\langle \mathbf{x}_{uu}, \mathbf{x}_v\rangle = 0$. Thus

$$\begin{aligned} 0 = E_v &= \partial_v \langle \mathbf{x}_u, \mathbf{x}_u\rangle = 2\langle \mathbf{x}_u, \mathbf{x}_{uv}\rangle \\ &= 2(\langle \mathbf{x}_v, \mathbf{x}_{uu}\rangle + \langle \mathbf{x}_u, \mathbf{x}_{uv}\rangle) = 2\partial_u \langle \mathbf{x}_u, \mathbf{x}_v\rangle = 2F_u. \end{aligned}$$

This implies $F \equiv 0$ since we know F vanishes along α. □

The special case where the initial curve α is itself a unit-speed geodesic gives what are called *Fermi coordinates* (along α), often used in Lorenzian geometry for general relativity (choosing α to be the wordline of some particle). In this case, not only is the first fundamental form the identity matrix along the whole starting curve, but also its derivative in the conormal direction vanishes, so all the Christoffel symbols vanish along that curve. (In particular our claim is that $E_u = 0 = G_u$ along α and this follows from calculations like the one above for F_u, using the additional fact that α is a geodesic. These are left as an exercise.)

Theorem B13.3 (Minding)**.** *Two surfaces with the same constant Gauss curvature K are locally isometric. Indeed, give a point in each surface and orthonormal frames at these points, the local isometry can be chosen to map the one frame to the other.*

Note that this implies that a surface of constant Gauss curvature has lots of (local) intrinsic symmetries: any two points have isometric neighborhoods.

Proof. Construct Fermi coordinates in each surface, starting with a unit-speed geodesic (through the given point in the direction of the first frame vector). The first fundamental form in these coordinates will be given by $\left(\begin{smallmatrix}1&0\\0&a^2\end{smallmatrix}\right)$ and we have that $K = -a_{uu}/a$ with $a(0, v) \equiv 1$. This can be viewed as the (ordinary!) differential equation $a_{uu} = -Ka$ for a, which of course has a unique solution given the initial conditions $a(0) = 1$, $a_u(0) = 0$. Thus a is independent of v and is the same on both surfaces. □

Note that of course we can solve this ODE explicitly. For $K = 0$ we have $a \equiv 1$ (as for the standard coordinates on $\mathbb{R}^2$, while for $K > 0$ we get $a = \cos(\sqrt{K}u)$, and for $K < 0$ we get $a = \cosh(\sqrt{-K}u)$. (Parallels on a sphere get shorter as we move away from the initial geodesic, whereas parallels in the hyperbolic plane get longer.)

Note that ordinary spherical coordinates (latitude, longitude) are Fermi coordinates (around the equator) for the round sphere with $K \equiv 1$. The pseudosphere is an example of a surface with $K \equiv -1$, but no such surface is complete.

Lemma B13.4. *In curvature-line coordinates with $g = \left(\begin{smallmatrix}E&0\\0&G\end{smallmatrix}\right)$ and $h = \left(\begin{smallmatrix}L&0\\0&N\end{smallmatrix}\right)$, the Codazzi equations become*

$$L_v = HE_v, \qquad N_u = HG_u.$$

In terms of the principal curvatures, with $L = k_1E$, $N = k_2G$, the equations can be written as

$$E_v = \frac{2E\partial_v k_1}{k_2 - k_1}, \quad G_u = \frac{2G\partial_u k_2}{k_1 - k_2}.$$

Proof. Dropping the terms involving h_{12}, the Codazzi equations are

$$\partial_2 h_{11} + \Gamma^2_{11}h_{22} = \Gamma^1_{12}h_{11}, \quad \partial_1 h_{22} + \Gamma^1_{22}h_{11} = \Gamma^2_{12}h_{22}.$$

Substituting the values we computed for the Christoffel symbols in arbitrary orthogonal coordinates gives

$$L_v - NE_v/2G = LE_v/2E, \quad N_u - LG_u/2E = NG_u/2G.$$

Recalling that the mean curvature is $H = L/2E + N/2G$ gives the first form.

Now differentiating the equations $L = k_1E$, $N = k_2G$ yields $(\partial_v k_1)E + k_1E_v = L_v = HE_v$, etc., which simplifies to the final equations in the statment. □

Lemma B13.5 (Hilbert)**.** *Suppose p is a nonumbilic point with $k_1 > k_2$ and suppose k_1 has a local maximum while k_2 has a local minimum at p. Then $K(p) \le 0$.*

Proof. Choose curvature-line coordinates in a neighborhood of p and use the classical notation of the last lemma. By assumption the derivatives of the principal curvatures vanish at p, so by the final formulas of the last lemma, $E_v = 0 = G_u$ there. Differentiating those formulas (and dropping terms involving first derivatives of k_i to evaluate at p), we find that

$$E_{vv} = \frac{2E\partial^2_{vv}k_1}{k_2 - k_1}, \quad G_{uu} = \frac{2G\partial^2_{uu}k_2}{k_1 - k_2}$$

at p. By assumption, at p we have $k_1 > k_2$ and also $\partial^2_{vv}k_1 \le 0 \le \partial^2_{uu}k_2$. Thus $E_{vv} \ge 0$ and $G_{uu} \ge 0$ at p. The Gauss equation in orthogonal coordinates gave a nice formula for K involving $\sqrt{EG}$. At a point where $E_v = 0 = G_u$, it is easy to reduce this formula to $-2EGK = E_{vv} + G_{uu}$. It follows immediately that $K(p) \le 0$. □

Recall our earlier claim that the only surfaces with constant H and K (or equivalently, with constant principal curvatures) are (pieces of) planes, spheres and cylinders. If $k_1 = k_2$ then we are in the totally umbilic case of planes or spheres. Otherwise, we can use the calculations from the proof of Hilbert's lemma. We have $\partial^2_{vv}k_1 = 0 \implies E_{vv} = 0$ and similarly $G_{uu} = 0$, giving $K \equiv 0$. Surfaces with $K \equiv 0$ are called *developable*. We will study these later and our first few results will suffice to conclude that the developable surfaces with constant H are round cylinders.

So far our study of surfaces has been local, working in one coordinate chart and ignoring the global topology of the surface. We now turn to some global results about *closed surfaces*, meaning connected compact surfaces (without boundary).

Lemma B13.6. *Any closed surface in $\mathbb{R}^3$ has at least one point (and hence an open set) where $K > 0$.*

Proof. Since M is compact, it is contained in some ball around the origin. Let $B_R(0)$ be the smallest such ball. Its boundary sphere (with normal curvatures $1/R$) must be tangent to M. Since M stays inside, its normal curvatures – in particular both principal curvatures – are at least $1/R$. Thus at the point of tangency $K > 1/R^2 > 0$. □

Theorem B13.7 (Liebmann 1899)**.** *A closed surface $M \subset \mathbb{R}^3$ with constant Gauss curvature K is necessarily a round sphere (of radius $1/\sqrt{K}$).*

Proof. By the lemma, we have $K > 0$. Denote the two principal curvatures of M by $k_1 \ge k_2$. By compactness k_1 attains a maximum at some $p \in M$ (where k_2 has a minimum, since $k_1k_2 \equiv K$). If $k_1 > k_2$ at p, then we are in the situation of Hilbert's lemma, so $K \le 0$, a contradiction. Thus we may assume $k_1(p) = \sqrt{K} = k_2(p)$, By the choice of p we then have $\sqrt{K} \ge k_1 \ge k_2 \ge \sqrt{K}$ everywhere, meaning that equality holds and the surface M is totally umbilic. As we have seen already, M is thus a piece of a sphere, indeed the whole sphere since it is closed. □

Theorem B13.8 (Liebmann 1900)**.** *A smooth closed surface $M \subset \mathbb{R}^3$ with $K > 0$ and constant mean curvature H is necessarily a round sphere (of radius $1/H$).*

Proof. We proceed exactly as before, letting k_1 attain its maximum at p. Again we just need to rule out the nonumbilic case $k_1 > k_2$. But here again Hilbert's lemma applies to give the contradiction $K \le 0$. □

One might ask if there are any other closed surfaces of constant mean curvature (CMC), perhaps even allowing "immersed" surfaces with self-intersections. Heinz Hopf (1955) conjectured not, and proved this in the case the surface is simply connected (topologically a sphere). A.D. Alexandrov proved (1962) there are no embedded examples of any genus. It was a surprise then in 1986 when Henry Wente found an immersed CMC torus. Since then, general methods for finding many such examples have been developed.

B14. The Umlaufsatz for smooth curves

At the beginning of the semester, we mentioned the "theorem on turning tangents", also known (even in English) as the Umlaufsatz: a simple closed plane curve has turning number 1 (or equivalently total signed curvature 2π). This is actually the planar case of the Gauss–Bonnet Theorem. The local version of Gauss–Bonnet talks about a simple closed curve γ enclosing a disk R on a surface M, and says the total geodesic curvature of γ in M equals 2π minus the total Gauss curvature of R:

$$\int_\gamma \kappa_g \, ds = 2\pi - \int_R K \, dA.$$

When $M = \mathbb{R}^2$ then of course $K \equiv 0$ and the geodesic curvature κ_g is the signed curvature of the plane curve, so Gauss–Bonnet does reduce to the Umlaufsatz; conversely, we use the Umlaufsatz as a lemma in our proof of Gauss–Bonnet.

Hopf was not the first to prove the Umlaufsatz, but it is his proof that we will sketch. See Do Carmo's book for more details.

Theorem B14.1 (Umlaufsatz)**.** *Let γ be a simple closed plane curve bounding a region $R \subset \mathbb{R}^2$. Orient γ so that R is to its left (that is, so that its normal vector $N = J(T)$ points into R). Then the turning number of γ is* 1.

Sketch of proof. Let the curve γ be parametrized at unit speed as an L-periodic map $\gamma\colon \mathbb{R} \to \mathbb{R}^2$. Shift the parameter if necessary, to ensure that $\gamma(0)$ is an extreme point on the convex hull. For convenience, rotate so that $\gamma(0)$ is a point with lowest y-coordinate along γ. Then $T(0) = \gamma'(0) = e_1$ is horizontal.

Now define the *secant map*

$$f(s,t) := \frac{\gamma(t) - \gamma(s)}{\left\|\gamma(t) - \gamma(s)\right\|} \in \mathbb{S}^1.$$

Because γ has no self-intersections, this is well defined on the diagonal strip $s < t < s + L$ in the (s,t)-plane. On the lower boundary $s = t$ it is extended smoothly by $f(s,s) = T(s)$, while on the upper boundary it is extended smoothly by $f(s, s + L) = -T(s)$. We will be interested in f restricted to the triangle

$$\Delta := \{(s,t) : 0 \le s \le t \le L\}.$$

Just as we lifted the $\mathbb{S}^1$-valued map $T(s)$ to a real-valued angle function $\theta(s)$ when we defined turning number, here we can lift the map $f\colon \Delta \to \mathbb{S}^1$ to a map $\theta\colon \Delta \to \mathbb{R}$ such that $f(s,t) = (\cos\theta(s,t), \sin\theta(s,t))$. (For smooth functions like we have here, this is easiest to do by considering what the derivatives of θ must be. But such a lift exists for any continuous f, as one learns in algebraic topology.) The lift is unique up to adding a constant multiple of 2π. We choose the lift for which $\theta(0,0) = 0$.

Along the diagonal, this $\theta(s,s)$ is the lift of T, so $\theta(L,L)$ is by definition 2π times the turning number: our goal is to show $\theta(L,L) = 2\pi$. Now consider the other sides of the triangle Δ, recalling that $\gamma(0)$ was chosen to be a lowest point on the curve.

Along the vertical side, $f(0,t)$ must point upwards, that is, it stays in the (closed) upper semicircle. Thus the angle function $\theta(0,t)$, starting at $\theta(0,0) = 0$ must stay in the interval $[0,\pi]$. When we reach $t = L$, where $f(0,L) = -e_1$ we know $\theta(0,L)$ is π (modulo 2π) so the only possibility in the interval is $\theta(0,L) = \pi$.

Continuing along the horizontal side, $f(s,L)$ always points downwards, staying in the lower semicircle. Thus, starting at $\theta(0,L) = \pi$, we see that $\theta(s,L)$ stays in the interal $[\pi, 2\pi]$. When we reach $\theta(L,L)$, which must be a multiple of 2π, we see it must be 2π as desired. □

It is known that there are no knotted curves in the plane: any simple closed curve can be deformed into a round circle while staying embedded. (An explicit deformation can be found for instance with a rescaled curve-shortening flow.) Using the basic idea here is that a continuous function with discrete (integer) values is constant, this gives another proof of the Umlaufsatz.

This is a recurring theme in topology. Algebraic topology associates to any topological space X various algebraic objects (fundamental groups, homology groups, etc.) and to any continuous map $f\colon X \to Y$ homomorphisms between the associated groups. For the circle $\mathbb{S}^1$, the fundamental group (or the first homology group) is the integers; a map $f\colon \mathbb{S}^1 \to \mathbb{S}^1$ induces a homomorphism $f_*\colon \mathbb{Z} \to \mathbb{Z}$ – this must be given by multiplication by some $d \in \mathbb{Z}$, called the *degree* of f.

Given a smooth closed plane curve α, its tangent vector can be viewed as a map $T\colon \mathbb{S}^1 \to \mathbb{S}^1$. The turning number of α is the degree of this map. For smooth maps like this, one can also bypass the machinery of algebraic topology and define the degree via differential topology, as we did by lifting to the angle function. Equivalently, the degree of f can be computed by integrating the derivative of f around the circle – just as we computed turning number as $\frac{1}{2\pi}\int \kappa \, ds$.

An alternative approach to degree is to note that almost every value in the range of f is a regular value, attained only at points where the differential of f is surjective. In particular, it is attained only finitely many times, and each time has a well-defined sign ± 1 depending on the orientation. The degree can be computed at any regular value as the sum of these signs.

(Another possibly familiar example of degree is the *winding number* used in complex analysis. If α is a closed curve in $\mathbb{R}^2 \smallsetminus \{0\}$, then the winding number of α around 0 is the degree of $\alpha/|\alpha|\colon \mathbb{S}^1 \to \mathbb{S}^1$.)

`End of Lecture 2019 June 21`

B15. The Umlaufsatz for piecewise smooth curves

We will want the Umlaufsatz not just for smooth curves but for all piecewise smooth curves. As a warmup, consider what it should say for simple polygons. The tangent vector T is constant along each side of a polygon but then jumps at each corner. If the interior angle is θ_i, then the *exterior* or *turning angle* is $\tau_i := \pi - \theta_i \in (-\pi, \pi)$, the signed angle between the incoming and outgoing tangent vectors. The total signed curvature should be replaced by $\sum \tau_i$; this will be 2π times the turning number.

Of course, the fact that $\sum \tau_i = 2\pi$ (or equivalently, $\sum \theta_i = (n-2)\pi$ for an n-gon) is standard in elementary geometry. To prove it by induction starting from the known case of a triangle, we just need to cut a larger polygon into pieces. (The one nontrivial lemma is that any simple polygon has some interior diagonal that can be drawn without crossing any edges. In fact one can always find an "ear", such a diagonal that cuts of just a triangle. Note that this is also the key lemma for a polygonal version of the Jordan curve theorem.)

How about piecewise smooth curves? (Note: important to have smooth on *closed* subintervals to get well-defined one-sided tangents.) Definition: $\int \kappa$ over smooth parts plus turning angles at junctions. But how do we choose the sign at cusps where $\tau = \pm\pi$? (Note: do Carmo's definition here doesn't always work.) Right definition: $+\pi$ if cusp points out from bounded region, $-\pi$ if cusp points into bounded region. From now on, we will write $\tau_i \in [-\pi, \pi]$ for this turning angle at the ith corner (with the correct choice of sign at cusps).

To prove the Umlaufsatz in this generality, one could try to take a limit of smooth or polygonal approximations – but preserving embeddedness and controlling the limiting value of total curvature is quite tricky. Hopf does it as follows: the secant map on the open triangle still limits to the tangent direction on all smooth points of the diagonal. The lifted map θ will jump at the points (s, s) corresponding to corners. We just need to show that the lifted map jumps by no more than π at each corner – and that the sign is right where the jump is $\pm\pi$. (The sign that naturally comes up here is given by the orientation of a triangle consisting of the cusp and two nearby points chosen such that the curve avoids the segment between them. One can check that this is equivalent to our definition above.)

B16. More on parallel transport

Now consider two nonvanishing vector fields X and Y, tangent to M along some curve α. As in our discussion of the total curvature of plane curves, even though the angle θ between X and Y is only defined up to multiples of 2π, given a choice of θ at the basepoint $\alpha(0)$, there is a unique smooth angle function along α. For consideration of the angle, we might as well assume that both vector fields have unit length.

So suppose X is a unit parallel field along α, meaning $\frac{\nabla}{dt}X = 0$, and Y is any unit field. Let $\theta(t)$ be the angle from X to Y at $\alpha(t)$. As we observed earlier, parallel transport preserves the angle between two vectors, so θ is constant if Y is also parallel. In general, we claim that the rate of change in angle θ measures the covariant derivative of Y, in the sense we will now explain.

Any unit Y is perpendicular to its derivative. In particular, the covariant derivative $\frac{\nabla}{dt}Y$, being perpendicular to $\boldsymbol{\nu}$ as well, is a scalar multiple of $\boldsymbol{\nu} \times Y$, the scalar being given by the triple product $\langle \frac{\nabla}{dt}Y, \boldsymbol{\nu} \times Y \rangle = \langle \dot{Y}, \boldsymbol{\nu} \times Y \rangle$.

In terms of the parallel field X, we can write $Y = \cos\theta X + \sin\theta(\boldsymbol{\nu} \times X)$. Taking the covariant derivative, we get

$$\frac{\nabla}{dt}Y = \dot{\theta}\Big(-\sin\theta X + \cos\theta(\boldsymbol{\nu} \times X)\Big) = \dot{\theta}(\boldsymbol{\nu} \times Y).$$

Equivalently, $\dot{\theta} = \langle \frac{\nabla}{dt}Y, \boldsymbol{\nu} \times Y \rangle$.

B17. Gauss–Bonnet: local form

Now we want to turn towards Gauss–Bonnet. As a first result for curves on surfaces, consider a parametrized surface patch $\mathbf{x}\colon U \to M = \mathbf{x}(U)$ and a simple closed (piecewise smooth) curve $\gamma \in M$. Consider the angle that $T = \gamma'$ makes with $\mathbf{x}_u$. Then we claim this changes by 2π as we go around γ. (Same convention as above for corners.)

Proof: pull everything back to U. If the metric g were standard, this would just be the Umlaufsatz. But we can continuously deform g to the Euclidean metric (through convex combinations) – an integer value must remain constant.

Let us define the total geodesic curvature $TC(\gamma)$ of a piecewise smooth curve γ as $\int \kappa_g\, ds + \sum \tau_i$, where the integral is taken over each smooth subarc. We have not yet formally defined integrals over regions in a surface, but for any coordinate chart $\mathbf{x}(U)$ and any region $D \subset U$, we have (just as for surface area)

$$\int_{\mathbf{x}(D)} K\, dA = \iint_D K\sqrt{\det g}\, du\, dv,$$

and can check that this is independent of coordinates. The local version of Gauss–Bonnet then says (for a piecewise smooth curve γ enclosing a disk $R \subset M$ to its left):

$$\int_R K\, dA = 2\pi - TC(\gamma).$$

Additivity under splitting a region.

Following do Carmo, we prove this directly under the assumption that R is contained in a orthogonally parametrized neighborhood $\mathbf{x}(U)$. (The case of larger disks is then a special case of one of the global versions of Gauss–Bonnet.) We write $D := \mathbf{x}^{-1}(R) \subset U$ and write α for the piecewise smooth boundary curve of D with $\gamma = \mathbf{x} \circ \alpha$.

Let us first examine the right hand side. By the extension to the Umlaufsatz, 2π is the total turning of the tangent vector relative to $\mathbf{x}_u$ (with the appropriate convention at corners). By definition, TC is the total turning of the tangent vector relative to a parallel field (with the same convention at corners). Thus $2\pi - TC$ is the total turning of $\mathbf{x}_u$ relative to a parallel field. Passing to the orthogonal unit vectors $e_1 := \mathbf{x}_u/\sqrt{E}$ and $e_2 :=$

$\mathbf{x}_v/\sqrt{G}$, we find that the rate of turning of $\mathbf{x}_u$ (relative to a parallel field) is

$$\langle \nabla_T e_1, e_2 \rangle = \frac{1}{\sqrt{EG}} \langle \nabla_T \mathbf{x}_u, \mathbf{x}_v \rangle.$$

With $T = \dot{u}\mathbf{x}_u + \dot{v}\mathbf{x}_v$, we have

$$\langle \nabla_T \mathbf{x}_u, \mathbf{x}_v \rangle = \dot{u}\Gamma_{112} + \dot{v}\Gamma_{122}.$$

But in orthogonal coordinates we computed $\Gamma_{112} = -E_v/2$ and $\Gamma_{122} = G_u/2$. Thus the total turning, the integral of the rate of turning, is

$$\int \frac{G_u \dot{v} - E_v \dot{u}}{2\sqrt{EG}}\, dt.$$

On the other hand, using our formula for K, we get

$$\begin{aligned} \int_D K\, dA &= \iint_R K\sqrt{EG}\, du\, dv \\ &= \iint_R \partial_v\Big(\frac{E_v}{2\sqrt{EG}}\Big) + \partial_u\Big(\frac{G_u}{2\sqrt{EG}}\Big)\, du\, dv. \end{aligned}$$

Now we recall Green's theorem in the plane: if P and Q are two functions on a region R bounded by a (piecewise smooth) curve α, then

$$\iint_R \partial_u Q - \partial_v P\, du\, dv = \int_\alpha P\, du + Q\, dv = \int_\alpha (P\dot{u} + Q\dot{v})\, dt.$$

(This is a special case of Stokes' Theorem, of course already known to Gauss in other forms.) We apply Green's theorem with $P = -E_v/2\sqrt{EG}$ and $Q = G_u/2\sqrt{EG}$ to give

$$\int_D K\, dA = \int_\alpha \frac{-E_v\dot{u} + G_u\dot{v}}{2\sqrt{EG}}\, dt.$$

This agrees with the expression we got above for $2\pi - TC$, so we have proved the Gauss–Bonnet theorem.

Interpretation in terms of holonomy of parallel transport. Gauss curvature as density – limit over small disks around p. Special case of sphere – spherical polygons (esp. triangles).

So far we have Gauss–Bonnet in the local form $\int_R K\, dA = 2\pi - TC(\partial R)$ for any disk R contained in an orthogonal coordinate patch and with peicewise smooth boundary. A comment on orientation is in order. Of course any coordinate patch is orintable. If we switch orientation, then K is unchanged, so the whole Gauss–Bonnet equation must be unchanged. Indeed, our convention that ∂R is oriented with R to the left depends on the surface normal, so T switches sign. That means however that the conormal η is unchanged, so the total geodesic curvature of ∂R is unchanged. A better way to express the orientation convention for ∂R might be to simply say the conormal η should point inwards towards R.

B18. Global topology of surfaces

To consider the global forms of Gauss–Bonnet, we need to discuss the topology of surfaces. A *regular region* R on a smooth surface will mean a compact subset R which is the closure of its interior and whose boundary is a finite disjoint union (possibly empty) of simple closed, piecewise smooth curves. Topologically, R is therefore a compact 2-manifold with boundary, a compact Hausdoff space locally homeomorphic to the closed half-plane $\{(x, y) : y \geq 0\}$. (That is, each point $p \in R$ has a neighborhood homeomorphic either to the plane or to the half-plane.)

One way to build such a topological space is to start with a finite collection of triangles and "zip" certain pairs of edges together. (Any edge that is not paired remains part of the boundary.) This is called a *triangulation* of the surface. A difficult theorem of Radó (1925) says that any topological 2-manifold with boundary can be triangulated. For our smooth regions R in space, this is not so difficult. Although we skip the details, the idea is that if we tile $\mathbb{R}^3$ with a fine enough cubic lattice – adjusted to be transverse to R – then each small cube contains just a single disk of R with piecewise smooth boundary: a polygon. Of course it is easy to cut an n-gon into $n-2$ triangles. Given an atlas of coordinate charts for a surface, note that we may assume the triangulation is fine enough that each triangle lies in one of the charts.

We now want to give a topological classification of regular regions R, that is, of compact 2-manifolds with boundary, that is, of spaces obtainable by zipping triangles together. For this discussion, we use the word *surface* to mean such a topological space (rather than a smooth surface in $\mathbb{R}^3$ as usual). We follow the description by Francis and Weeks of Conway's "ZIP proof".

Let us first describe the statement. To "perforate" a surface is to delete an open disk. (A sphere with one perforation is a closed disk; a sphere with two perforations is an annulus – also called a cylinder.) To add a "handle" or "cross-handle" to a surface is to perforate it twice, and then sew in an annulus connecting the new boundaries. (Or equivalently, then zip these boundaries to each other.) To add a "cross-cap" to a surface is to perforate it once and then sew in a Möbius band along the new boundary. (Or equivalently, then zip the two halves of the boundary together.)

Adding a cross-handle is the same as adding two cross-caps. (The Klein bottle is the union of two Möbius bands.)

Thm: Any surface is a finite union of components, each being a sphere with a certain number of perforations, handles, and cross-caps.

Pf: The starting collection of triangles is of this form. It suffices to show that a single zipping (involving one or two components) preserves this. First consider the case where entire boundary components are zipped together: either join two components or add a (cross-)handle. If the two edges to be zipped are instead the two halves of one boundary component, we add a cross-cap or remove a perforation. Finally, if the edges to be zipped are just subarcs of the cases considered so far, then the effect is the same except that we are left with one or two more perforations.

Note that adding a cross-cap (or a cross-handle) makes a component non-orientable. An orientable component is thus $\Sigma_{g,k}$, a sphere with $g \geq 0$ handles and $k \geq 0$ perforations. On a non-orientable component there is no way to distinguish handles from cross-handles. Thus it is has the form $N_{h,k}$, a sphere with $h \geq 1$ cross-caps and $k \geq 0$ perforations. We can restate the classification as follows.

Thm: A connected surface has the form $\Sigma_{g,k}$ if orientable or $N_{h,k}$ if nonorientable.

Note that the closed nonorientable surfaces $N_{h,0}$ cannot be embedded in $\mathbb{R}^3$, while all the other types can be.

The *Euler number* of a triangulation is $\chi := V - E + F$. This is clearly the sum of the Euler numbers of the components. For a single triangle $\chi = 1$. Zipping a pair of edges leaves F unchanged and decreases E by one; the effect on V varies. But tracing through the cases considered above shows that the Euler number depends only on the topology. Indeed $\chi(\Sigma_{g,k}) = 2 - 2g - k$ while $\chi(N_{h,k}) = 2 - h - k$. When thought of as an invariant of a topological space, χ is called the *Euler characteristic*.

Note that the topological type of a connected surface is thus determined by its orientability, its Euler number χ, and its number k of boundary components.

B19. Gauss–Bonnet: global forms

Recall that a regular region R on a smooth surface M in $\mathbb{R}^3$ is compact with piecewise smooth boundary. Every point in R has a neighborhood with an orthogonal parametrization, and R can be triangulated by triangles, each containted in such a parametrized neighborhood. To integrate a function f over R, we sum over the triangles:

$$\int_R f\,dA = \sum \int_{T_k} f\,dA = \sum \int_{\mathbf{x}_k^{-1}(T_k)} f\sqrt{\det g_k}\,du_k\,dv_k.$$

Note that our local form of Gauss–Bonnet applies to each triangle: $\int_T K\,dA = 2\pi - TC(\partial T)$.

Consider first a closed surface $R = M$. Counting edges of triangles gives $2E = 3F$, so $\chi(M) = V - F/2$. Now we sum the Gauss–Bonnet relation over all triangles. Each edge is used twice, with opposite orientations, so the terms $\int \kappa_g\,ds$ cancel out. Thus $\int_M K\,dA$ equals the sum over all triangles of $a + b + c - \pi$, where a, b, c are the interior angles. This is the sum of *all* interior angles minus πF. But grouping the angle sum around the vertices, it is $2\pi V$. Thus we get

$$\int_M K\,dA = 2\pi(V - F/2) = 2\pi\chi(M).$$

An alternative way of doing the bookkeeping is to start with total charge $2\pi\chi(M)$ by putting charges $+2\pi$ at each vertex and in each face and -2π on each edge. Then move the charges from the vertices and edges into the faces, based on angles and total curvatures. We are left with charge in each face, equal (by the local form of Gauss–Bonnet) to $\int K$.

For a general region R, we do the same thing. But along boundary edges, $\int \kappa_g\,ds$ does not cancel out. Similarly, at boundary vertices, the sum of interior angles is not 2π. Also $2E = 3F$ must be corrected by the number of boundary edges. Putting it all together, we get:

$$\int_R K\,dA = 2\pi\chi(R) - TC(\partial R).$$

This is the most general form of Gauss–Bonnet, with the previous versions (for disks and closed surfaces) as special cases. There are several immediate applications.

- Any closed surface in $\mathbb{R}^3$ with $K > 0$ has $\chi > 0$ so it is homeomorphic to a sphere or a projective plane; if it is embedded it must be a sphere.
- If there are two closed geodesics on a surface with $K > 0$, then they intersect (because otherwise they would bound an annulus with $\chi = 0$). (Note that Lyusternik and Shnirelman proved that any sphere has at least three different simple closed geodesics.)
- A simple closed geodesic on a surface with $K \leq 0$ cannot bound a disk to either side (because such a disk has $\int K = 2\pi$).
- There is no geodesic 1-gon or 2-gon (disk) on a surface with $K \leq 0$ (because geodesics that are tangent coincide, so the exterior angles are not π but strictly less).
- The angle excess of a geodesic triangle has the same sign as the average (or total) Gauss curvature in the triangle.

B20. The Gauss image and total absolute curvature

The Gauss–Bonnet theorem shows that K is a density – important is its integral over a region. There is also an extrinsic interpretation.

Near any point $p \in M$ where $K \neq 0$ the Gauss map $\boldsymbol{\nu}$ is locally an immersion $M \to \mathbb{S}^2$ – orientation-preserving if $K > 0$ and orientation-reversing if $K < 0$. Its Jacobian determinant $K = \det D\boldsymbol{\nu}$ gives the factor by which area is stretched – here we mean an "algebraic" signed area. (Think about curvature-line coordinates.)

We can say $\int_R K\,dA$ equals the signed area of $\boldsymbol{\nu}(R) \subset \mathbb{S}^2$. (And we can recover $K(p)$ as the limit of $\operatorname{area}(\boldsymbol{\nu}(R))/\operatorname{area}(R)$ as the region R shrinks down to the point p.)

Using the appropriate notion of area with multiplicities, we can say $\int_R K\,dA = \operatorname{area}(\boldsymbol{\nu}|_R)$ for any region $R \subset M$.

Consider a closed surface $M \subset \mathbb{R}^3$. An analog of the Jordan curve theorem says that it divides space into one bounded and one unbounded region; we can orient it with the outward unit normal (pointing into the unbounded region). Thus M is necessarily an orientable surface of some genus $g \geq 0$.

By Gauss–Bonnet, the area of the Gauss image $\boldsymbol{\nu}(M)$ in $\mathbb{S}^2$ equals $2\pi\chi = 4\pi(1 - g)$. That is, the Gauss map covers the sphere $1 - g$ times (in an oriented sense) – the *degree* of the Gauss map is $1 - g$. By degree theory, almost every point $w \in \mathbb{S}^2$ has finitely many preimages, $1 - g$ of them if counted with signs. That is, if there are k preimages with $K > 0$ then there are $k + g - 1$ with $K < 0$.

(Note that the points with normal $\boldsymbol{\nu} = \pm w$ are exactly the critical points of the function $\langle \cdot, w\rangle : M \to \mathbb{R}$.)

Now consider bringing a plane with given normal $w \in \mathbb{S}^2$ in from infinity until it first touches M. Any point of contact

$p \in M$ has $K \geq 0$, since M stays stays to one side of the plane. This means every point $w \in \mathbb{S}^2$ has at least one preimage with $K \geq 0$. (That is, $k \geq 1$ in the counts above.) Let $M_\pm$ denote the regions where $\pm K > 0$. Then we have $\int_{M_+} K\,dA \geq 4\pi$. With Gauss–Bonnet, this gives $\int_{M_-} K\,dA \leq -4\pi g$. Subtracting these gives $\int_M |K|\,dA \geq 4\pi(1+g)$.

Equality holds in these last three inequalities if and only if all points $p \in M$ with $K > 0$ are extreme on the convex hull. Such a surface is called *tight*. An interesting theory of tight surfaces characterizes them, for instance, as exactly those surfaces having the *two-piece property* of being cut into no more than two pieces by any plane.

Now consider the case that $K > 0$ holds on all of M. (So by Gauss–Bonnet, M is a sphere and $\int K\,dA = \int |K|\,dA = 4\pi$.) The Gauss map is bijective, indeed a diffeomorphism $M \to \mathbb{S}^2$. Each point $p \in M$ is an extreme point on the convex hull, that is, M is globally convex (as we mentioned before.

The results above on total absolute curvature do not hold for abstract surfaces. For instance, $\mathbb{R}^2/\mathbb{Z}^2$ is a torus with a flat ($K \equiv 0$) metric. An abstract surface of genus $g > 1$ can be given a *hyperbolic* metric with $K \equiv -1$. By Gauss–Bonnet its area is then $-2\pi\chi = 4\pi(g-1)$.

B21. Developable Surfaces

We now want to consider a surface $M \subset \mathbb{R}^3$ with $K \equiv 0$. Each point on M is parabolic – a closed subset $P \subset M$ consists of planar points; the open complement $U := M \setminus P$ consists of nonumbilic points. (Note the example of a triangle joined to three cylinders as in do Carmo; the example of a cylinder where P is, say, a Cantor set ($\times\mathbb{R}$); and the example of two cones where P is a single line at which the ruling is Lipschitz but not C^1 as in Kühnel.)

We follow a paper of Massey, as summarized in do Carmo's book.

At each $p \in U$ there is a unique asymptotic direction. Integrating these, we foliate U by a unique family of asymptotic lines. The first claim is that these curves are straight lines in space. Of course, along these asymptotic curves which are also lines of curvature, the surface normal $\boldsymbol{\nu}$ is constant. (But compare top curve of round torus – asymptotic=curvature line of parabolic points, but not straight – need to know conormal derivative of K vanishes.)

Proof: Locally on U we can use curvature-line coordinates where, say, the u-curves are the asymptotics. Then $\boldsymbol{\nu}$ is a function of v alone. Now consider the real-valued function $\langle \mathbf{x}, \boldsymbol{\nu}\rangle$ on U. Since $\mathbf{x}_u \perp \boldsymbol{\nu}$ and $\boldsymbol{\nu}_u \equiv 0$, its u-derivative vanishes, so it is a also function of v alone: $\langle \mathbf{x}, \boldsymbol{\nu}\rangle = \varphi(v)$. Differentiating gives $\langle \mathbf{x}, \boldsymbol{\nu}_v\rangle = \varphi'(v)$. Note that $\boldsymbol{\nu}_v \neq 0$ since we are at nonplanar points; of course $\boldsymbol{\nu}_v$ (like $\boldsymbol{\nu}$) is constant along each asymptotic curve. Thus each of these equations is the equation of some plane in space (depending on v); the planes are orthogonal. The u-curve $v =$ const. lies in the intersection line of these planes.

We next claim that we may assume u is the arclength parameter along each u-curve (asymptotic line).

Proof: Recall that in curvature-line coordinates with $g = \left(\begin{smallmatrix} E & 0 \\ 0 & G\end{smallmatrix}\right)$ and $h = \left(\begin{smallmatrix} L & 0 \\ 0 & N\end{smallmatrix}\right)$, the Codazzi equations become $L_v = HE_v$ and $N_u = HG_u$. Here we have $L \equiv 0$, so $E_v \equiv 0$, meaning that E is a function of u alone. Thus defining $s = \int \sqrt{E}\,du$ (independent of v), this is arclength along each u-curve. Then (s, v) are equally valid curvature-line coordinates, where $g = \left(\begin{smallmatrix} 1 & 0 \\ 0 & G\end{smallmatrix}\right)$ and $h = \left(\begin{smallmatrix} 0 & 0 \\ 0 & N\end{smallmatrix}\right)$.

The condition $E \equiv 1$ means that the new coordinates are not only curvature-line coordinates but also simultaneously geodesic parallel coordinates. As before we write $G = a^2$ (with $a = |\mathbf{x}_v|$) and have $K = -a_{uu}/a$. Thus here $a_{uu} = 0$, meaning that a is a linear function along each asymptotic line (that is, for each v).

Now consider again the Codazzi equation $N_u = HG_u = 2Haa_u$, where the mean curvature satisfies $2H = N/G = N/a^2$. Combining these gives

$$\frac{N_u}{N} = \frac{a_u}{a},$$

meaning that for each v (i.e., along each asymptotic curve), N is a constant multiple of a. That is $N = c(v)a$. Finally consider the principal radius of curvature $r = \frac{1}{2H} = \frac{a^2}{N} = \frac{a}{c(v)}$ in the $\mathbf{x}_v$ direction. This (like a and N) is a linear function along each asymptotic line.

Lemma: An asymptotic line through a point $p \in U$, even if extended indefinitely, never reaches P.

Proof: The mean curvature H is continuous on M. It vanishes on P but along any asymptotic line in U is the (nowhere vanishing) reciprocal of a linear function.

Prop: The boundary $\partial U = \partial P \subset M$ is a union of open line segments. (Note: might be infinite union!)

Proof: Consider a boundary point p and some neighborhood V parametrized as a graph over T_pM. The set $U \cap V$ is foliated by lines, which in the projection do not cross. Thus in a smaller neighborhood V' of p their directions are given by a Lipschitz function. Thus (consider the lines ℓ_{p_i} through points in $U \cap V'$ approaching p) there is a a well-defined limiting direction at p. Any point in V along the resulting line ℓ_p is a limit of points along the ℓ_{p_i} and in particular is in ∂U (but it cannot be in U because then all of ℓ_p would be).

Thm: A complete surface with $K = 0$ is a cylinder over some plane curve.

Proof: First we claim that the direction of the asymptotic lines is locally constant on U. Along any line, the radius of curvature is a linear function, which can never vanish on a smooth surface. Thus r is constant on each line. It follows from the equations above (in local coordinates) that $a = |\mathbf{x}_v|$ is also constant, implying that $\mathbf{x}_v$ is constant along each line. That is, $0 = \mathbf{x}_{vu} = \mathbf{x}_{uv}$. Thus $\mathbf{x}_u$ (which we know is constant along each line, of course) is locally constant as claimed.

At points of ∂U we defined lines whose direction was a limit; using the Lipschitz condition, we find the direction is actually locally constant on $U \cup \partial U$. (We don't get, say, a Cantor function!)

Finally consider the interior $P \setminus \partial U$ of P. It consists of open pieces of planes, bounded by complete lines. Each such piece must then be an infinite strip, bounded by two parallel lines. We can foliate it by further parallel lines.

Thus, as claimed, all of M is foliated in this way by parallel lines.