Distances, volumes, and integration

Scribe: Aric Bartle

1 Local Shape of a Surface

A question that we may ask ourselves is what significance does the second fundamental form play in the geometric characteristics of surface upon which it is defined. Let us consider a surface S and some point p on it. We may express locally at p the surface in terms of the graph over its tangent plane (from the homework). This means, that in a coordinate system with origin p, and axes $E_1, E_2 \in T_pS, E_3 = N_p$ with E_1, E_2 linearly independent, locally $S = \{(v^1, v^2, f(v^1, v^2)) : (v^1, v^2) \in \mathcal{V}\}$ for some sufficiently small $\mathcal{V} \subset \mathbb{R}^2$. This is equivalent to saying that we take S and translate it so that p is at the origin as well as rotating S so that tangent plane at p is coplanar to, say, the x, y axes. No matter which way we look at it, $f(0,0) = 0$, $\frac{\partial f(0,0)}{\partial x} = 0$, $\frac{\partial f(0,0)}{\partial y} = 0$. So without loss of generality, we may gain intuition of the second fundamental form by analysing it in this convenient frame.

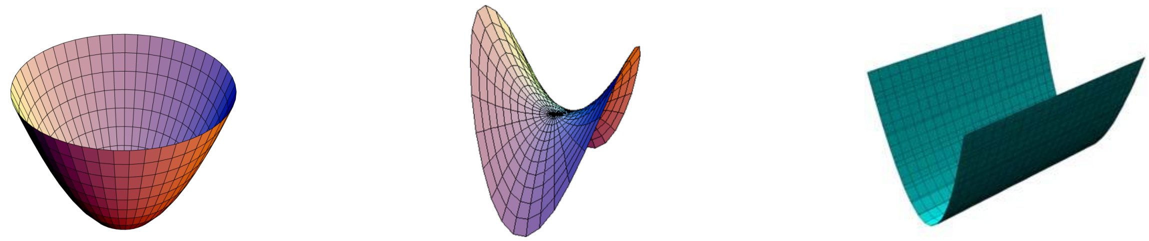
In such a frame, the second fundamental form at the origin becomes

$$A_{(0,0)} = \left.\begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}\right|_{(0,0)} \tag{1}$$

Of particular note, $A_{(0,0)}$ is the Hessian matrix. From Taylor's Theorem, we know that Hessian locally characterizes the surface at the origin. The Hessian or $A_{(0,0)}$ is a quadratic form, and so we may determine the shape based upon the eigenvalues of it:

- *Elliptic:* Both are greater than or less than zero
- *Hyperbolic:* One of each sign
- *Parabolic:* One is zero
- *Planar:* Both are zero
- *Umbilic:* Both are equal

Note, however, that if one or both of the eigenvalues are zero, then the surface locally need not be quadratic in the direction of the eigenvector corresponding to the zero eigenvalue. This follows from Talyor's Theorem in which if all terms up to and including the second degree term vanish, then the surface must be characterized by higher order terms. For instance, $h(x) = x^3$ vanishes at the origin through second degree. In other words, the second fundamental form tells us nothing about the surface in the direction of an eigenvector if the corresponding eigenvalue vanishes.



(a) Elliptic (b) Hyperbolic (c) Parabolic

Figure 1: The shapes of quadratic surfaces

2 Interpretation of Mean Curvature

We had previously defined the mean curvature as $k_{min} + k_{max}$ or $\mathsf{Tr}(A_p)$ w.r.t an orthonormal basis. Another way to express it is as the gradient of the surface area. This definition is a bit puzzling at first as what is the gradient of surface area. The surface area function takes as input a surface, S and gives as output a positive real number. As it turns out, the gradient of the surface area is the differential of the surface area as the input varies. To formalize this, let S_ϵ be the varied surface and $\phi : \mathcal{U} \to \mathbb{R},\ \phi_\epsilon : \mathcal{U} \to \mathbb{R}$ be parameterizations of the regular and varied surfaces, respectively. We would like a nice form of the parameterization ϕ_ϵ that is convenient to work with. Imagine starting at any point p on S and walking in the normal direction. At some distance, we will hit the surface S_ϵ and call this distance $\epsilon f(p)$, so, in fact, we can describe ϕ_ϵ as $\phi_\epsilon = \phi + \epsilon f \cdot N$ where ϵf can be thought of as the distance to the varied surface in the normal direction at a given point. The first variation of the area (or gradient of surface area) is:

$$\begin{aligned}
\frac{d}{d\epsilon}\mathsf{Area}(\phi_\epsilon(\mathcal{U}))\Big|_{\epsilon=0} &= \frac{d}{d\epsilon}\int_{\mathcal{U}} \sqrt{\det(g_\epsilon(u))}du\Big|_{\epsilon=0} \\
&= \frac{1}{2}\int_{\mathcal{U}} \mathsf{Tr}\Big(g^{-1}\frac{dg_\epsilon(u)}{d\epsilon}\Big|_{\epsilon=0}\Big)\sqrt{\det(g(u))}\,du \\
&= -\int_{\mathcal{U}} H(u)f(u)\sqrt{\det(g(u))}\,du
\end{aligned}$$

where $g_\epsilon(u) := [D\phi_\epsilon]_u^\top [D\phi_\epsilon]_u$ and $g := g_0$. The detailed derivation of this result has been deferred to the lecture supplement. Nonetheless, let us still gain an understanding of $H(u)$. It can be equivalently expressed as $\mathsf{Tr}([g(u)]^{-1}A(u))$ where $A(u)$ is the second fundamental form. Previously, we had learned that the trace of the second fundamental form yielded the mean curvature if expressed w.r.t. an orthonormal basis. But in this formula, the g^{-1} accounts for any non-orthogonality in a choice of basis for A. Form this it follows that the trace yields the mean curvature.

Now, if we wish to make the surface area decrease as fast as possible, we could, of course, just choose a large f, but if we restrict the norm of f, then we may instead choose $f(u) = H(u)$. We can see why this gives the best "direction" for f in Figures 1(a),(b). In Figure 1(a), to minimize the area of the surface we want to flatten out the surface (push the sides inwards). In fact, the mean curvature points inwards, following our intuition of how to flatten the surface. In Figure 2(a) there is a positive curvature and a negative curvature. We can imagine that if they are equal then the mean curvature would be zero, and the surface area would already be minimal. Indeed, such a surface minimizes area and is called a minimal surface. We can thus conclude that the mean curvature is the gradient of the surface area functional because it makes it decrease the fastest.

3 The Induced Metric and Covariance

Let $\phi : \mathcal{U} \to \mathbb{R}^3$ parametrize a surface S. The object

$$g := [D\phi_u]^\top D\phi_u \qquad \text{for } u \in \mathcal{U}$$

has appeared often, but what is the interpretation of g?

Definition: The object g is the induced metric of the surface.

To understand this is, set $E_i = \partial\phi/\partial u_i$ to be the tangent vectors of S at $\phi(u)$. Then, g is given by

$$g = [D\phi_u]^\top D\phi_u = \begin{bmatrix} E_1^\top \\ E_2^\top \end{bmatrix} [E_1 E_2] = \begin{bmatrix} \langle E_1, E_1\rangle & \langle E_1, E_2\rangle \\ \langle E_2, E_1\rangle & \langle E_2, E_2\rangle \end{bmatrix},$$

or in component form $g_{ij} = \langle E_i, E_j\rangle$. We may say that the induced metric represents the the Euclidean inner product of $\mathbb{R}^3$ restricted to $T_{\phi(u)}S$ (the Euclidean inner product of tangent vectors at p), pulled back to $\mathcal{U}$ by ϕ. Whenever you have a parameterization, you get a representation of the induced metric as a matrix of coefficients where the matrix of coefficients represents the inner product in the parameter space (the space of the coordinates u, v).

Now that we have defined the induced metric and given it an interpretation we will show that it is covariant. A scalar quantity defined on a surface S is “geometric” if its value computed w.r.t. any parametrization is always the same. A different property holds for vector and tensor quantities. For these quantities, the components of a “geometric” vector quantity computed w.r.t. two different parameterizations can be different. For instance, consider again the induced metric but computed by two different parameterizations of a surface. Each metric is given by the inner products of the tangent vectors for that parameterization. We can clearly imagine that the corresponding inner products of the two metrics would be different. Thus, for vector and tensor quantities, we must abandon our idea of invariance under a change of parameters. We instead say that a vector or tensor quantity is covariant if there exists a certain form of a transformation formula (defined below) to transform between the two vector or tensor quantities computed w.r.t. to the two different parameterizations.

Let $\phi : \mathcal{U} \to \mathbb{R}^3$ and $\psi : \mathcal{V} \to \mathbb{R}^3$ both parametrize S with $\phi(u) = \psi(v) = p \in S$. In addition,

- $e_i := [0 \dots 1 \dots 0]^\top$ are the standard basis vectors in $\mathcal{U}$.
- $f_i := [0 \dots 1 \dots 0]^\top$ are the standard basis vectors in $\mathcal{V}$.
- $E_i := \frac{\partial \phi}{\partial u^i} = D\phi_u \cdot e_i$ and $F_i := \frac{\partial \psi}{\partial v^i} = D\psi_v \cdot f_i$ are bases for T_pS.

We have then:

$$F_i = \frac{\partial \psi}{\partial v^i} = \frac{\partial \phi \circ \phi^{-1} \circ \psi}{\partial v^i} = \sum_j \frac{\partial [\phi^{-1} \circ \psi]^j}{\partial v^i} \frac{\partial \phi}{\partial u^j} = \underbrace{\sum_j \frac{\partial u^j}{\partial v^i}}_{A} E_j$$

And

$$\langle F_k, F_\ell \rangle = \left\langle \sum_i \frac{\partial u^i}{\partial v^k} E_i, \sum_j \frac{\partial u^j}{\partial v^\ell} E_j \right\rangle = \sum_{ij} \frac{\partial u^i}{\partial v^k} \frac{\partial u^j}{\partial v^\ell} g_{ij}$$

The A in the expression for F_i represents the change of basis matrix. We notice that this matrix is also present in the formula for the coefficient $\langle F_k, F_\ell \rangle$ in the induced metric. We say that the components of the induced metric change covariantly if when you change parameterizations (change bases), you get the change of basis matrix multiplied in the way seen for the transformation. In principal, whenever you define a vector quantity, you have to check that it is covariant. This is similar to showing that scalar quantities like arc-length do not change under different parameterizations.

4 Geodesics

To define geodesics, let us first consider the shortest path in a surface, S between two points p, q. We may give an equation satisfied by this shortest path: Let $\gamma : I \to S$ be the shortest path and γ_ϵ a variation with variation vector field V that is tangent to S. Then

$$0 = \frac{d}{d\epsilon} \mathsf{Length}(\gamma_\epsilon)\bigg|_{\epsilon=0} = \frac{d}{d\epsilon} \int_I \|\dot{\gamma}_\epsilon(t)\| \, dt \bigg|_{\epsilon=0} \quad \forall \text{ variations}, \tag{2}$$

and from the homework, this implies

$$0 = \langle \vec{k}_\gamma, V \rangle \quad \forall \text{ variations} \qquad \Leftrightarrow \qquad \vec{k}_\gamma \perp S. \tag{3}$$

Thus, a geodesic is any curve satisfying (3). We can now make a few comments about (2) and (3). Concerning (2), given a shortest path any small variation of this path with the same endpoints must result in a longer curve as γ is a minimum. Hence, $d\gamma_\epsilon/d\epsilon = 0$ at $\epsilon = 0$. As for (3), we may interpret as stating if a curve is a geodesic, then it has curvature vector orthogonal to the surface, and if the curvature vector of a curve is orthogonal to the surface, then it is a geodesic. It follows that if a curve is the shortest path, then it is a geodesic. But we would like to make the connection that if a curve is a geodesic, then it is a shortest path. Unfortunately, this is generally not as shown below. We shall prove a weaker statement that shows this statement is valid for a sufficiently small curve before proving a global statement about geodesics and length minimization. But first, we must introduce some machinery in order to accomplish these two goals.

The equation (3) is in fact a second-order ODE for γ. It follows from Existence and Uniqueness Theorem of Second-Order ODEs (EU) that given any initial position and velocity to this ODE there exists a unique solution, γ to the ODE within some local space. From this we can define the exponential map: $\exp_p : T_pS \to S$ such that $\exp_p(X) = \gamma(1)$ where $p \in S$, $X \in T_pS$, and γ the solution to the geodesic equation (3) with initial conditions $\gamma(0) = p$, $\dot{\gamma}(0) = X$. Furthermore, by the "uniqueness" part of EU, we can show that $\exp_p(tX) = \gamma(t), t \in [0,\ 1]$. We can interpret the exponential map as starting from p and integrating a unit of time along the geodesic γ to arrive at a new point on S. The geodesic is not arc-length parameterized, but at least the velocity of the geodesic is constant and proportional to the arc-length. A geodesic is constant speed from the fact that its acceleration is orthogonal to its velocity, implying that the velocity is constant. In addition, the distance traveled is proportional to X. In fact, X need be chosen so that $\gamma(1)$ still lies inside the space that EU guarantees that γ is the unique solution. At this point, we are still not quite done; we need to introduce two more results before for finally stating our theorem about short geodesics.

Proposition: It is easy to see that $[D \exp_p]_0 = id$. Hence, $\exp_p$ is a diffeomorphism near the origin in T_pM.

Proposition: ("Gauss lemma") Let $v, w \in T_v(T_pS)$. Then

$$\big\langle [D \exp_p]_v(v), [D \exp_p]_v(w) \big\rangle = \langle v, w \rangle$$

Theorem 1: *Geodesics locally minimize length: if γ is a sufficiently short geodesic and c is a curve with the same endpoints as γ, then $\mathsf{Length}(\gamma) \leq \mathsf{Length}(c)$ with equality if and only if $\gamma = c$.*

Proof: See the lecture supplement.

Let us now briefly review what we know about geodesics. Firstly, length-minimizing curves are geodesics. Secondly, short geodesics are length minimizing. Thirdly, there may exist long geodesics that are not length-minimizing. Indeed, it is easy to construct an example of a geodesic that is not length-minimizing. Consider a sphere and two points on it: p at the north pole and q somewhere else but not at the south pole. There are actually two geodesics that connect these two points. One that is simply the path of shortest distance between them and the other being the path that wraps all the way around the sphere.

Despite a seemingly hopeless situation, we can still provide a satisfying result by turning S into a metric space with distance function

$$d(p, q) := \inf_{\gamma \text{ from } p \text{ to } q} \mathsf{Length}(\gamma)$$

It can be show that this distance function is continuous, satisfies the triangle inequality, etc. for it to be metric space. We now turn to the Hopf-Rinow Theorem to finally give our result.

Theorem Hopf-Rinow: *If exp is globally defined then any two points p, q can be connected by a geodesic with length $d(p, q)$.*