GEODESICS AND THE HOPF-RINOW THEOREM

MINGHAN SUN

Abstract. In this paper, we define the notion of geodesics and prove the Hopf-Rinow Theorem, which states that any two points on a complete surface are joined by a minimal geodesic. Our main tool in the proof is the theory of exponential maps, which we will develop in detail.

Contents

1. Introductory Remarks

Throughout this paper, we assume that the reader has a solid understanding of basic differential geometry of curves and surfaces, and is familiar with ideas such as regular surfaces, first and second fundamental forms, the Gauss map, and principal and Gaussian curvatures.

Throughout this paper, we adopt the following conventions:

(1) We use the term "regular surface" to refer to a *connected* smooth 2-manifold embedded in $\mathbb{R}^3$. We commonly denote a regular surface by the letter S.

(2) We use the term "local parameterization" to denote a map $\mathbf{x} : U \to S$, where U is an open set in $\mathbb{R}^2$ and $\mathbf{x}$ is a diffeomorphism such that $d\mathbf{x}$ is nonsingular everywhere. Since S is a regular surface, for all $p \in S$, there exists a local parameterization $\mathbf{x}_p : U \to S$ such that for some $w \in U$, we have $\mathbf{x}(w) = p$. We call $\mathbf{x}_p$ a *local parameterization around* p.

(3) We use the term "smooth curve" to refer to a a connected smooth 1-manifold embedded in $\mathbb{R}^3$.

(4) We use $T_p(S)$ to denote the tangent plane of a regular surface S at point p.

Date: AUGUST 28TH, 2021.

(5) We use $B(x,r)$ to denote the open unit ball with radius r centered at point x of an Euclidean space.

2. Geodesics

This section aims to define what geodesics are. To do this, we first need to define the notion of a vector field and thence the notion of a "covariant derivative."

Definition 2.1. Suppose S is a regular surface and U is an open subset of S. A *vector field* w on U is a function which sends each $p \in U$ to a vector in $T_p(S)$. We shall denote this vector by $w(p)$.

Definition 2.2. Let U be an open subset of a regular surface S, and let w be a vector field on U. We say that w is *differentiable* at some $p \in U$ if the following statement holds. Let $\mathbf{x} : U \to S$ be a local parameterization for S around p. Then there exists some neighborhood W of p such that for all $q \in W$, if we write

$$w(q) = a(q)\mathbf{x}_u + b(q)\mathbf{x}_v, \tag{2.3}$$

for some functions $a(q)$ and $b(q)$, then both a and b are differentiable functions. Furthermore, we say that w is *differentiable* if it is differentiable at points in U.

Definition 2.4. Suppose that $p \in S$, that U is an open subset of S containing p, and that w is a vector field on U. Suppose further that $v \in T_p(S)$. Let $\alpha : (-\varepsilon, +\varepsilon) \to S$ be a smooth curve such that $\alpha(0) = p$ and $\alpha'(0) = v$. Let $\tilde{w} : (-\varepsilon, +\varepsilon) \to S$ be the restriction of the vector field w to the curve α, that is, for all $t \in (-\varepsilon, +\varepsilon)$, we have $\tilde{w}(t) = w(\alpha(t))$. Let $\frac{d\tilde{w}}{dt}(p)$ denote the usual vector derivative of $\tilde{w}$ at point p.

The *covariant derivative of w at p relative to the vector v* is defined as the projection of $\frac{d\tilde{w}}{dt}(p)$ onto $T_p(S)$. We denote this covariant derivative as $\left(\frac{Dw}{dt}\right)_v(p)$,

Remark 2.5. Very informally, the covariant derivative gives the *acceleration* of the vector field at a point in a certain direction, *as seen from the surface.* It is as seen from the surface because of the projection onto the tangent plane that we took in the definition above. We shall take up this point again in Remark 2.18.

This definition of the covariant derivative may seem slightly unintuitive at first, but the following two propositions, especially Proposition 2.7, should begin to shed light on its importance.

Proposition 2.6. *The covariant derivative is well-defined, that is, it does not depend on the choice of curve α that was made in Definition 2.4.*

Proposition 2.7. *The covariant derivative is intrinsic, that is, it depends only on the first fundamental form coefficients of a regular surface.*

Proof of Propositions 2.6 and 2.7. Our strategy is to derive a formula for $\left(\frac{Dw}{dt}\right)_v(p)$.

Let $\alpha : (-\varepsilon, +\varepsilon) \to S$ be a regular curve with $a(0) = 0$ and $\alpha'(0) = v$. Let $\mathbf{x} : U \to S$ be a local parameterization for S around p. Let $\phi : (-\varepsilon, +\varepsilon) \to U$ be a curve such that if we write $\phi(t) = (a(t), b(t))$, then we have $\alpha(t) = \mathbf{x}(a(t), b(t))$ for all $t \in (-\varepsilon, +\varepsilon)$. Since w is differentiable, for all $t \in (-\varepsilon, +\varepsilon)$, we can write

$$w(t) = f(a(t), b(t))\mathbf{x}_a + g(a(t), b(t))\mathbf{x}_b,$$

for some differentiable functions f and g.

We can now write down a formula for $\left(\frac{Dw}{dt}\right)_v(p)$. We will present this formula without proof, since the proof involves technical manipulations which the reader may lack some context to completely understand. The formula is as follows:

$$\begin{aligned}\left(\tfrac{Dw}{dt}\right)_v(p) &= (f' + \Gamma^1_{11} fa' + \Gamma^1_{12} fb' + \Gamma^1_{12} ga' + \Gamma^1_{22} gb')\mathbf{x}_a \\ &\quad + (g' + \Gamma^2_{11} fa' + \Gamma^2_{12} fb' + \Gamma^2_{12} ga' + \Gamma^2_{22} gb')\mathbf{x}_b\end{aligned} \tag{2.8}$$

The coefficients with a "Γ" are called the *Christoffel symbols*. The crucial fact is that they depend only on the first fundamental form coefficients of a regular surface. This proves that the covariant derivative is an intrinsic concept (Proposition 2.7). As a side remark, one proof of Gauss' Theorema Egregium involves writing the Gaussian curvature solely in terms of these symbols, thereby concluding it is intrinsic; the reader who is interested in this can peruse Chapter 4 of [1].

Also notice that α appears nowhere in (2.8): this means it does not matter what specific choice of the curve α we make, as long as it has the requisite properties as prescribed in Definition 2.4. Hence Proposition 2.6 follows. □

Definition 2.4 defines the covariant derivative of a vector field at a point with respect to a *direction*. This can be somewhat cumbersome to work with. Instead, we might as well first restrict the vector field to just a curve, and then define the covariant derivative on a curve without explicitly referring to a direction.

Definition 2.9. Let $\alpha : (0-\varepsilon, l+\varepsilon) \to S$ be a differentiable map. A *parameterized curve* $\alpha : [0,l] \to S$ is the restriction of ϕ to the closed interval $[0,l]$. We say that the curve α is *regular* if $\alpha'(t)$ is nonzero for all $t \in [0,l]$.

Notations 2.10. From now on, we shall refer to a parameterized regular curve as just a "regular curve," while we shall still refer to a parameterized curve that is not necessarily regular as a "parameterized curve."

Definition 2.11. Let $\alpha : [0,l] \to S$ be a continuous function. We say that α is *piecewise regular* if there exists a partition $[t_0 = 0, t_1, t_2, \cdots, t_N = l]$ of the interval $[0,l]$ such that α is regular on $[t_i, t_{i+1}]$ for all $0 \le i \le N-1$.

Notations 2.12. We say that a curve $\alpha : [0,l] \to S$ *joins* two points p, q on a regular surface S if $\alpha(0) = p$ and $\alpha(l) = q$.

Definition 2.13. Let $\alpha : [0,l] \to S$ be a parameterized curve. A *vector field* w on α is defined as a function which sends every $t \in [0,l]$ to a vector in $T_{\alpha(t)}(S)$. Let us denote this vector by $w(p)$.

Definition 2.14. A vector field w on a parameterized curve α is said to be *differentiable* at some $t_0 \in (0,l)$ if the following holds: if $\mathbf{x} : U \to S$ is a parameterization for S around p, then there exists some $\delta > 0$ such that for all $t \in (t_0 - \delta, t_0 + \delta)$, if we write

$$w(t) = a(t)\mathbf{x}_u + b(t)\mathbf{x}_v,$$

for some functions $a(t)$ and $b(t)$, then both a and b are differentiable functions. Furthermore, we say that w is *differentiable* if it is differentiable at points in $(0,l)$.

Now we can define the covariant derivative on a curve.

Definition 2.15. Let $\alpha : [0,l] \to S$ be a parameterized curve, and let $w : (0,l) \to S$ be a vector field on α. The covariant derivative of α at some $t \in [0,l]$ is defined as projection of $\frac{dw}{dt}$ onto $T_p(S)$. We denote the covariant derivative as $\frac{Dw}{dt}$, Now we have all we need to define what a geodesic is.

Definition 2.16. Let $\alpha : [0, l] \to S$ be a nonconstant regular curve. Suppose $w : (0, l) \to S$ be its tangent vector field, i.e. we have $w(t) = \alpha'(t)$ for $t \in (0, l)$. We say α is *geodesic* at some $t_0 \in (0, l)$ if we have

$$\tfrac{D\alpha'}{dt_0}|_{t=t_0} = 0.$$

We say that α is a *geodesic* if it is geodesic for all $t \in (0, l)$.

Remark 2.17. One might wonder what the notion of a geodesic is actually good for. At the end of Section 4, we shall prove the important result that geodesics are "locally length minimizing" in some precise sense. Loosely speaking, geodesics generalize the notion of straight lines, which minimize length in $\mathbb{R}^2$, to all regular surfaces.

Remark 2.18. The previous remark serves as a great motivation for the definition of geodesics. Back in Remark 2.5, we have noted that the covariant derivative is roughly the second derivative as seen from the surface. We observe that in $\mathbb{R}^2$, length-minimizers, which are straight lines, are characterized by having identically zero second derivative. So to generalize straight lines to a general regular surface, it is natural for us to define them as having acceleration zero, and crucially, this acceleration must be from from the perspective of the surface itself.

3. The Hopf-Rinow Theorem

In this section, we define what it means for a regular surface to be "geodesically complete" and then go on to state the Hopf-Rinow Theorem. The proof of the theorem will be deferred to Section 8, since we still need to develop some machinery needed for the proof.

Definition 3.1. A regular surface S is said to be *geodesically complete* if for all $p \in S$ and all geodesics $\alpha : [0, \varepsilon) \to S$ such that $\alpha(0) = p$, there exists some geodesic $\tilde{\alpha} : \mathbb{R} \to S$ such that $\tilde{\alpha}(t) = \alpha(t)$ for all $t \in [0, \varepsilon)$.

Remark 3.2. More informally, a regular surface is geodesically complete if every geodesic can be extended to a geodesic defined on the entire real line.

Example 3.3. Let us give an example of a geodesically incomplete surface. Consider the cone

$$C = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = z^2 \text{ and } z > 0\}. \tag{3.4}$$

From basic differential geometry, C is a regular surface. Consider the curve $\alpha : [0, 1) \to (1, 1, \sqrt{2}) - t(1, 1, \sqrt{2})$. Since α' is constant, the curve α is a geodesic. But α cannot be extended beyond $t \geq 1$, since it would have had to go throught the "vertex," which is not in the surface.

Definition 3.5. Suppose $\alpha : [0, l] \to S$ is a piecewise regular curve. Then there exists a partition $[t_0 = 0, t_1, t_2, \cdots, t_N = l]$ of the closed interval $[0, l]$ such that α is regular on $[t_i, t_{i+1}]$ for $0 \leq i \leq N - 1$. We define the *length* of α to be

$$l_\alpha = \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} |\alpha'(t)| \, \mathrm{d}t \tag{3.6}$$

Definition 3.7. Let $\alpha : [0, l] \to S$ be a geodesic joining $p, q \in S$. We say α is a *minimal geodesic* joining p and q if the length of α is less than or equal to the length of any other piecewise regular curve joining p and q.

At this point, we can state the main theorem of this paper, due to Heinz Hopf and Willi Rinow in 1931.

Theorem 3.8 (Hopf-Rinow, 1931)**.** *Suppose S is a geodesically complete regular surface, and let $p, q \in S$. There exists a minimal geodesic joining p and q.*

Again, the proof will be deferred until Section 8.

Example 3.9. Let us give an example of a geodesically incomplete surface that fails to admit a minimal geodesic joining all pairs of points. Consider the regular surface

$$S = \{(x, y, z) \in \mathbb{R}^3 : z = 0 \text{ and } 0 < x^2 + y^2 < 1\}. \tag{3.10}$$

S is not geodesically complete because the geodesic $\alpha : [\frac{1}{4}, \frac{1}{2}] \to S$ given by $\alpha(t) = (t, 0, 0)$ cannot be extended to all of $\mathbb{R}$. More precisely, it fails to extend beyond $t \geq 1$ or $t \leq 0$.

Consider the two points $(\frac{1}{2}, 0, 0)$ and $(-\frac{1}{2}, 0, 0)$. Since the origin is not in S, and the geodesics of S are all lines, no minimal geodesic can join these two points.

4. The Exponential Map

Now we begin to develop the machinery necessary for a proof of Hopf-Rinow Theorem. In this section, we define exponential maps, which will be our most important tool for working with geodesics.

We first need to establish some preparatory results.

Proposition 4.1. *Let S be a regular surface. For all $p \in S$ and for all $v \in T_p(S)$, there exists a unique geodesic $\alpha : [0, \varepsilon) \to S$ with $\alpha(0) = p$ and $\alpha'(0) = v$.*

Proof. We are looking for some function α that satisfies the ordinary differential equation

$$\tfrac{D\alpha}{dt} = 0 \tag{4.2}$$

with initial conditions

$$\alpha(0) = p \tag{4.3}$$

$$\alpha'(0) = v. \tag{4.4}$$

The existence of such a function follows directly from the existence and uniqueness of solutions to ordinary differential equations. □

Notation 4.5. A word on notation. Fix some $p \in S$ and some $v \in T_p(S)$. Let us denote the unique geodesic referred to in Proposition 4.1 as $\alpha(t, v)$, to indicate its dependence on v.

Observation 4.6. Suppose $\alpha(t, v) : (-\varepsilon, +\varepsilon) \to S$ is a geodesic. Then $\alpha(t, \lambda v) : (-\frac{\varepsilon}{\lambda}, +\frac{\varepsilon}{\lambda}) \to S$ is also a geodesic. Moreover, $\alpha(\lambda t, v) = \alpha(t, \lambda v)$ for all t and v.

We now define the exponential map.

Definition 4.7. Let $p \in S$ and $v \in T_p(S)$. If $\alpha(1, v)$ is defined, then we define

$$\exp_p(v) = \alpha(1, v).$$

In addition, we define $\exp_p(0) = p$.

This definition would be quite useless if the exponential map turns out to be ill-behaved. And at first sight, it is not at all clear that the exponential map would even be differentiable. Fortunately, we have the following fundamental result.

Theorem 4.8 (Fundamental Theorem of The Exponential Map)**.** *Let $p \in S$. There exists $\varepsilon > 0$ such that $\exp_p(v)|_{B(0,\varepsilon)}$ is a diffeomorphism from $B(0, \varepsilon)$ to $\exp_p(B(0, \varepsilon))$.*

The proof of this theorem and its corollaries deserve a separate section.

5. Proof of the Fundamental Theorem of Exponential Maps

To prove this theorem, we first need to show the weaker statement that $\exp_p(v)$ is defined and is differentiable in an neighborhood of 0 in $T_p(S)$.

Lemma 5.1. *Let $p \in S$. There exists $\varepsilon > 0$ such that if $v \in B(0, \varepsilon) \subset T_p(S)$, then $\exp_p(v)$ exists; moreover, $\exp_p(v)$ is differentiable at v.*

Remark 5.2. Proposition 4.1 says that for all $v \in T_p(S)$, there exists a unique geodesic α with $\alpha(0) = p$ and $\alpha'(0) = v$. Therefore, for v with sufficiently small norm, $\exp_p(v) = \alpha(1, v)$ is guaranteed to exist, because we have $\alpha(1, v) = \alpha(|v|, \frac{v}{|v|})$ by Observation 4.6. The difficulty, roughly speaking, is that even though we can find geodesics in all directions, their lengths might not have a positive lower bound, and hence the exponential map still may not be defined on any *open* neighborhood of zero in $T_p(S)$.

In order to establish Lemma 5.1, we must surmount the difficulty discussed in Remark 5.2. To do this, we need a result from the theory of ordinary differential equations, which we will state without proof.

Lemma 5.3. *Suppose $p \in S$. There exist $\varepsilon_1, \varepsilon_2 > 0$, and a differentiable map*

$$\alpha : (-\varepsilon_1, +\varepsilon_1) \times B(0, \varepsilon_2) \subset \mathbb{R} \times T_p(S) \to S \tag{5.4}$$

such that if we fix some $v \in B(0, \varepsilon_2)$, then $\alpha(t, v) : (-\varepsilon_1, +\varepsilon_1) \to S$ is the unique geodesic with $\alpha(0, v) = p$ and $\alpha'(0, v) = v$.

Proof of Lemma 5.1. We carry on the notations in Lemma 5.3. Let $\varepsilon = \frac{\varepsilon_1 \varepsilon_2}{2}$. We will show that $\exp_p$ is defined and differentiable in $B(0, \varepsilon) \subset T_p(S)$.

Suppose $u \in B(0, \varepsilon)$, and let $w = \frac{\varepsilon_1 u}{2}$. We observe that $|w| < \varepsilon_2$. By Lemma 5.3, there exists a geodesic $\alpha(t, w) : (-\varepsilon_1, +\varepsilon_1) \to S$ such that $\alpha(0, w) = p$ and $\alpha'(0, w) = w$. By Observation 4.6, the curve $\beta : (-\varepsilon_1, \varepsilon_1) \to S$ with $\beta(t) = \alpha(\frac{t}{\varepsilon_1/2}, \left(\frac{\varepsilon_1}{2}\right) w) = \alpha(\frac{t}{\varepsilon_1/2}, u)$ is also a geodesic. In particular, when $t = \frac{\varepsilon_1}{2}$, $\alpha(1, u) = \alpha(\frac{\varepsilon_1/2}{\varepsilon_1/2}, u)$ is a well-defined real number. Thus the exponential map is defined for the vector u. Since u is arbitrary, we conclude that the exponential map is defined on $B(0, \varepsilon)$.

Finally, the differentiability of $\exp_p$ follows from the differentiability of the function α given in Lemma 5.3. $\square$

Now we have all we need to prove Theorem 4.8.

Proof of Theorem 4.8. The key idea is to show that the exponential map $\exp_p$ is nonsingular at the origin of $T_p(S)$; if this is the case, we can apply the Inverse Function Theorem to finish the proof.

To show that $D(\exp_p)(0) \neq 0$, let us consider some unit vector $v \in T_p(S)$. We want to compute $D(\exp_p)(0)(v)$, which is the directional derivative of $\exp_p$ in the direction given by v.

To do this, we first let α be the map as defined in Lemma 5.3. By definition of the directional derivative, we have the following

$$\begin{aligned} D(\exp_p)(0)(v) &= \frac{d}{dt}\,\alpha(1, tv) \\ &= \frac{d}{dt}\,\alpha(t, v) \\ &= v \neq 0. \end{aligned} \tag{5.5}$$

Since v is arbitrary, we conclude that $D(\exp_p)(0) \neq 0$, and applying the Inverse Function Theorem, we obtain some neighborhood U of 0 in $T_p(S)$ such that $\exp_p$ restricted to U is a diffeomorphism. ☐

The fact that the exponential map is a local diffeomorphism has an important corollary, that geodesics are locally unique.

Corollary 5.6. *Suppose $p \in S$. Then there exists some open neighborhood U of p such that every $q \in U$ can be joined to p by a unique geodesic.*

Proof. By Theorem 4.8, we have some neighborhood V of p such that $\exp_p$ restricted to V is a diffeomorphism from V to its image.

Suppose $\gamma : [0, l] \to S$ and $\phi : [0, l] \to S$ are two geodesics joining p and q.

Let $\alpha : (-\varepsilon_1, +\varepsilon_1) \times B(0, \varepsilon_2) \to S$ be the map defined in Lemma 5.1. Consider two cases.

Case 1: Suppose $\gamma'(0) = \phi'(0)$. By Lemma 5.3, both γ and ϕ coincide with $\alpha(t, w)$, which is the unique geodesic joining p and q such that $\alpha'(0, w) = \gamma'(0) = \phi'(0)$.

Case 2: Suppose $\gamma'(0) \neq \phi'(0)$. By definition, we have $\gamma(l) = \phi(l) = q$. We define $v = \exp_p(q)$. Note that v is well-defined because $\exp_p$ is a diffeomorphism.

By Lemma 5.3, we can write $\gamma(t) = \alpha(t, \gamma'(0))$ and $\phi(t) = \alpha(t, \phi'(0))$. Hence, we have $q = \gamma(l) = \phi(l) = \alpha(l, \gamma'(0)) = \alpha(l, \phi'(0))$. In addition, by definition, we have $\alpha(l, \gamma'(0)) = \exp_p(\frac{\gamma'(0)}{l})$ and $\alpha(l, \gamma'(0)) = \exp_p(\frac{\phi'(0)}{l})$. But since $\phi'(0) \neq \gamma'(0)$, we have derived a contradiction, since we have assumed that $\exp_p$ is a diffeomorphism, and in particular, a bijection.

In both cases, the corollary has been proven. ☐

Remark 5.7. With the help of Corollary 5.6, we can justify our naming of "geodesically complete" surfaces. After all, why are they so *complete* anyways? It turns out that they are in a very precise sense "non-extendable." Let us formulate this rigorously below.

Definition 5.8. A regular surface S is said to be *extendable* if there exists some regular surface $\tilde{S}$ such that S is a proper subset of $\tilde{S}$. A regular surface S is said to be *non-extendable* if it is not extendable.

Proposition 5.9. *A geodesically complete surface is non-extendable.*

To prove this proposition, we need a lemma from basic differential geometry of curves and surfaces. We shall only sketch the proof and omit some of the topological details.

Lemma 5.10. *Let $\tilde{S}$ be a regular surface, and let S be a subset of $\tilde{S}$. Then S is a regular surface if and only if it is open in $\tilde{S}$.*

Proof Sketch. If S is open in $\tilde{S}$, then S is regular because it naturally inherits the local parameterizations of $\tilde{S}$.

Let ∂S denote the boundary of S in $\tilde{S}$. If S is not open in $\tilde{S}$, then there exists $x \in S \cap \partial S$. No open set in $\mathbb{R}^2$ can be homeomorphic to an open neighborhood of x in S, and therefore there exists no local parameterization of S near x. Hence S cannot be regular. □

Proof of Proposition 5.9. Let S be a geodesically complete surface. Suppose there exists a surface $\tilde{S}$ such that $S \subset \tilde{S}$ properly.

Let ∂S denote the boundary of S in $\tilde{S}$. Let $p \in \partial S$. We observe that $p \in \tilde{S} \setminus S$, since S is open in $\tilde{S}$. By Corollary 5.6, there exists a neighborhood U of p such that every $q \in U$ can be joined to p by a unique geodesic. Note that since $p \in \partial S$, there exists $q \in U$ such that $q \in S$.

Let $\alpha : [0, l] \to S$ be the unique geodesic joining p to q. Let $\gamma : [0, l] \to S$ be defined as

$$\gamma(t) = \alpha(l - t) \tag{5.11}$$

for all $t \in [0, l]$. Intuitively, γ is the reverse of α. Consider the restriction of γ to the interval $[0, l)$. Since $p \notin S$, the geodesic γ cannot be extended to all of $\mathbb{R}$, meaning that S is not geodesically complete. Hence, we have established the contrapositive of the proposition. □

6. Geodesics as Length-Minimizers

To fulfill our promise at the end of Section 2, we shall now prove that geodesics are locally length-minimizing, using exponential maps as our main tool. We first need several definitions.

Definition 6.1. Let $p \in S$, and let U be a neighborhood of the origin in $T_p(S)$ such that $\exp_p|_U$ is a diffeomorphism from U to $\exp_p(U)$. Note that the existence of U is guaranteed by Theorem 4.8. Let r be a positive real number such that $B(0, r)$ is contained in U. We call $\exp_p(B(0, r))$ the *geodesic circle* of radius r at p.

Notation 6.2. Suppose $\alpha : [0, l] \to S$ is a piecewise regular curve, and suppose $I \subset [0, l]$ is an interval. The notation "$l_{\alpha(I)}$" denotes the length of α when restricted to I.

Definition 6.3. Suppose $\alpha : [0, l] \to S$ is a piecewise regular curve. We call the set

$$\{\alpha(t) : t \in [0, l]\}$$

the *trace* of α. We denote this set as $\mathrm{tr}(\alpha)$.

With the above preparations made, we can proceed to state and prove our main result, that geodesics are locally length-minimizing.

Theorem 6.4. *Suppose $p \in S$. There exists some neighborhood W of p in S such that the following is true.*

Suppose that $\alpha : [0,l] \to S$ is a geodesic with $\alpha(0) = p$ and $\alpha(l) = q$ and that $\alpha(t) \in U$ for all $t \in [0,l]$. Let $\beta : [0,l] \to S$ be a piecewise regular curve with $\beta(0) = p$ and $\beta(l) = q$ ($\beta(t)$ does not necessarily lie in U for all $t \in [0,l]$). Then we have

$$l_\alpha \leq l_\beta, \tag{6.5}$$

where l_α denotes the length of α, and and l_β that of β.

Proof. We first assume that β is a regular curve, and we will pass to the general case at the end. We observe that since β is a piecewise regular curve defined on a compact interval, it must have finite length.

Let ε be a positive real number such that $\exp_p(B(0,\varepsilon))$ exists. To ease notation, let $W = \exp_p(B(0,\varepsilon))$. We will show that W is the neighborhood we seek in the theorem.

Let us consider two cases.

***Case 1**: β lies entirely within W, i.e. for all $t \in [0,l]$, we have $\beta(t) \in W$.*

Our main strategy is to put "polar coordinates" on $B(0,\varepsilon)$, and therefore on W.

We need some preparatory steps. We consider some real numbers t_1, t_2 with $0 < t_1 < t_2 < l$. We do this because here we are dealing with lengths of curves, and thus with the derivative of β. But β' is undefined at $t = 0$ and $t = l$. We can avoid this issue by doing calculations on the open interval (t_1, t_2), and then taking $t_1 \to 0$ and $t_2 \to l$.

Since $l_\beta < \infty$, we trivially have $l_{\beta([t_1,t_2])} < \infty$, and so we have that

$$l_{\exp_p^{-1}(\beta([t_1,t_2]))} < \infty \tag{6.6}$$

because $\exp_p$ is a diffeomorphism. Consequently, there exists some $v \in T_p(S)$ with $|v| = \varepsilon$ such that if we consider the curve $\gamma : (0,1) \to T_p(S)$ such that the trace of the curve $\gamma : (0,1) \to T_p(S)$ such that $\gamma(v) = tv$ intersects $\exp_p^{-1}(\beta([t_1,t_2]))$ at finitely many points. Consequently, if we let $R = \exp_p(L)$, then R intersects β at finitely many points as well. Denote these points as $p_1, \cdots, p_{N-1}$, and suppose we have $p_i = \beta(\tau_i)$, for some $\tau_1, \cdots, \tau_N$. We set $t_1 = \tau_0$ and $t_2 = \tau_N$.

Let us put our usual polar coordinates on $B(0,\varepsilon) \setminus L \subset T_p(S)$. Note that we have to exclude L because the angle component takes value in the open interval $(0, 2\pi)$).

With these coordinates, we can prove that geodesics are length-minimizing in a similar way as we can prove that a line is length-minimizing in the plane. For all $t \in (\tau_0, \tau_N)$, we can write

$$\beta(t) = \exp_p(\rho(t), \theta(t)), \tag{6.7}$$

for functions $\rho : (\tau_0, \tau_N) \to T_p(S)$ and $\theta : (\tau_0, \tau_N) \to T_p(S)$. With this in mind, we can find the length of $\beta((\tau_0, \tau_N))$, which is

$$\begin{aligned} l_{\beta(\tau_0,\tau_N)} &= \sum_{i=0}^{N} l_{\beta(\tau_i,\tau_{i+1})} \\ &= \sum_{i=0}^{N} \int_{\tau_i}^{\tau_{i+1}} |\beta'(t)| \, \mathrm{d}t \\ &= \sum_{i=0}^{N} \int_{\tau_i}^{\tau_{i+1}} \sqrt{G(\theta'(t))^2 + E(\rho'(t))^2} \, \mathrm{d}t, \end{aligned} \tag{6.8}$$

where E and G are the standard first fundamental form coefficients, i.e. $E = \langle \frac{\partial \exp_p}{\partial \rho}, \frac{\partial \exp_p}{\partial \rho} \rangle$ and $G = \langle \frac{\partial \exp_p}{\partial \theta}, \frac{\partial \exp_p}{\partial \theta} \rangle$. Note that the last step in (6.8) follows directly from applying the chain rule to (6.7).

Two remarks about the first fundamental form coefficients. First, Since $\mathbf{x}$ is a local parameterization, we must have $G \neq 0$. Second, by a theorem which we shall not prove here, for the polar coordinates we have introduced, we have $E \equiv 1$. The interested reader may find a proof for this Proposition 3, Section 4.6 of [1].

To finish proving *Case 1*, the crucial observation is this:

$$\begin{aligned} l_{\beta((\tau_0,\tau_N))} &= \sum_{i=0}^{N} \int_{\tau_i}^{\tau_{i+1}} \sqrt{G(\theta'(t))^2 + (\rho'(t))^2} \, \mathrm{d}t \\ &\geq \sum_{i=0}^{N} \int_{\tau_i}^{\tau_{i+1}} |\rho(t)| \, \mathrm{d}t \\ &= \sum_{i=0}^{N} |\rho(\tau_{i+1}) - \rho(\tau_i)| \\ &\geq |\rho(\tau_N) - \rho(\tau_0)| \, . \end{aligned} \tag{6.9}$$

Now letting τ_0 tend to 0 and let τ_N tend to l, we see that

$$l_\beta \geq |\rho(l) - \rho(0)| \, . \tag{6.10}$$

Let $u = \exp_p^{-1}(q)$. Let us define $\gamma : [0, l] \to S$ such that $\gamma(t) = (|tu/l| \, , \arg u)$ in polar coordinates. The factor of l is slightly awkward, but we still use it to maintain consistency. The key observation is that since α is a geodesic in W, by the uniqueness assertion in Corollary 5.6, we see that we can write $\alpha = \exp_p \circ \gamma$. So in polar coordinates, we have

$$\alpha(t) = \exp_p(|tu/l| \, , \arg u). \tag{6.11}$$

The key observation is that $\arg u$ is a constant. This allows us to do the same calculations in (6.8) and (6.9) for the curve α to conclude that

$$l_\alpha = |\rho(\tau_N) - \rho(\tau_0)| \, . \tag{6.12}$$

Thus we have proven that $l_\beta \geq l_\alpha$, which is what we want in *Case 1*.

Case 2: *β does not lie entirely within W.* This case is much simpler. First suppose γ is a geodesic connecting p and q. We shall show that $l_\beta \geq l_\gamma$.

Define T to be the smallest positive number such that $\beta(T)$ lies on ∂W. We trivially have

$$l_\beta \geq l_{\beta([0,T])}. \tag{6.13}$$

Define $v = \exp_p^{-1}(\beta(T))$. Define $\Gamma : [0,1] \to S$ such that $\Gamma(t) = \exp_p(vt)$. By definition, $\Gamma(1) = \exp_p(\exp_p^{-1}(\beta(T))) = \beta(T)$. By definition of the exponential map, Γ is a geodesic, and so by the argument in *Case 1*, we have

$$l_{\beta([0,T])} \geq l_\Gamma. \tag{6.14}$$

Finally, observe that $\beta(T) \in \partial W$ and $q \in W$, and so we have

$$l_\Gamma \geq l_\gamma. \tag{6.15}$$

Combining the three inequalities finishes the proof of *Case 2*.

Finally, we pass to the general case, where β is a *piecewise* regular curve. We can approximate β by a sequence of regular curves $(\beta_i)_{i\in\mathbb{N}}$, and by what has been proven we have for all i

$$l_\alpha \geq l_{\beta_i} \tag{6.16}$$

By continuity, we must have $l_\alpha \geq l_\beta$, which proves the theorem. □

Observation 6.17. Let S be a regular surface and $p \in S$. We have established three related results:

(1) The Fundamental Theorem of Exponential Maps (Theorem 4.8) gives us a neighborhood W_1 of p such that $\exp_p$ is a diffeomorphism from W_1 to $\exp_p(W_1)$.
(2) Corollary 5.6 gives us a neighborhood W_2 of p such that all $q \in W_2$ can be joined to p by a unique geodesic.
(3) Theorem 6.4 gives a neighborhood W_3 of p such that all geodesics in W_3 are length-minimizing.

We observe that by examining the proof of Corollary 5.6, one sees that we can always choose W_2 to be the same as W_1. In addition, by examining the proof of Theorem 6.4, one sees that we can always choose W_3 so that $W_3 \subset W_2$.

These observations, aside from providing some clarification, will come in handy in the proof of Proposition 7.7.

Remark 6.18. One might naturally wonder whether the converse of Theorem 6.4 is true, i.e. whether we have that if a piecewise regular curve α joining p and q has smaller length than any other curve joining p and q, then α is a geodesic. It turns out, perhaps slightly unexpectedly, that this is true! The following theorem gives a precise formulation.

Theorem 6.19. *Suppose $\alpha : [0,l] \to S$ is a piecewise regular curve joining p and q and is parameterized by arc length. Suppose for all $0 < t_1 < t_2 < l$, and for all piecewise regular curves β joining $\alpha(t_1)$ and $\alpha(t_2)$, we have*

$$l_{\alpha([t_1,t_2])} \leq l_\beta. \tag{6.20}$$

Then α is a geodesic.

Proof. We refer the reader to Proposition 2 of Section 4-7 of [1]. □

7. Metric Structure on Regular Surfaces

The Hopf-Rinow Theorem guarantees the existence of a *minimal* geodesic joining two points of a complete surface. Although we gave a definition for minimality in Section 2, that alone will not provide sufficient machinery for us to prove the theorem. Consequently, we need to develop a metric on a regular surface and establish some of its foundational properties.

Notation 7.1. Suppose $W \subset S$ is open. We call W a *coordinate neighborhood* for S if there exists a local parameterization $\mathbf{x} : U \to S$ such that $\mathbf{x}(U) = W$. The reader is likely to be familiar with this terminology, but some clarification is always helpful.

Proposition 7.2. *Let p, q be two points on the regular surface S. Then there exists a parameterized piecewise differentiable curve joining p and q.*

Proof. Since S is assumed to be connected, there exists a continuous map $\alpha : [0, l] \to S$ such that $\alpha(0) = p$ and $\alpha(1) = q$.

For each $t \in [0, l]$, let I_t be an open set containing t in $(0, i)$ such that $\alpha(I_t)$ lies in a single coordinate neighborhood of the regular surface S. By compactness of the trace of α, there exists a finite collection $(I_j)_{j=1}^N$ such that $(\alpha(I_j))_{j=1}^N$ covers $\mathrm{tr}(\alpha)$.

As a result, we can obtain a partition $[t_0, t_1, \cdots, t_m]$ of the interval $[0, l]$, where $t_0 = 0$ and $t_m = l$, such that $\alpha([t_i, t_{i+1}])$ is contained in a single coordinate neighborhood of S for all $0 \le i \le m - 1$.

We observe for all $0 \le i \le m - 1$, the point $\alpha(t_i)$ can be joined to $\alpha(t_{i+1})$ by a differentiable path γ_i. To prove this, suppose $\mathbf{x} : U \to S$ is a coordinate neighborhood such that $\alpha(t_i), \alpha(t_{i+1})$ are both contained in $\mathbf{x}(U)$. Suppose β is a differentiable path in U joining $\mathbf{x}^{-1}(\alpha(t_i))$ and $\mathbf{x}^{-1}(\alpha(t^{i+1}))$ (we can find such a path because U is simply a subset of $\mathbb{R}^2$). Let us define $\gamma_i = \mathbf{x} \circ \beta$, and γ is differentiable because both $\mathbf{x}$ and β are. This gives a differentiable path from $\alpha(t_i)$ to $\alpha(t_{i+1})$.

Applying the previous paragraph to all $0 \le i \le m - 1$, we find differentiable paths γ_i joining $\alpha(t_i)$ to $\alpha(t_{i+1})$. This gives a piecewise differentiable path joining $\alpha(t_0) = p$ to $\alpha(t_m) = q$, which is what we want. □

Definition 7.3. Let p, q be two points on the regular surface S. The *intrinsic distance* between p and q is defined as

$$\mathrm{d}(p, q) = \inf \mathrm{length}(\alpha), \tag{7.4}$$

where the infimum is taken over all parameterized piecewise differentiable curves α joining p and q.

Remark 7.5. Proposition 7.2 clearly guarantees that any two points on S have a well-defined distance.

Remark 7.6. It is not difficult to check that the distance function in Definition 7.3 satisfies these three properties: for all $p, q, r \in S$

(1) $\mathrm{d}(p, q) \ge 0$.
(2) $\mathrm{d}(p, q) = \mathrm{d}(q, p)$,
(3) $\mathrm{d}(p, q) + \mathrm{d}(q, r) \ge \mathrm{d}(p, r)$.

As a result, in order for d to be metric on S, we still have to establish one additional property.

Proposition 7.7. *For all $p, q \in S$, we have $d(p, q) = 0$ if and only if $p = q$.*

To prove this proposition, we first need a simple lemma.

Lemma 7.8. *Suppose $p \in S$, and let $\varepsilon > 0$ be sufficiently small that $\exp_p(B(0,\varepsilon))$ exists. Then for all $q \in \partial \exp_p(B(0,\varepsilon))$, we have $\mathrm{d}(p,q) = \varepsilon$.*

Proof. By Observation 6.17, there exists a unique minimal geodesic joining p and q. This geodesic must be given by $\alpha : [0,\varepsilon] \to S$ such that $\alpha(t) = \exp_p(vt)$, for some unit vector $v \in T_p(S)$, and so $l_\alpha = \varepsilon$. This proves the lemma. □

Proof of Proposition 7.7. From the definition of the distance function, it is trivial that if $p = q$, then $\mathrm{d}(p,q) = 0$.

Let us establish the other direction. By Theorem 6.4, there exists some $\delta > 0$ such that all geodesics that lie in $B(0,\delta) \subset T_p(S)$ are length-minimizing. By Observation 6.17, all $q \in B(0,\delta)$ can be joined to p by a unique geodesic.

Consider two cases.

Case 1: If $q \in B(0,\delta)$. Let γ be a geodesic joining p and q, and γ is by definition length-minimizing. Since by construction, γ is nonconstant, it must have positive length. This proves that $\mathrm{d}(p,q) > 0$.

Case 2: If $q \notin B(0,\delta)$. Again, let $\gamma : [0,l] \to S$ be a piecewise regular curve joining p and q. Observe that by the continuity of γ, the set $\gamma([0,l])$ must be connected. As a result, it must intersect $\partial B(0,\delta)$ at some point y. Let t_0 be the smallest positive number such that $\alpha(t_0) \in \partial B(0,\delta)$. By Lemma 7.8, we have $l_{\gamma([0,t_0])} = \delta$. This implies that $l_\gamma \geq \delta$, and therefore $\mathrm{d}(p,q) > 0$, as desired. □

Remark 7.9. Having defined the intrinsic metric, we can reformulate the definition of minimal geodesic (Definition 3.7) as follows: suppose $p, q \in S$, and $\alpha : [0,l] \to S$ is a geodesic joining p and q. We say that α is a *minimal geodesic* if $l_\alpha = \mathrm{d}(p,q)$.

The next proposition will be essential in the proof of the Hopf-Rinow Theorem.

Proposition 7.10. *Suppose $p \in S$ is fixed. Then the function $g : S \to \mathbb{R}_{\geq 0}$ given by $g(q) = \mathrm{d}(p,q)$ is continuous.*

Proof. Fix some $q \in S$. We can find some sufficiently small $\varepsilon > 0$ so that $\exp_p(B(0,\varepsilon))$ exists.

Our objective is to find some neighborhood of U of q such that if $w \in U$, then $|\mathrm{d}(p,q) - \mathrm{d}(p,w)| < \varepsilon$. We observe that $\exp_p(B(0,\varepsilon))$ is the desired neighborhood, because for all $w \in U$, we have

$$\mathrm{d}(w,q) < \varepsilon, \tag{7.11}$$

and so

$$|\mathrm{d}(p,q) - \mathrm{d}(p,w)| \leq \mathrm{d}(w,q) < \varepsilon, \tag{7.12}$$

which is what we want. □

Remark 7.13. This is a side remark. In Proposition 5.9, we proved that a geodesically complete surface is non-extendable. In fact, we can make this statement more general and more satisfying, as follows.

Theorem 7.14. *A regular surface S is geodesically complete if and only if it is complete with respect to the intrinsic metric.*

Proof. The proof is somewhat involved, and the interested reader is referred to Section 1.4 of [2]. □

8. Proof of Hopf-Rinow

Now we have all we need to prove the Hopf-Rinow Theorem. For clarity, let us first restate the theorem.

Theorem 8.1 (Hopf-Rinow, 1931)**.** *Suppose S is a geodesically complete regular surface. Let $p, q \in S$. There exists a minimal geodesic joining p and q.*

Proof. Let ε be a positive real number such that $\exp_p(B(0,\varepsilon))$ exists. To ease notation, denote this set by W.

Consider the function $D : \partial W \to \mathbb{R}_{\geq 0}$ such that $D(y) = \mathrm{d}(y, q)$ for all $y \in \partial W$. Since ∂W is compact, the function D has a minimum on ∂W, i.e. there exists some $x \in \partial D$ such that for all $y \in \partial W$, we have $D(x) \leq D(y)$.

Define $v = \frac{\exp_p^{-1}(x)}{\left|\exp_p^{-1}(x)\right|}$. We define a geodesic $\alpha : [0, \varepsilon] \to S$ given by

$$\alpha(t) = \exp_p(tv). \tag{8.2}$$

We note that $\alpha(\varepsilon) = x$. Since S is complete, we can extend the domain of definition of α to all of $\mathbb{R}$.

Let $r = d(p, q)$. If we can show that

$$\alpha(r) = q, \tag{8.3}$$

then we have proven the theorem, since α is a minimal geodesic connecting p and q. In fact, it is convenient to prove (8.3) by establishing something slightly stronger: we will show that for all $t \in [\varepsilon, r]$, we have

$$\mathrm{d}(\alpha(t), q) = r - t. \tag{8.4}$$

Let us prove (8.4) in two steps.

Step 1 Let us first show that $\mathrm{d}(\alpha(\varepsilon), q) = r - \varepsilon$.

We have $\mathrm{d}(\alpha(\varepsilon, q)) \geq r - \varepsilon$ because otherwise

$$\begin{aligned} \mathrm{d}(p, q) &\leq \mathrm{d}(p, x) + \mathrm{d}(q, x) \\ &< \varepsilon + (r - \varepsilon) \\ &= r, \end{aligned} \tag{8.5}$$

which is a contradiction.

We also have $\mathrm{d}(\alpha(t), q) \leq r - \varepsilon$. Suppose the contrary is true (i.e. $\mathrm{d}(\alpha(t), q) > r - \varepsilon$). Let γ be any piecewise regular curve joining p and q, and let t_0 be the smallest positive real number such that $\gamma(t_0) \in \partial W$. Then

$$\begin{aligned} l_\gamma &= l_{\gamma([0,t_0])} + l_{\gamma([t_0,1])} \\ &> \varepsilon + (r - \varepsilon) \\ &> r, \end{aligned} \tag{8.6}$$

which is again a contradiction. Note that in the first inequality above we used Lemma 7.8. Having shown that $\mathrm{d}(x, q) \geq r - \varepsilon$ and $\mathrm{d}(x, q) \leq r - \varepsilon$, we are done with Step 1.

Step 2 We will show that if we have $\mathrm{d}(\alpha(T), q) = r - T$ for some $T \in [\varepsilon, r]$, then there exists a sufficiently small $\delta > 0$ such that if $t \in (T, T + \delta)$, then $\mathrm{d}(\alpha(t), q) = r - t$. Moreover, the number δ is independent of T.

To prove this, we first let $z = \alpha(T)$. Let $\delta > 0$ be sufficiently small that $\exp_z(B(0, \delta))$ exists. I claim that δ is the constant we seek in *Step 2*.

To ease notation, denote $\exp_p(B(0,\delta))$ by W'. Again, since $\partial W'$ is compact, there exists some $x' \in W'$ such that for all $y' \in W'$, we have $\mathrm{d}(x', q) \leq \mathrm{d}(y', q)$.

Since we assumed that $\mathrm{d}(z, q) = r - T$, we can apply the exact argument in Step 1 to conclude that

$$\mathrm{d}(x', q) = r - T - \delta. \tag{8.7}$$

As a result, we have

$$\begin{aligned} \mathrm{d}(p, x') &\geq \mathrm{d}(p, q) - \mathrm{d}(q, x') \\ &= r - (r - T - \delta) \\ &= T + \delta. \end{aligned} \tag{8.8}$$

Let $w = \frac{\exp_z^{-1}(x')}{|\exp_z^{-1}(x')|}$. Define a piecewise regular curve β such that $\beta(t) = \alpha(t)$ for $t \in [0, T]$ and that $\beta(t) = \exp_z((t - T)w)$ for $t \in (T, T + \delta)$. By definition, we have $l_\beta = T + \delta$. Since $\mathrm{d}(p, x') \geq T + \delta$, Theorem 6.19 tells us that β is a geodesic, and hence regular at all points. By this regularity, and the fact that β and α agree when $t \in [0, T]$, we conclude that β and α in fact agree for all $t \in [0, T + \delta]$. As a result, we see that

$$x' = \beta(T + \delta) = \alpha(T + \delta). \tag{8.9}$$

Consequently, by (8.7), we see that

$$\mathrm{d}(\alpha(T + \delta), q) = r - T - \delta, \tag{8.10}$$

as desired. Thus we are done with Step 2.

To finish the proof, we observe that since the constant δ in Step 2 is independent of T, Steps 1 and 2 together imply (8.4), which proves the theorem. □

We give one immediate corollary of the Hopf-Rinow Theorem.

Corollary 8.11. *Let S be complete. Then for all $p \in S$, the exponential map $\exp_p : T_p(S) \to S$ is an onto map.*

Proof. Fix some $p \in S$. For all $q \in S$, we have a minimal geodesic α that joins p and q. Then we have

$$q = \exp_p(\mathrm{d}(p, q) \cdot \alpha'(0)), \tag{8.12}$$

and since q is arbitrary, we conclude that $\exp_p$ is onto. □

Corollary 8.11 has an important application.

Theorem 8.13. *If a regular surface S is geodesically complete and is bounded in the metric d, then it is compact.*

Proof. Fix some $p \in S$, and since S is complete, the corollary above says that $\exp_p$ is onto. Since S is bounded in the intrinsic metric, we can find some compact set $W \subset T_p(S)$ such that $\exp_p(W) = S$. Since W is bounded and $\exp_p$ is continuous, we conclude that S is compact, as desired. □

9. Generalizations to Hopf-Rinow

In this paper, we proved the Hopf-Rinow Theorem for 2-manifolds embedded in $\mathbb{R}^3$. Using quite similar techniques, we can generalize this theorem to an abstract Riemannian manifold of any dimension. The interested reader can refer to Chapter 1 of [2].

Furthermore, one may wonder whether the infinite dimensional case of the Hopf-Rinow Theorem is true. In 1975, Atkin proved that the answer is no; the interested reader with the necessary background can peruse his paper [4].

Acknowledgments

I sincerely and whole-heartedly thank my mentor, Adrian Chu, for all his guidance and insight. In addition, I thank the organizers, staff, and instructors of the REU program for providing me with this opportunity to have a wonderful summer of mathematics. I also thank my co-mentees, Chen Lin and Hakan Solak, for helping me resolve many points of confusion, and for their great humor and friendship throughout the summer.

References

[1] do Carmo, Manfredo P.. Differential geometry of curves and surfaces. Prentice Hall, 1976.
[2] Jost, Jürgen. Riemannian Geometry and Geometric Analysis. Springer, 2017.
[3] H. Hopf, W. Rinow, "Ueber den Begriff der vollständigen differentialgeometrischen Flächen" Comm. Math. Helv. , 3 (1931) pp. 209–225
[4] Atkin, C. J. (1975), "The Hopf–Rinow theorem is false in infinite dimensions", The Bulletin of the London Mathematical Society, 7 (3): 261–266