Geodesics

Peter Kagey

April 16, 2019

1 Preliminaries

The running intuition in this paper is to think of a creature who lives in the manifold walking in a "straight" line in some sense at a constant (nonzero) speed–in other words, the creature isn't accelerating.

In the case where $\mathbb{R}^n$, the notion of acceleration along a path at some point in time is easy to naively define, since the notion of adding velocity vectors at two different points along the curve can be done naively, and so the second derivative works in the typical sense

$$\gamma''(t) = \lim_{h \to 0} \frac{\gamma'(t+h) - \gamma'(t)}{h}.$$

However, one has to be much more careful when talking about acceleration in a general manifold. In particular, $T_{\gamma(t)}$ and $T_{\gamma(t+h)}$ are generally different tangent spaces, and the notion of adding their vectors is decidedly more subtle. In particular, if a manifold is isometrically embedded into Euclidean space via $f\colon M \to \mathbb{R}^n$, the velocity along the image of a curve is preserved, but the acceleration is not quite the same—instead, the acceleration in M corresponds to the acceleration in $f(M)$ of the projection of the acceleration to the tangent plane. However, capturing this notion intrinsically is more subtle.

Definition 1.1 (Connection)**.** Let $\pi\colon E \to M$ be a smooth vector bundle over M. Then a connection in E is any map that sends a tangent vector field and a section of E to another section of E

$$\nabla\colon \mathfrak{X}(M) \times \Gamma(E) \to \Gamma(E)$$

denoted $\nabla_X Y := \nabla(X, Y)$ and satisfying the following three properties:

(i) $\nabla_{f_1 X_1 + f_2 X_2} Y = f_1 \nabla_{X_1} Y + f_2 \nabla_{X_2} Y$,

(ii) $\nabla_X (a_1 Y_1 + a_2 Y_2) = a_1 \nabla_X Y_1 + a_2 \nabla_X Y_2$, and

(iii) $\nabla_X (fY) = f \nabla_X Y + \underbrace{(Xf)}_{\in C^\infty(M)} Y$.

Connections are generalizations of directional derivatives with respect to X, a tangent vector field.

Note 1.1. In the case where $\Gamma(E) = C^\infty(M)$, $\nabla(X, f) := Xf$ satisfies the above properties. In particular, ∇ can be thought of as a prescription for extending $X\colon C^\infty(M) \to C^\infty(M)$ to act on other vector bundles.

Example 1.1. Consider the case of $M = \mathbb{R}^3$ with the section of the tangent bundle given by $(x, 1, e^y) = X \in \mathfrak{X}(\mathbb{R}^3)$ and the rank 2 vector bundle $(x^2, y+z) = Y \in \Gamma(E)$. Then taking the derivative of Y along X yields

$$\nabla_X Y = \underbrace{\begin{bmatrix} \dfrac{\partial x^2}{\partial x} & \dfrac{\partial x^2}{\partial y} & \dfrac{\partial x^2}{\partial z} \\[2ex] \dfrac{\partial (y+z)}{\partial x} & \dfrac{\partial (y+z)}{\partial y} & \dfrac{\partial (y+z)}{\partial z} \end{bmatrix}}_{dY(x,y,z)} \underbrace{\begin{bmatrix} x \\ 1 \\ e^y \end{bmatrix}}_{X} = \begin{bmatrix} 2x^2 \\ 1 + e^y \end{bmatrix}.$$

The matrix multiplication perspective makes it clearer to see that in this case (i) ∇ is $C^\infty(\mathbb{R}^3)$-linear in the first term, because the column vector is $C^\infty(\mathbb{R}^3)$-linear, (ii) ∇ is $\mathbb{R}$-linear in the second term because both partial derivatives and matrices are $\mathbb{R}$-linear, and (iii) holds by doing the product rule entrywise in the Jacobian matrix $d(fY)(x, y, z)$.

These conditions capture, in some sense, the "essential qualities" of taking the directional derivative of a vector bundle over a given section of the tangent bundle.

Definition 1.2 (Covariant derivative along a curve)**.** Let ∇ be a connection in tangent bundle TM, and let γ be a curve in M. Then the **covariant derivative along γ** is the (unique) map

$$D_\gamma \colon \mathfrak{X}(\gamma) \to \mathfrak{X}(\gamma)$$

satisfying the following three properties

(i) $D_\gamma(aV + bW) = aD_\gamma V + bD_\gamma W$

(ii) $D_\gamma(fV) = f'V + fD_\gamma V$

(iii) $D_\gamma V(t) = \nabla_{\gamma'(t)}\widetilde{V}$ for every extension $\widetilde{V}$ of V.

Note 1.2. This definition is sensible, because (i) and (ii) correspond to (ii) and (iii) in the definition of a connection, and the last condition is roughly equivalent to (i) in the definition of a connection: that is, the covariant derivative is $C^\infty(M)$-linear in the first component, so it only depends on the vector field pointwise.

When on a Riemannian manifold, it is common to use a particular connection called the Levi-Civita connection, which is compatible with the metric in the expected way, and which has some other nice properties.

Definition 1.3 (Levi-Civita connection)**.** The **Levi-Civita connection** is a connection ∇ such that

(a) the metric is preserved under ∇, that is, $\nabla_{(-)}\, g \equiv 0$, and

(b) for any vector fields $X, Y \in (M)$, $\nabla_X Y - \nabla_Y X = [X, Y]$.

Note 1.3. When the Levi-Civita connection exists, it is unique.

Note 1.4. A connection that satisfies the second condition in the definition of the Levi-Civita connection is called"torsion-free", because the torsion is defined as

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

Definition 1.4 (Acceleration of a curve)**.** For every smooth map from an interval to the manifold, $\gamma\colon I \to M$, define the **acceleration of** γ as the vector field $D_\gamma\gamma'$ along γ.

Definition 1.5 (Geodesic)**.** A smooth curve γ is called a **geodesic** (with respect to ∇) if it's acceleration is zero: $D_\gamma\gamma' \equiv 0$.

Theorem 1.1 (Existence and uniqueness for geodesics)*.* Let M be a smooth manifold and ∇ a connection. Then for every $p \in M$, $v \in T_pM$ and $t_0 \in \mathbb{R}$, there exists an open interval $I \subset \mathbb{R}$ containing t_0 and a geodesic $\gamma\colon I \to M$ satisfying $\gamma(t_0) = p$ and $\gamma'(t_0) = v$.

Furthermore, any two geodesics with the same initial conditions agree on their common domain.

Proof idea. This result follows from moving to local coordinates in a way that respects the connection, and then making an appeal to uniqueness and existence for (second order) ordinary differential equations. ☐

Definition 1.6 (Geodesic extension)**.** Given a geodesic $\gamma\colon I \to M$, an **extension of γ** is a map $\widetilde{\gamma}\colon \widetilde{I} \to M$ such that

(i) I is a subset of $\widetilde{I}$,

(ii) the map $\widetilde{\gamma}$ is itself a geodesic, that is $D_{\widetilde{\gamma}}\widetilde{\gamma}' \equiv 0$, and

(iii) the map agrees on the original interval, that is $\widetilde{\gamma}|_I = \gamma$.

Note 1.5. A *proper* extension is an extension in which condition (i) is strengthened such that I is a *proper* subset of $\widetilde{I}$.

Note 1.6. A geodesic with no proper extension is called a **maximal** geodesic.

Definition 1.7 (Geodesically complete manifold)**.** A manifold is called geodesically complete (with respect to ∇) if every geodesic curve γ has an "all-time" extension to $\widetilde{\gamma}\colon \mathbb{R} \to M$.

Definition 1.8 (Closed geodesic)**.** A geodesic $\gamma\colon I \to M$ is called **closed** if there exist two points $t_1 \neq t_2 \in I$ such that $\gamma(t_1) = \gamma(t_2)$ and $\gamma'(t_1) = \gamma'(t_2)$.

Note 1.7. Every closed geodesic γ has an all-time extension $\widetilde{\gamma}\colon \mathbb{R} \to M$.

2 Properties of geodesics

An arbitrary manifold is not endowed with any "measuring tape" and creatures living on the manifold don't have any notion of their speed when they travel. Of course, it's natural to believe that having one gives the other—that is if our creature has a speedometer, it also has an odometer, and vice versa.

A Riemannian manifold comes equipped with an inner product on the tangent space at each point, and an inner product is equivalent to assigning a magnitude to each vector (by the polarization identity). That is, a Riemannian manifold is special because each tangent vector in TM has a notion of its magnitude or speed. This section will develop the natural way of turning the speed at each point to the length of a curve, and a natural way to define the distance between two points derived from the length of the *curves* between two points.

Definition 2.1. The **length of a vector** $v \in TM$ in a Riemannian manifold with inner product g is

$$||v||_g = \sqrt{g(v,v)} = \sqrt{\langle v, v\rangle}.$$

Definition 2.2. The **length of a** (piecewise differentiable) **curve** $\gamma\colon I \to M$ in a Riemannian manifold is

$$L(\gamma) = \int_\gamma ||\gamma'(t)||_g \, dt$$

Theorem 2.1. Let (M, g) be a Riemannian manifold. Then g induces a metric $d\colon M \times M \to \mathbb{R}$ on M, namely

$$d(a,b) = \inf \{L(\gamma) : \gamma \text{ is a path from } a \text{ to } b\}$$

Proof. In order to show that d is a metric, it is necessary and sufficient to show that

1. $d(a,b) = 0$ if and only if $a = b$.

   Suppose $a \neq b$. Then $d(a,b) > 0$ because of the exponential map.

   The other direction is much simpler. If $a = b$, the constant path $\gamma(t) = a = b$ has length 0, and the length of a path is strictly non-negative because g is positive definite.

2. $d(x,y) = d(y,x)$

   If $\gamma\colon (t_0,t_1) \to M$ is a path from a to b, then $\overline{\gamma}\colon (t_0,t_1) \to M$ defined by $\overline{\gamma}(t) = \gamma(t_1 + t_0 - t)$ is a path from b to a which has the same image and same length.

3. $d(a,b) \leq d(a,c) + d(c,b)$

   The set of pairs of paths from a to c and c to a is in length preserving bijection (via concatenation) with the set of paths from a to b through c, which is a subset of the set of paths from a to b.

□

The distance function on M is defined as the infimum under a (generally) uncountable number of paths, however in a geodesically complete (connected) manifold, it is equivalent to use the minimum over the paths instead of the infimum.

Theorem 2.2. In a geodesically complete Riemannian manifold, there exists a geodesic (with respect to the Levi-Civita connection) γ from (a, b) such that $L(\gamma) = d(a, b)$.

Since (M, g) is a metric space (M, d) via the metric induced by g, it is also a topological space $(M, \mathcal{O})$ induced by the metric.

Theorem 2.3. The metric topology (M, d) induced by g agrees with the manifold topology $(M, \mathcal{A})$.

3 Geodesically complete manifolds

We're interested in when the creature can travel forever in a "straight line" starting at any point and heading in any direction. Clearly the creature does not run into any trouble flying through space ($\mathbb{R}^3$) or walking on the surface of a sphere (S^2), but the creature clearly *does* run into trouble if it is on the surface of a disc (D^2) or if it is flying toward the origin, but that point has been removed ($\mathbb{R}^3 \setminus \{0\}$).

The Hopf-Rinow Theorem says that these two examples capture the extent of what can go wrong during the creature's journey. In particular, the notion of being able to fly through space is captured by the notion of a manifold being *geodesically complete*.

But before we can give a proper definition of geodesic completeness, we must first introduce the exponential map. The exponential map gives a way of precisely specifying where a point ends up after flowing in some tangent direction for a unit of time.

Definition 3.1 (Exponential Map)**.**

Definition 3.2. A connected Riemannian manifold (M, g) is called **geodesically complete** if each of its geodesics extends indefinitely, that is if for each $p \in M$ the geodesic exponential map $\exp\colon T_pX \to X$ is defined on the full tangent space at that point.

Definition 3.3 (Complete metric space)**.** A complete metric space is a metric space in which every Cauchy sequence is convergent.

Theorem 3.1 (Hopf-Rinow). Let (M, g) be a Riemannian manifold, and let (M, d) be the induced metric space. Then the following are equivalent:

1. (M, d) is metrically complete.
2. (M, g) is geodesically complete.
3. (M, g) is geodesically complete at some point.
4. Every closed, bounded set in M is compact.

And furthermore, between any two points $p, q \in M$, there exists a length-minimizing geodesic connecting these two points.

Proof. A proof will be sketched. □

4 Closed geodesics

Theorem 4.1. Let $\pi_1(M)$ be the fundamental group of M. Then each equivalence class of loops $[\alpha] \in \pi_1$ has a (closed) geodesic representative, in particular it contains a curve of minimal length which is a geodesic.

Proof idea. Aloizio Macedo summaries the proof as

> One way to see why this is true is via a Morse-theory approach, with a full proof being given in Klingenberg - Lectures on Closed Geodesics. The gist of it is to pick a loop close to the infimum on the class and flow it under (minus) the gradient flow of the energy functional in the free loop space (this has an obvious analogy with the finite-dimensional case when we flow via gradient and converge to a critical point).

□

Note 4.1. In a survey paper, Alexandru Oancea claims

> one of the first successes of the calculus of variations was to establish rigorously that such a minimizing procedure is effective and produces a closed geodesic. The situation is subtler if the manifold is simply connected, and the question was answered in the affirmative by Lyusternik and Fet in their celebrated 1951 paper

Theorem 4.2 (Lyusternik, Fet, 1951). Every compact Riemannian manifold (M, g) has at least one closed geodesic.

Proof. Lyusternik and Fet considered the energy functional on the loop space ΛM and showed that the topology of ΛM is complicated enough so that the energy functional must have critical points with nonzero energy (which are non trivial closed geodesics). □

Conjecture 4.1. Every compact Riemannian manifold of dimension greater than 1 contains infinitely many geometrically distinct non-constant closed geodesics.

Note 4.2. This particular conjecture is discussed in a report from the International Workshop on Geodesics in August 2010:

> How many closed geodesics must exist on a closed manifold? For surfaces, the answer is known: there are always infinitely many geometrically different closed geodesics. This is easily proved using Birkhoffs first argument when the fundamental group is infinite. The remaining cases of the sphere and the projective plane were settled by Bangert (1993) and Franks (1992). In higher dimensions Rademacher has shown that a closed manifold with a generic Riemannian metric admits infinitely many geometrically different closed geodesics Rademacher (1989).

The report gives a heuristic for why this is hard to prove:

> It is important to distinguish geometrically different geodesics from repetitions of the same geodesic. This distinction is difficult to make.

Note 4.3. As of (at least) 1999, it appears that

> The question of which Riemannian manifolds admit simple closed geodesics is still a mystery. It is not known whether all closed Riemannian manifolds contain simple closed geodesics.

Theorem 4.3 (Lyusternik, Schnirelmann (1929); Ballman (1978)). Every Riemannian manifold with the topology of a sphere has at least three simple closed geodesics.

Proof. Proof of theorem of three geodesics □

Note 4.4. This is classical result of M. Morse, who proved that the fourth periodic [not necessarily simple] geodesic becomes uncontrollably large for ellipsoids with distinct but very close semi-axes.k

In particular, a recent theorem shows an even stronger result: the lengths of all three geodesics on a sphere are bounded by the "size" of the manifold.

Definition 4.1. The **diameter** of a Riemannian manifold (M, g) is defined by

$$\mathrm{diam}(M) := \sup_{p,q \in M} d(p, q) \in \mathbb{R}_{\geq 0} \cup \{+\infty\}$$

Example 4.1. Consider the unit circle $S^1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ as a sub-manifold of $\mathbb{R}^2$ with the induced inner product. Then $\operatorname{diam}(S^1) = \pi$, with the distance between two antipodal points, say $p = (1, 0)$ and $q = (-1, 0)$, being half of the circumference of the circle (in the ordinary sense).

Example 4.2. Consider the torus $T^2 = \mathbb{R}^2/\mathbb{Z}^2$ with the inner product inherited from $\mathbb{R}^2$. Then the diameter of the torus $\operatorname{diam}(T^2) = \sqrt{2}/2$, with this distance achieved by $p = (0, 0)$ and $q = (\frac{1}{2}, \frac{1}{2})$.

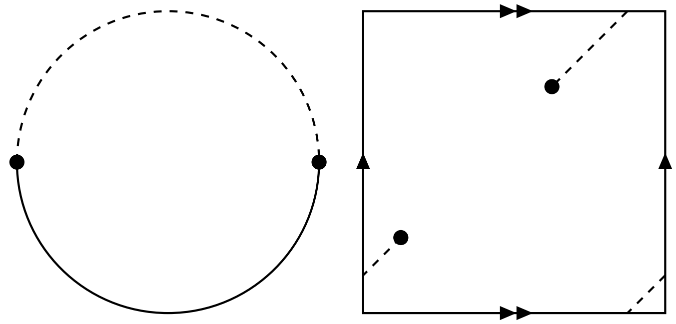

Figure 1: Illustrations of diameters on S^1 and T^2.

Theorem 4.4 (Liokumovich, Nabutovsky, Rotman, 2014). Let M be a Riemannian 2-sphere. Then there exist three distinct simple geodesics with lengths that do not exceed $20d$ where d is the diameter of M. Furthermore, if no simple closed geodesics of length $\leq 2d$, then there are three distinct simple periodic geodesics on M with lengths $\leq 5d$, $10d$, and $20d$ respectively.

Proof idea. The above authors give their proof idea as follows:

> The main idea of the proof of [the above theorem] is to express three homology classes of the space of non-parametrized curves that are used in classical proofs of the Lyusternik-Schnirelmann theorem by cycles that consist of simple closed curves "mainly made" of curves in a meridian-like family that connects two fixed points of M. [...] We attempt to construct such a family where the lengths of all meridians are bounded by const d for an appropriate const. Our repeated attempts can be blocked only by appearance of different short simple periodic geodesics of index 0. So, we either get three short simple periodic geodesics of index 0, or our third attempt to construct a meridional slicing succeeds. Once one of our attempts succeeds, and we get a slicing of M into short meridians, the original proof of the Lyusternik-Schnirelmann theorem yields the desired upper bounds.

□