Path Space and Variation of Energy*

Hangjun Xu

October 23, 2011

1 Motivation and Background

The notion of geodesics on a surface was first introduced by Johann Bernoulli, who defined it as the curve of minimal length on the surfaces between any two of its points. The **local** theory of geodesics was well understood due to the work of geometers on surfaces in the nineteenth century (Gauss, Jacobi, Bonnet), and later on Riemannian manifolds (Riemann, Christoffel, Levi-Civita). For the **global** aspects of geodesics on a Riemannian manifold M of dimension n, there are two main problems: (1) Does an arc of a geodesics with end points p, q actually minimize length among all smooth curves joining p and q? (2) How many geodesic arcs are there joining any two points of M? Locally, these two problems have complete answers: each point of M has an open neighborhood U such that for any distinct points p, q in U there is a unique arc of geodesic contained in U joining p and q, and it is unique *minimal* geodesic between p and q.
Until 1920, the only general results on the global problems came from Jacobi's work on problem (1). He had shown that on a geodesic γ with starting point x_0, there exists in general a sequence of points $x_1, x_2, \cdots$, known as the conjugate points of x_0[1], such that any arc C of γ that does not contain any of the $x_j's$ for $j \geq 1$ is a minimal geodesic arc; but if C does contain any x_j, say p and q are the endpoints of C, then in *every* neighborhood of γ, there exist piecewise smooth arcs joining p and q with length strictly smaller than

*This is a talk I gave at the Morse Theory Seminar, Fall 2011, Duke

[1] The problem of finding those conjugate points explicitly on a general Riemannian manifold is still open today.

the length of C. In other words, γ will stay minimizing until it reaches a conjugate point; and once it passes through a conjugate point, it will again be minimizing until it reaches the next conjugate point. The sphere with conjugate points the north pole and the south pole is a very good example to illustrate this.

Beginning in 1928, Marston Morse published a series of papers attacking this problem by a bold combination of differential geometry and algebraic topology applied to suitable function spaces, developing a technique he called "calculus of variation in the large". Given a Riemannian manifold (M, g), and two points p and q on M, he considered the space of all piecewise smooth paths on M joining p and q: $\Omega(M; p, q) = \Omega$. If $\gamma \in \Omega(M; p, q)$, we can always reparametrize so that γ is defined on $[0, 1]$, and there exists a partition of $[0, 1] : 0 = t_0 < t_1 < \cdots < t_k = 1$ such that $\gamma|_{[t_i, t_{i+1}]}$ is smooth for $i = 0, \cdots, k-1$. Morse then considered real-valued function F defined on Ω and studied the critical points of F, and tried to related that with the global problem of geodesics. For example, we can let F be the energy functional E. The energy E of $\gamma \in \Omega(M; p, q)$ from a to b where $0 \leq a < b \leq 1$ is defined to be:

$$E_a^b(\gamma) := \int_a^b \langle \frac{d\gamma}{dt}, \frac{d\gamma}{dt} \rangle_g dt.$$

We will make two conventions: (1) We let $E(\gamma)$ to denote $E_0^1(\gamma)$; (2) We will drop the metric g whenever it is clear that with respect to which metric the inner product is taken. Another example is to take F to be the length functional defined in a similar manner as

$$L_a^b(\gamma) := \int_a^b \sqrt{\langle \frac{d\gamma}{dt}, \frac{d\gamma}{dt} \rangle}\, dt.$$

The remarkable contribution of Morse which was the essence of his theory was that it is possible to give Ω a differentiable structure so that it is meaningful to talk about smooth functions on Ω and their critical points, hence his general results on critical points could be applied. The presentation that we are studying today was simplified by Bott, and was further simplified by Milnor.

2 Step I: First Variation of Energy

Applying Schwarz's inequality $\left(\int_a^b fgdt\right)^2 \leq \left(\int_a^b f^2 dt\right)\left(\int_a^b g^2 dt\right)$ with $f(t) = 1$ and $g(t) = \sqrt{\langle \frac{d\gamma}{dt}, \frac{d\gamma}{dt}\rangle}$, we see that

$$(L_a^b)^2 \leq (b-a)E_a^b,$$

where equality holds if and only if g is constant; that is if and only if γ is parametrized proportional to arc length. Now suppose there is a minimal geodesic $\widetilde{\gamma}$ from p to q, then we have

$$E(\widetilde{\gamma}) = L(\widetilde{\gamma})^2 \leq L(\gamma)^2 \leq E(\gamma).$$

Note that $L(\widetilde{\gamma}) = L(\gamma)$ if and only if γ is also a minimal geodesic, possibly reparametrized. On the other hand $L(\gamma)^2 = E(\gamma)$ if and only if γ is parametrized proportional to arc length. Therefore $E(\widetilde{\gamma}) < E(\gamma)$ unless γ is also a minimal geodesic. Suppose M is complete, then any two points p, q can be joined by a minimal geodesic. Therefore the energy functional $E : \Omega \longrightarrow \mathbb{R}$ achieves its minima precisely on the set of minimal geodesics from p to q. Now we have two questions: (1) How can we detect those minimal geodesics using E? (2) Can we at least detect the geodesics from p to q using E?

Definition 1. *Given a path $\gamma \in \Omega(M; p, q)$, an end point fixed variation α of γ is a continuous map $\alpha : (-\epsilon, \epsilon) \times [0, 1] \longrightarrow M$ with the following properties:*
(1) $\alpha(0, t) = \gamma(t)$,
(2) $\alpha(u, 0) = p, \alpha(u, 1) = q$ for all $-\epsilon < u < \epsilon$, and
(3) there is a partition of $[0, 1] : 0 = t_0 < t_1 < \cdots < t_k = 1$ such that α is smooth on each $(-\epsilon, \epsilon) \times [t_i, t_{i+1}]$, $i = 0, \cdots, k-1$.

A *variational vector field* $t \mapsto W(t)$ is associated to each variation α, where $W(t)$ is a tangent vector in the tangent space $T_{\gamma(t)}M$, defined as

$$W(t) := \frac{\partial \alpha}{\partial u}(0, t).$$

$W(t)$ is a continuous map from $[0, 1]$ into the tangent bundle TM, and is smooth in each $[t_i, t_{i+1}]$, for $i = 0, \cdots, k-1$. All these piecewise smooth vector fields along γ vanishing at end points form a vector space, $T\Omega(\gamma)$, which can be naturally thought as the tangent space of $\Omega(M;p,q)$ at the path γ. This is an infinite dimensional vector space. Conversely, given a piecewise smooth vector field along γ, we can realize it as the variational vector field of some variation α, using exponential maps. Therefore there is a one-to-one correspondence:

$$\{\text{tangent vectors in } \ T\Omega(\gamma)\} \longleftrightarrow \{\text{variational vector fields along } \gamma\}$$

More generally, one can replace $(-\epsilon, \epsilon)$ in the above definition with a neighborhood of 0 in some $\mathbb{R}^n$, defining an *n-parameter variation.* We can also talk about a variation without end points fixed, and the discussions below follow with possible slight modifications to take care of the end points.
Let $F : \Omega \longrightarrow \mathbb{R}$ be a real-valued function. We wish to define a map

$$F_*|_\gamma : T\Omega(\gamma) \longrightarrow T_{F(\gamma)}\mathbb{R}.$$

How do we define F_*? Naively, we could pick a curve in Ω that starts at γ, which is exactly a variation α of γ, then pick a tangent vector of that curve, which corresponds to the variational vector field $W(t)$. Thus we would like to define

$$F_*|_\gamma(W(t)) := F_*|_\gamma(\frac{\partial \alpha}{\partial u}(0,t)) = \frac{d}{du}|_{u=0}\big(F \circ \alpha \ (u,t)\big).$$

We will not discuss the conditions that F must satisfy in order for $F \circ \alpha$ to be differentiable, we only indicate how F_* could be defined to motivate the notion of critical points of F.

Definition 2. *A path γ_0 is a critical path or a critical point of the functional F if and only if $\frac{d}{du}|_{u=0}(F(\alpha(u,t))) = 0$ for every variation α of γ_0.*

Example 1. *Suppose F achieves a minimum at a path $\gamma_0 \in \Omega$, and if the derivatives $\frac{d}{du}F(\alpha(u,t))$ are all defined, then γ_0 is a critical point of F.*

Suppose M is a complete manifold. If F is the energy functional E we talked about earlier, then from the above we know that the minima of E on Ω are precisely the minimizing geodesics joining p to q. Therefore every minimal geodesic in $\Omega(M;p,q)$ is a critical point of E. The question is: are critical points of E also minimal geodesics? If so, then this will give us an answer to the previously raised questions about using E to detect minimal geodesics. It turns out that this is a bit too much to ask for.

Theorem 1 (First Variation Formula). *Let $\gamma : [0,1] \longrightarrow M$ be a path in $\Omega(M;p,q)$, and let $\alpha : (-\epsilon,\epsilon) \times [0,1] \longrightarrow M$ be a end points fixed variation of γ. Let $W(t)$ denote the variational vector field, and $V(t) = \frac{d\gamma}{dt}(t)$ be the velocity vector field along γ. Let $A(t) := \nabla_{\frac{\partial}{\partial t}} V(t)$ be vector field along γ that is the covariant derivative of $V(t)$. Finally, let $\Delta_t V = V(t^+) - V(t^-)$ be the discontinuity in the velocity vector at t where $0 < t < 1$. Then we have*

$$\frac{1}{2}\frac{d}{du}|_{u=0} E(\alpha(u,t)) = -\sum_t \langle W(t), \Delta_t V\rangle - \int_0^1 \langle W(t), A(t)\rangle dt.$$

Proof. Choose $0 = t_0 < t_1 < \cdots < t_k = 1$ so that α is differentiable on each $(-\epsilon,\epsilon) \times [t_{i-1}, t_i]$, $i = 1, \cdots, k$. By metric compatibilitiy, we have

$$\frac{\partial}{\partial u}\langle \frac{\partial \alpha}{\partial t}, \frac{\partial \alpha}{\partial t}\rangle = 2\langle \nabla_{\frac{\partial}{\partial u}} \frac{\partial \alpha}{\partial t}, \frac{\partial \alpha}{\partial t}\rangle.$$

Therefore

$$\begin{aligned}
\frac{1}{2}\frac{d}{du} E(\alpha(u,t)) &= \frac{1}{2}\frac{d}{du}\int_0^1 \langle \frac{\partial \alpha}{\partial t}, \frac{\partial \alpha}{\partial t}\rangle dt \\
&= \int_0^1 \langle \nabla_{\frac{\partial}{\partial u}} \frac{\partial \alpha}{\partial t}, \frac{\partial \alpha}{\partial t}\rangle dt \\
&= \int_0^1 \langle \nabla_{\frac{\partial}{\partial t}} \frac{\partial \alpha}{\partial u}, \frac{\partial \alpha}{\partial t}\rangle dt \\
&= \int_0^1 \left(\frac{\partial}{\partial t}\langle \frac{\partial \alpha}{\partial u}, \frac{\partial \alpha}{\partial t}\rangle - \langle \frac{\partial \alpha}{\partial u}, \nabla_{\frac{\partial}{\partial t}} \frac{\partial \alpha}{\partial t}\rangle \right) dt \\
&= \sum_{i=1}^k \int_{t_{i-1}}^{t_i} \left(\frac{\partial}{\partial t}\langle \frac{\partial \alpha}{\partial u}, \frac{\partial \alpha}{\partial t}\rangle - \langle \frac{\partial \alpha}{\partial u}, \nabla_{\frac{\partial}{\partial t}} \frac{\partial \alpha}{\partial t}\rangle \right) dt \\
&= \sum_{i=1}^k \langle \frac{\partial \alpha}{\partial u}, \frac{\partial \alpha}{\partial t}\rangle |_{t_{i-1}^+}^{t_i^-} - \sum_{i=1}^k \int_{t_{i-1}}^{t_i} \langle \frac{\partial \alpha}{\partial u}, \nabla_{\frac{\partial}{\partial t}} \frac{\partial \alpha}{\partial t}\rangle dt \\
&= -\sum_{i=1}^{k-1} \langle \frac{\partial \alpha}{\partial u}(t_i), \Delta_{t_i} \frac{\partial \alpha}{\partial t}\rangle - \int_0^1 \langle \frac{\partial \alpha}{\partial u}, \nabla_{\frac{\partial}{\partial t}} \frac{\partial \alpha}{\partial t}\rangle dt
\end{aligned}$$

Setting $u = 0$, we have the desired formula. $\square$

Intuitively, in order to decrease energy, we should vary the path γ in the direction of decreasing "kink" (see diagram below); and in the direction of the acceleration vector $A(t)$.

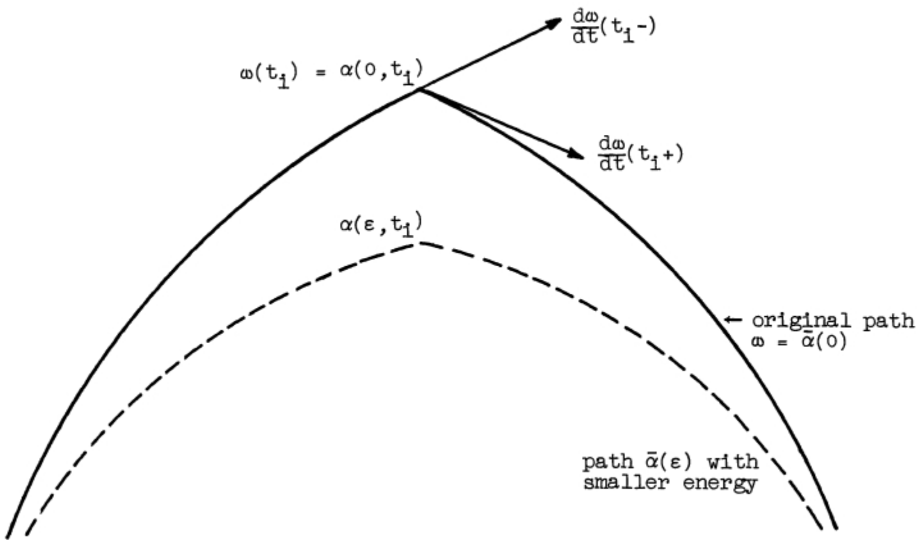

Figure 1: Picture Taken from Milnor's Book [2]

Corollary 2. *A path γ_0 is a critical point for E if and only if γ_0 is a geodesic.*

Proof. Suppose γ_0 is a geodesic joining p to q, then it is smooth on $[0,1]$ with zero acceleration. Hence for any variation α of γ_0, we have $\Delta_t V = 0$ and $A(t) = 0$, thus the first variation formula implies that γ_0 is a critical point. Conversely, suppose γ_0 is a critical point. Let α be the variation of γ_0 with variational vector field $W(t) := f(t)A(t)$, where $f(t)$ is a positive function except that it vanishes that the $t_i's$. Then

$$0 = \frac{1}{2}\frac{d}{du}|_{u=0}E(\alpha(u,t)) = -\int_0^1 f(t)\langle W(t), W(t)\rangle dt.$$

This is zero if and only if $A(t) = 0$ for all t. Hence each $\gamma_0|_{[t_i,t_{i+1}]}$ is a geodesic. Now let α be a variation such that $W(t_i) = \Delta_{t_i}V$. Then

$$0 = \frac{1}{2}\frac{d}{du}|_{u=0} = -\sum_t \langle \Delta_{t_i}V, \Delta_{t_i}V\rangle.$$

Hence all $\Delta_{t_i} V(t) = 0$, and γ_0 is thus C^1, even at the points t_i. Now it follows from the uniqueness theorem for differential equations that γ_0 is in fact C^∞ everywhere. $\square$

Thus by computing the critical points of the energy functional, we are able to find all the geodesics joining p and q. But this tells us nothing about the minimizing property of those geodesics. From calculus we know that if we are to find the minimum of some function, computing the first derivative can only tell us the critical points, and in order to determine minimality, we need to take the second derivative. This is the second step.

References

[1] Carmo, Do; *Riemannian Geometry*. Mathematics: Theory and Applications, Birkhäuser, 1992.

[2] Milnor, John; *Morse Theory*. Annals of Mathematics Studies No. 51, Princeton Unversity Press, 1973.