Calculus of Variations and Elliptic Equations
–
1st class

Geodesics

A very natural problem in optimization consists in finding the shortest path to connect two points. A simple but interesting issue is the following: we are given a set $M \subset \mathbb{R}^d$, two points $x_0, x_1 \in M$, and we look for the shortest path connecting x_0 to x_1 and staying in M. For the sake of generality, we can replace the set M, supposed to be a subset of the Euclidean space and inheriting its metric structure (i.e., the Euclidean distance) with a more general metric space (X, d).

First of all, let us define the length of a curve ω in a general metric space (X, d).

Definition - For a curve $\omega : [0,1] \to X$, let us define

$$\text{Length}(\omega) := \sup \left\{ \sum_{k=0}^{n-1} d(\omega(t_k), \omega(t_{k+1})) \ : \ n \geq 1,\, 0 = t_0 < t_1 < \cdots < t_n = 1 \right\}.$$

Note that the same definition could be given for functions defined on $[0,1]$ and valued in X, not necessarily continuous.

It is well-known and not difficult to prove, when $X = M \subset \mathbb{R}^d$ and $\omega \in C^1$, that we have the equality $\text{Length}(\omega) = \int_0^1 |\omega'|(t)\mathrm{d}t$. If we want to stay at a more general level, we need some notions to generalize this equality to the metric setting. You could skip this part if you are only interested in the Euclidean case (but, anyway, one would need to generalize to curves which are not C^1).

Let us start from some properties about Lipschitz curves in metric spaces.

A curve ω is a continuous function defined on a interval, say $[0,1]$ and valued in a metric space (X, d). As it is a map between metric spaces, it is meaningful to say whether it is Lipschitz or not, but its speed $\omega'(t)$ has no meaning, unless X is a vector space.

Surprisingly, it is possible to give a meaning to the modulus of the velocity, $|\omega'|(t)$.

Definition - If $\omega : [0,1] \to X$ is a curve valued in the metric space (X, d) we define the metric derivative of ω at time t, denoted by $|\omega'|(t)$ through

$$|\omega'|(t) := \lim_{h \to 0} \frac{d(\omega(t+h), \omega(t))}{|h|},$$

provided this limit exists.

The following theorem, in the spirit of Rademacher Theorem guarantees the existence of the metric derivative for Lipschitz curves.

Theorem - Suppose that $\omega : [0,1] \to X$ is Lipschitz continuous. Then the metric derivative $|\omega'|(t)$ exists for a.e. $t \in [0,1]$. Moreover we have, for $t < s$,

$$d(\omega(t), \omega(s)) \leq \int_t^s |\omega'|(\tau)\,\mathrm{d}\tau.$$

The above theorem can be proved by using the fact that every compact metric space (and $\omega([0,1])$ is the image of a compact set through a continuous map, hence it is compact) can be isometrically embedded in ℓ^∞, where one can work componentwise. For all the notions and the proofs about metric derivatives, we refer for instance to [**?**].

We also need to consider more general curves, not only Lipschitz continuous.

Definition - A curve $\omega : [0,1] \to X$ is defined *absolutely continuous* whenever there exists $g \in L^1([0,1])$ such that $d(\omega(t_0), \omega(t_1)) \leq \int_{t_0}^{t_1} g(s)\mathrm{d}s$ for every $t_0 < t_1$. The set of absolutely continuous curves defined on $[0,1]$ and valued in X is denoted by $\text{AC}(X)$.

It is well-known that every absolutely continuous curve can be reparametrized in time (through a monotone-increasing reparametrization) and become Lipschitz continuous. A possible way to achieve this goal is the following: let $G(t) := \int_0^t g(s)\mathrm{d}s$, then set $S(t) = \varepsilon t + G(t)$ (for any $\varepsilon > 0$), which is continuous and strictly increasing, and valued in an interval $[0, L]$; for $t \in [0, L]$, set $\widetilde{\omega}(t) = \omega(S^{-1}(t))$. It is easy to check that $\widetilde{\omega} \in \mathrm{Lip}_1$. If we add the assumption that ω has no interval $(t_0, t_1) \subset [0, 1]$ on which it is constant, then there is no need to use $\varepsilon > 0$ and G is already strictly increasing. in this case we use $\varepsilon = 0$ and obtain a unit-speed reparametrization of ω. If we want to have parametrizations defined on the same interval $[0, 1]$, we just need to rescale by a factor L.

In particular the above Rademacher Theorem is also true for $\omega \in \mathrm{AC}(X)$ (since the reparametrization that we defined is differentiable a.e.).

With these definitions in mind, we can state the following general equality.

Proposition - For any curve $\omega \in \mathrm{AC}(X)$ we have

$$\mathrm{Length}(\omega) = \int_0^1 |\omega'|(t)\mathrm{d}t.$$

We collect now some more definitions.

Definition - A curve $\omega : [0, 1] \to X$ is said to be a geodesic between x_0 and $x_1 \in X$ if it minimizes the length among all curves such that $\omega(0) = x_0$ and $\omega(1) = x_1$.

Definition - A curve $\omega \in \mathrm{AC}(X)$ defined on the interval $[0, 1]$ is said to be a *constant speed geodesic* between $\omega(0)$ and $\omega(1) \in X$ if it is a gesdesic and $|\omega'|(t)$ equals $\mathrm{Length}(\omega)$ for a.e. t (this is equivalent to saying that $|\omega'|$ is constant, since the only possible constant is the length of the curve).

The following characterization is useful.

Proposition 1. *Fix an exponent $p > 1$ and consider curves connecting x_0 to x_1. The following facts are equivalent*

1. *ω is a constant speed geodesic,*
2. *ω solves* $\min\left\{\int_0^1 |\omega'|(t)^p\mathrm{d}t \,:\, \omega(0) = x_0, \omega(1) = x_1\right\}$.

Proof. Set $L_0 = \inf\{\mathrm{Length}(\omega) \,:\, \omega \in \mathrm{AC}(X), \omega(0) = x_0\omega(1) = x_1\}$. We then have, for an arbitrary curve ω joining x_0 and x_1:

$$\int_0^1 |\omega'|(t)^p\mathrm{d}t \geq \left(\int_0^1 |\omega'|(t)\mathrm{d}t\right)^p \geq L_0^p,$$

where the first inequality comes from Jensen inequality and the second from the definition of L_0. The first inequality is an equality if and only if $|\omega'|$ is constant. This shows that the infimum in the second condition ($\inf\left\{\int_0^1 |\omega'|(t)^p\mathrm{d}t \,:\, \omega(0) = x_0, \omega(1) = x_1\right\}$) can be reduced to constant-speed curves, and in this case we only need to look at the power p of the length. In particular, the infimum equals L_0^p and it is attained if and only if a geodesic exists. Let us suppose that a geodesic exists. Then the value of the inf is attained and the solutions are all the curves realizing both inequalities, which means geodesics with constant speed, which shows the equivalence between the two conditions. If no geodesic exists, then no curve satisfies either condition. □

Geodesics on the sphere

We consider the case $M : \mathbb{S}^2 \subset \mathbb{R}^3$, i.e. the unit sphere in the three-dimensional space. Let us take two vectors $v_0, v_1 \in M$ such that $v_0 \cdot v_1 = 0$, and the curve

$$\gamma(t) = v_0 \cos(t) + v_1 \sin(t).$$

This curve is a great circle in M. It is C^∞ and parametrized at constant unit speed, since $\gamma'(t) = -v_0 \sin(t) + v_1 \cos(t)$ and $|\gamma'(t)|^2 = |v_0|^2 \sin^2(t) + |v_1|^2 \cos^2(t) - 2v_0 \cdot v_1 \sin(t)\cos(t) = \sin^2(t) + \cos^2(t) = 1$. Moreover, differentiating twice, se see that we have $\gamma''(t) = -\gamma(t)$.

We want to prove that, for small T, the curve γ is a geodesic in M, i.e. it minimizes the length among those curves valued in M with same initial and final points. We will take advantage of Proposition **??** and prove that γ is a constant-speed geodesic (even if we already knew that its speed is constant).

Proposition 2. *For $T \le \pi$, the arcs of great circles defined above are constant-speed geodesics.*

Proof. We will compare γ of the form $\gamma(t) = v_0 \cos(t) + v_1 \sin(t)$ to a competitor that we write as $\tilde\gamma = \gamma + \eta$. The vector-valued function $\eta : [0, T] \to \mathbb{R}^3$ will be such that $\tilde\gamma \in \mathbb{S}^2$ and $\eta(0) = \eta(T) = 0$. We will prove $\int_0^T |\gamma'|^2 \le \int_0^T |\gamma' + \eta'|^2$. The condition $|\tilde\gamma(t) \in \mathbb{S}^2$ means $|\gamma + \eta|^2 = 1$, i.e. $|\eta|^2 + 2\gamma \cdot \eta = 0$.
We have

$$\int_0^T |\gamma' + \eta'|^2 = \int_0^T |\gamma''|^2 + \int_0^T |\eta'|^2 + \int_0^T 2\gamma' \cdot \eta'.$$

We can integrate by parts the last term and, using $\eta(0) = \eta(T) = 0$, we get $\int_0^T 2\gamma' \cdot \eta' = -\int_0^T 2\gamma'' \cdot \eta$. We then use $\gamma'' = -\gamma$ and obtain $\int_0^T 2\gamma \cdot \eta = -\int_0^T |\eta|^2$, after also using $|\eta|^2 + 2\gamma \cdot \eta = 0$.
Hence,

$$\int_0^T |\tilde\gamma'|^2 = \int_0^T |\gamma''|^2 + \int_0^T |\eta'|^2 - \int_0^T |\eta|^2.$$

and the claim is proven thanks to the next lemma, which provides $\int_0^T |\eta'|^2 \ge \int_0^T |\eta|^2$ for small T. □

The following lemma is a statement of the well-known Poincaré inequality in 1D.

Lemma 1. *There exists a constant C such that $\int_0^T |\eta|^2 \le CT^2 \int_0^T |\eta|^2$ for every H^1 function $\eta : [0, T] \to \mathbb{R}^n$ with $\eta(0) = \eta(T) = 0$. The optimal value of this constant is π^{-2}.*

Proof. It is easy to prove the inequality with $C = 1$. Indeed, from $\eta(0) = 0$ we have $\eta(t) = \int_0^t \eta'(s)ds$, hence

$$|\eta(t)| \le \int_0^t |\eta'(s)|ds \le \int_0^T |\eta'(s)|ds \le \sqrt{T}\sqrt{\int_0^T |\eta'(s)|^2 ds}.$$

This implies

$$\int_0^T |\eta(t)|^2 \le T \cdot T \cdot \int_0^T |\eta'(s)|^2 ds.$$

Using $\eta(t) = v_0 \sin(\pi t/T)$ (for an arbitrary vector $v_0 \in \mathbb{R}^n \setminus \{0\}$) we see that the constant cannot be better than π^{-2}. Proving the sharp inequality can be done via Fourier series. If $T = \pi$ we can write $\eta = \sum_{n\in\mathbb{Z}} a_n \sin(nt)$, $\eta' = \sum_{n\in\mathbb{Z}} n a_n \cos(nt)$. We then have $||\eta||^2_{L^2([0,\pi])} = \sum |a_n|^2 \le \sum n^2 |a_n|^2 = ||\eta'||^2_{L^2([0,\pi])}$. The inequality is hence proven in its sharp form for $T = \pi$. A simple scaling argument, applying the inequality on $[0, \pi]$ to the function $\tilde\eta$ given by $\tilde\eta(t) = \eta(tT/\pi)$ provides the general statement. □

Techniques for existence

The most common way to prove that a function admits a minimizer is called "direct method in calculus of variations". It simply consists in the classic Weierstrass Theorem, possibly replacing continuity with semi-continuity.

Definition - On a metric space X, a function $f : X \to \mathbb{R} \cup \{+\infty\}$ is said to be lower-semi-continuous (l.s.c. in short) if for every sequence $x_n \to x$ we have $f(x) \le \liminf_n f(x_n)$.

Definition - A metric space X is said to be compact if from any sequence x_n we can extract a converging subsequence $x_{n_k} \to x \in X$.

Theorem 1 (Weierstrass)**.** *If $f : X \to \mathbb{R} \cup \{+\infty\}$ is lower semi-continuous and X is compact, then there exists $\bar{x} \in X$ such that $f(\bar{x}) = \min\{f(x) : x \in X\}$.*

Proof. Define $\ell := \inf\{f(x) : x \in X\} \in \mathbb{R} \cup \{-\infty\}$ ($\ell = +\infty$ only if f is identically $+\infty$, but in this case any point in X minimizes f, or if X is empty, another no-interesting case). By definition there exists a minimizing sequence x_n, i.e. points in X such that $f(x_n) \to \ell$. By compactness we can assume $x_n \to \bar{x}$. By lower semi-continuity, we have $f(\bar{x}) \le \liminf_n f(x_n) = \ell$. On the other hand, we have $f(\bar{x}) \ge \ell$ since ℓ is the infimum. This proves $\ell = f(\bar{x}) \in \mathbb{R}$ and this value is the minimum of f, realized at $\bar{x}$. ☐

Note that there is no need to check in advance that f is bounded from below and $\ell \neq -\infty$: this will be proven as a consequence of $\ell = f(\bar{x})$.

We will apply the same ideas so as to prove existence of a minimizer in a toy-example.

Consider an interval $I = [a, b] \subset \mathbb{R}$ and a countinuous function $F : I \times \mathbb{R} \to \mathbb{R}$. We suppose $F \ge 0$ (bounded from below by a different constant could be treated in a similar way).

Theorem 2. *Consider the problem*

$$\min\left\{J(u) := \int_a^b \left(F(t, u(t)) + |u'(t)|^2\right) dt \quad : \quad u \in H^1(I),\ u(a) = A,\ u(b) = B\right\}.$$

This minimization problem admits a solution.

Proof. Take a minimizing sequence u_n, i.e. $J(u_n) \to \inf J$. The sum of the two terms of J (the part with F and the one with $|u'|^2$) is then bounded, but the two terms are positive, hence they are both separatey bounded from above. In particular, $||u'_n||_{L^2}$ is bounded. Moreover, since the boundary values of u_n are fixed, an easy application of the Poincaré inequality shows that $||u_n||_{H^1}$ is bounded.

Hence, $(u_n)_n$ is bounded in H^1 and we can extract a weakly converging subsequence, since H^1 is a separable and reflexif space. We have $u_n \rightharpoonup u$ in H^1. This convergence also implies uniform convergence, because of the compact injection of H^1 into C^0. In particular, we have pointwise convergence and the boundary values are preserved, so that we obtain $u(a) = A$ and $u(b) = B$. the limit u is then an admissible competitor in our minimization problem. We just need to prove $J(u) \le \liminf_n J(u_n) = \inf J$ in order to deduce that u minimizes J.

since $u_n \to u$ uniformly, the functions v_n defined by $v_n(t) = F(t, u_n(t))$ uniformly converge to v defined via $v(t) = F(t, u(t))$ (using the continuity of F). In particular, $\int_a^b v_n \to \int_a^b v$. This provides the continuity (w.r.t. the weak convergence in H^1) of the first term.

We then note that the map $H^1 \ni v \mapsto v' \in L^2$ is continous, so that the weak convergence of u_n to u in H^1 implies that of u'_n to u' in L^2. We then use the well-known property of the lower semi-continuity of the norm for the weak convergence: we have $||u'||_{L^2} \le \liminf ||u'_n||_{L^2}$. Hence

$$\int_a^b |u(t)|^2 dt \le \liminf_n \int_a^b |u_n(t)|^2 dt.$$

This shows the lower semicontinuity of the second term and proves the claim. ☐

Remarks and variants

The same arguments could be applied with minor modifications if the constraints on $u(a)$ and/or $u(b)$ were replaced with penalizations, such as in

$$\min\left\{J(u) := \int_a^b \left(F(t, u(t)) + |u'(t)|^2\right) dt + g(u(a)) + h(u(b)) \quad : \quad u \in H^1(I)\right\}.$$

If g and h are continuous, the uniform convergence of u_n to u implies the continuity of the two additional terms. Yet, we need to guarantee that the sequence stays bounded in H^1. For this any of the following assumptions is enough: either $\lim_{x\to\pm\infty} g(x) = +\infty$ and h is bounded from below, or $\lim_{x\to\pm\infty} h(x) = +\infty$ and g is bounded from below, or g and h are bounded from below and $F(t,x) \geq c(|x|^2 - 1)$ for a positive constant. Indeed, if the conditions on the growth of g or h are satisfied we can obtain a bound on the value of $u_n(a)$ or $u_n(b)$; if the condition on the growth of F is satisfied, then J provides a bound on the L^2 norm of u_n and we can obtain the H^1 norm without using Poincaré.

Another easy variant concerns replacing the continuity assumption on F, g or h with lower semicontinuity (in the case of F, the semicontinuity of the integral term is obtained by using Fatou's lemma).

Finally, the term $|u'(t)|^2$ could be easily replaced with $|u'(t)|^p$, $p > 1$, and the result could be proven using the space $W^{1,p}(I)$. For $p = 1$, on the other hand, there is a difficulty because we would have a non-reflexive space.

Minimization among smooth functions

What about solving the same minimization problem in a more classical functional space, such as C^1?a possible strategy is the following: we extend the problem to a larger space such as H^1 or another Sobolev space, choosing a space in which the functional J is well-defined and we are able to prove existence; then, we write the Euler-Lagrange of the problem and we use it to prove that the minimizer is more regular than just Sobolev, and in particular C^1. this can depend on the regularity of the data of the problem, and in particular on F. If we suppose $F \in C^1$, then the Euler-Lagrange equation is of the form $u''(t) = F'(t, u(t))$ (where F' is the derivative w.r.t. the second variable): using $F' \in C^0$ and $u \in C^0$ (since sobolev functions in 1D are continuous) we deduce $u'' \in C^0$, hence $u \in C^2$. In case $F \in C^\infty$ then we also obtain, iterating the same argument, $u \in C^\infty$.

In the example of Theorem **??** every minimizing sequence satisfies he compactness property that we need to guarantee existence of a minimizer. This is not always the case, as we can see from the following result (see also [**?**], Chapter 4).

Theorem 3. *Suppose that (X, d) is a metric space in which closed balls are compact. Then for every two points $x_0, x_1 \in X$ which can be connected by a finite-length curve, there exists at least one geodesic connecting them.*

Proof. Following the usual strategy we consider a minimizing sequence $\gamma_n : [0,1] \to X$ of AC curves with $\gamma_n(i) = x_i$ for $i = 0, 1$, and $\text{Length}(\gamma_n) \to \ell := \inf\{\text{Length}(\omega) : \omega \in AC([0,1]; X), \omega(0) = x_0, \omega(1) = x_1\}$. Since $\text{Length}(\gamma_n) = \int_0^1 |\gamma_n'|(t)\mathrm{d}t$ we have a bound on the L^1 norm of the metric speed $|\gamma_n'|$. Unfortunately, L^1 is not a reflexive space, nor the dual of a separable space, and we cannot apply the Banach-Alaoglou theorem to extract a converging subsequence.

We will be able to extract a suitable subsequence only after improving the sequence γ_n. First, we can remove from each γ_n any possible interval on which γ_n is constant (and reparametrize the curve we obtain) since this does not change the lenth, nor the endpoints. Then, we can parametrize the obtained curve by constant speed. We thus obtain a sequence $\tilde{\gamma}_n$ with the property $|\tilde{\gamma}_n'| = \text{Length}(\tilde{\gamma}_n) = \text{Length}(\gamma_n) \to \ell$. For these curves, the L^∞ and the L^1 norms of the speed coincide. In particular, the sequence $\tilde{\gamma}_n$ is equi-Lipschitz since $\text{Lip}(\tilde{\gamma}_n) = |||\tilde{\gamma}_n'|||_{L^\infty}$.

Using $d(\tilde{\gamma}_n(t), x_0) \leq \text{Length}(\tilde{\gamma}_n) \leq C$ we easily see that all the curves $\tilde{\gamma}_n$ are contained in a given ball, hence in a compact set. We can then apply Ascoli-Arzelà's theorem, extracting a uniformly converging subsequence, that we will still call $\tilde{\gamma}_n$, and we denote by γ its uniform limit. As the limit is also pointwise, the endpoints are preserved. We write

$$d(\tilde{\gamma}_n(t), \tilde{\gamma}_n(s)) \leq \text{Lip}(\tilde{\gamma}_n)|t - s|$$

and, passing to the limit $n \to \infty$, we obtain

$$d(\gamma(t), \gamma(s)) \leq \ell|t - s|,$$

which shows that γ is ℓ-Lipschitz continuous. This implies $\text{Length}(\gamma) \leq \ell$ and shows that γ is a geodesic. $\square$

Note that in the above proof we could have used the L^1 and L^∞ bounds on $|\tilde{\gamma}_n'|$ to deduce an L^2 bound, and use the weak convergence in H^1, at least in the Euclidean case $X \subset \mathbb{R}^d$ (otherwise the H^1 space is not defined). An extra difficulty would be in this case to prove the semicontinuity of the L^1 norm w.r.t. the weak L^2 convergence, as the norm is not the same (but this can be easily be fixed, for instance noting that L^2 continuously embeds into L^1).

References

[1] L. Ambrosio and P. Tilli, *Topics on analysis in metric spaces*. Oxford Lecture Series in Mathematics and its Applications (25). Oxford University Press, Oxford, 2004.