Introduction to the Calculus of Variations
Lecture Notes
Lecture 9

Swarnendu Sil

Spring 2021, IISc

Chapter 1

Prelude to Direct Methods

Direct methods in a classical problem Now we intend to give a brief illustration of direct methods in a classical problem. We intend to study a problem that we have referred to a few times already, the problem of finding **geodesics**, i.e. curves of 'shortest' length between two given points on a manifold.

However, we are going to solve the problem using the **direct methods**. Although the setting is decidedly simpler here, but we would already see a remarkable number of features that would remain with us in different guises and would keep us busy till the end of the course. Roughly, these are the following.

- Sobolev spaces (we would see a baby version here and this would stay with us from chapter 4 onwards)
- direct methods for existence (this will return and stay with us from chapter 5 onwards)
- noncompactness due to group action and a possible way to overcome it (this would return when we study the area functional in the last chapter)
- regularity questions (we shall take it up again the chapter 6)

1.1 Geodesics: the problem

1.1.1 The variational problem for geodesics

Let M be an N-dimensional smooth embedded submanifold of $\mathbb{R}^d$. Let $c \in C^1([0,T]; M)$ be a C^1 curve on M. Let $p_1, p_2 \in M$ be two distinct points on M. We suppose that the curve begins at p_1 and ends at p_2, which translates to

$$c(0) = p_1 \qquad \text{and} \qquad c(T) = p_2.$$

The length of the curve is

$$L(c) := \int_0^T |\dot{c}(t)| \, \mathrm{d}t.$$

Our aim is to find a curve connecting p_1 and p_2 which has the shortest length.

So our first try for the variational problem is

$$\boxed{\inf \left\{ L(c) : c \in C^1 \left([0,T]; \mathbb{R}^d \right), c(0) = p_1, c(T) = p_2. \right\} = m.}$$

But clearly this can not be the variational problem. It has no reference to M whatsoever! In fact, we already know the solution to the above variational problem (though it is quite tricky proving it this way!). The straight line in $\mathbb{R}^d$ joining the points p_1 and p_2 is the unique path of shortest length. This path has no reason to lie in M. (Think of M as the N-sphere $\mathbb{S}^N$ in $\mathbb{R}^{N+1}$.)

Now there are two ways we can bring M into the picture. One is if M is given by some equations

$$M = \left\{ x \in \mathbb{R}^d : G_\alpha(x) = 0 \text{ for all } \alpha \in \mathcal{I} \right\},$$

then we can treat this as a variational problem with additional constraints

$$G_\alpha(c(t)) = 0 \qquad \text{for all } \alpha \in \mathcal{I}.$$

However, here we shall not take this path and instead introduce local charts in M.

Charts, length and the metric tensor

Local charts Let $p \in M$. A local chart around p is a map $f : U \subset \mathbb{R}^N \to V \subset \mathbb{R}^d$ such that

- U, V are open sets in the respective Euclidean spaces,
- $f(U) = M \cap V$,
- $p \in f(U)$ and
- f is a smooth diffeomorphism onto its image.

Now since f is a diffeomorphism, for any curve $c(t)$ which is contained inside a single chart, i.e. $c([0,T]) \subset f(U)$, there exists a curve γ in U such that

$$c(t) = f(\gamma(t)) \qquad \text{for every } t \in [0,T].$$

γ is also C^1 if c is and by the chain rule, we have

$$\dot{c}(t) = Df(\gamma(t))\, \dot{\gamma}(t).$$

Writing in components, this is

$$\dot{c^{\alpha}}(t) = \frac{\partial f^{\alpha}}{\partial z^{i}}(\gamma(t))\,\dot{\gamma^{i}}(t) \qquad \text{for every } 1 \leq \alpha \leq d.$$

Here we used the **Einstein summation convention**, where any repeated index would be summed over its respective range (i.e. here the only repeated index is i, which is to be summed over, here from 1 to N). Thus

$$\boxed{\begin{aligned} L(c) &= \int_0^T \left(\frac{\partial f^{\alpha}}{\partial z^{i}}(\gamma(t))\,\dot{\gamma^{i}}(t)\,\frac{\partial f^{\alpha}}{\partial z^{j}}(\gamma(t))\,\dot{\gamma^{j}}(t) \right)^{\frac{1}{2}}\,\mathrm{d}t \\ &= \int_0^T \left(g_{ij}(\gamma(t))\,\dot{\gamma^{i}}(t)\,\dot{\gamma^{j}}(t) \right)^{\frac{1}{2}}\,\mathrm{d}t, \end{aligned}}$$

where $G = (g_{ij})$ is a **positive definite symmetric** $N \times N$ matrix

$$\boxed{g_{ij}(z) = \frac{\partial f^{\alpha}}{\partial z^{i}}(z)\,\frac{\partial f^{\alpha}}{\partial z^{j}}(z),}$$

called the **metric tensor** of M with respect to the chart $f : U \to V$, where as usual we used the Einstein summation convention, i.e. the repeated index α is to be summed over from 1 to d. The symmetry of the metric tensor G can be proved immediately from the expressions for g_{ij} above. For proving positive definiteness, note that the expressions imply

$$G = (Df)^{T}\,Df, \tag{1.1}$$

where the superscript denotes the transpose and Df is invertible. The positive definiteness of G is now a simple exercise in linear algebra. Note that the symmetry of G can be immediately read off from this expression as well.

Chart overlaps, transition functions and the length of curves We shall always work with the simplifying assumption that the curve is contained in a single chart just for clarity. In general, a manifold would be covered by a collection of charts $\{(f_{\beta}, U_{\beta})\}_{\beta}$ (called an atlas). Given a curve c on M, we can always find a partition

$$0 = t_0 < t_1 < \ldots < t_r < T$$

such that $c([t_k, t_{k+1}])$ is contained in a single chart and then we would write the length functional as sum of the integrals. It might appear that the length of a curve depends on the chart chosen. But it does not. If $c(0,T) \subset f_1(U_1) \cap f_2(U_2)$, then one can check we have

$$\int_0^T \left(g^{1}_{ij}(\gamma_1(t))\,\dot{\gamma}^{i}_{1}(t)\,\dot{\gamma}^{j}_{1}(t) \right)^{\frac{1}{2}}\,\mathrm{d}t = \int_0^T \left(g^{2}_{ij}(\gamma_2(t))\,\dot{\gamma}^{i}_{2}(t)\,\dot{\gamma}^{j}_{2}(t) \right)^{\frac{1}{2}}\,\mathrm{d}t,$$

where

$$f_1 \circ \gamma_1 = c = f_2 \circ \gamma_2.$$

The diffeomorphism

$$f_2^{-1} \circ f_1 : f_1^{-1}\left(f_1\left(U_1\right) \cap f_2\left(U_2\right)\right) \to f_2^{-1}\left(f_1\left(U_1\right) \cap f_2\left(U_2\right)\right)$$

is called a **transition map**.

The variational problem Now our variational problem is

$$\boxed{\inf_{\gamma \in X} \left\{ \int_0^T \left(g_{ij}\left(\gamma\left(t\right)\right) \dot{\gamma}^i\left(t\right) \dot{\gamma}^j\left(t\right) \right)^{\frac{1}{2}} \ \mathrm{d}t \right\} = m.}$$

where

$$X = \left\{ : \gamma \in C^1\left([0,T]; U\right) : \gamma\left(0\right) = f^{-1}\left(p_1\right), \gamma\left(T\right) = f^{-1}\left(p_2\right) \right\}.$$

Now we attempt to solve it via direct methods. But it is a quite difficult one and we need to slowly move towards it.

1.1.2 Difficulties

To understand at least some of the difficulties of the problem, we first try to solve the simpler problem

$$\boxed{\inf \left\{ L\left(c\right) : c \in C^1\left([0,T]; \mathbb{R}^d\right), c(0) = p_1, c(T) = p_2. \right\} = m.}$$

This is just the first problem we wrote down today.

Now, as we did in the case of finding a minima, if $\{c_\nu\}$ is a minimizing sequence, i.e.

$$L\left(c_\nu\right) = \int_0^T \left|\dot{c_\nu}\left(t\right)\right| \ \mathrm{d}t \to m,$$

we deduce

$$\left\|\dot{c_\nu}\right\|_{L^1([0,T])} := \int_0^T \left|\dot{c_\nu}\left(t\right)\right| \ \mathrm{d}t \leq m + 1.$$

Now we see one of the first difficulties. We obtained an uniform bound for the L^1 norm of the derivatives and **not** the C^0 norm of the derivatives.

So we realize that C^1 is a terrible class from the point of view of direct methods. From integral functionals, uniform bounds for some integral norms of the derivatives are the best we can hope for. So, minimizing sequences would never be uniformly bounded in the C^1 norm!

However, we push ahead a bit more. Using the fundamental theorem of calculus, we obtain

$$\boxed{|c_\nu(t)| \le |c_\nu(0)| + \left| \int_0^t \dot{c_\nu}(t)\ \mathrm{d}t \right| \le |p_1| + \|\dot{c_\nu}\|_{L^1([0,T])} \le |p_1| + m + 1,}$$

and

$$\boxed{|c_\nu(t) - c_\nu(s)| = \left| \int_s^t \dot{c_\nu}(t)\ \mathrm{d}t \right| \le \int_s^t |\dot{c_\nu}(t)|\ \mathrm{d}t.}$$

So at least the C^0 norm of the minimizing sequences are uniformly bounded. However, this is not good enough for extracting a convergent sequence. (Thus showing C^0 is an equally bad space as C^1).

But we were very close. By virtue of the Ascoli-Arzela theorem, all we needed for compactness is **equicontinuity**, i.e.

$$|c_\nu(t) - c_\nu(s)| \to 0 \quad \text{uniformly in } \nu \quad \text{as } t - s \to 0.$$

From the second inequality, this would be the case if we can conclude

$$\int_s^t |\dot{c_\nu}(t)|\ \mathrm{d}t \to 0 \quad \text{uniformly in } \nu \quad \text{as } t - s \to 0.$$

This property is called **equiintegrability**. Unfortunately, a sequence which is uniformly bounded in L^1 need not be equiintegrable, showing L^1 is not a particularly nice space either.

An easier problem Let us now make our life a bit easier and try to solve the variational problem

$$\boxed{\inf_{c \in X} \left\{ E(c) = \int_0^T |\dot{c}(t)|^2\ \mathrm{d}t \right\} = m,}$$

where

$$X = \left\{ c \in C^1\left([0,T]; \mathbb{R}^d\right) : c(0) = p_1, c(T) = p_2 \right\}.$$

Arguing as before, for a minimizing sequence $\{c_\nu\}$, we now have

$$E(c_\nu) = \|\dot{c_\nu}\|^2_{L^2([0,T])} \le m + 1.$$

But this time we have a control of the L^2 norm of the derivatives instead of the L^1 norm.

Compactness in C^0 Using the fundamental theorem of calculus once again, this time we obtain

$$\begin{aligned}|c_\nu(t)| &\leq |c_\nu(0)| + \left|\int_0^t \dot{c_\nu}(t)\ \mathrm{d}t\right| \\ &\overset{\text{Hölder}}{\leq} |p_1| + \sqrt{t}\,\|\dot{c_\nu}\|_{L^2([0,T])} \\ &\leq |p_1| + \sqrt{t}\sqrt{m+1}.\end{aligned}$$

Thus $\{c_\nu\}$ is uniformly bounded in C^0. Now we have,

$$\begin{aligned}|c_\nu(t) - c_\nu(s)| = \left|\int_s^t \dot{c_\nu}(t)\ \mathrm{d}t\right| &\overset{\text{Hölder}}{\leq} \sqrt{(t-s)}\left(\int_s^t |\dot{c_\nu}(t)|^2\ \mathrm{d}t\right)^{\frac{1}{2}} \\ &\leq \sqrt{(t-s)}\,\|\dot{c_\nu}\|_{L^2([0,T])} \\ &\leq \sqrt{(t-s)}\sqrt{m+1}.\end{aligned}$$

Thus,

$$|c_\nu(t) - c_\nu(s)| \to 0 \quad \text{uniformly in } \nu \quad \text{as } t-s \to 0.$$

Hence by Ascoli-Arzela theorem, we deduce that up to the extraction of a subsequence which is not relabelled, we obtain

$$c_\nu \to c \qquad \text{in } C^0,$$

for some $c \in C^0\left([0,T];\mathbb{R}^d\right)$. Unfortunately, this tells us nothing about the derivatives of c.

c might not even be differentiable, let alone being C^1.

However, since $\{\dot{c_\nu}\}$ is uniformly bounded in L^2, which unlike L^1, is a reflexive space, we deduce, by Banach-Alaoglu theorem

$$\dot{c_\nu} \rightharpoonup v \quad \text{in } L^2, \tag{1.2}$$

for some $v \in L^2\left([0,T];\mathbb{R}^d\right)$. Is there a relation between v and c? In particular, is $v = \dot{c}$?

Note that (1.2) implies for any $\psi \in C_c^\infty\left([0,T];\mathbb{R}^d\right)$, we have

$$\int_0^T \langle \dot{c_\nu}, \psi\rangle \to \int_0^T \langle v, \psi\rangle. \tag{1.3}$$

1.1.3 Idea of weak derivatives

But integrating by parts, we obtain

$$\int_0^T \langle \dot{c_\nu}, \psi\rangle = -\int_0^T \left\langle c_\nu, \dot{\psi}\right\rangle. \tag{1.4}$$

By convergence of c_ν to c in C^0, the RHS above converges to

$$-\int_0^T \left\langle c_\nu, \dot{\psi} \right\rangle \to -\int_0^T \left\langle c, \dot{\psi} \right\rangle.$$

So, using this and (1.2) and (1.3) and (1.4), we deduce

$$\boxed{\int_0^T \langle v, \psi \rangle = -\int_0^T \left\langle c, \dot{\psi} \right\rangle \qquad \text{for any } \psi \in C_c^\infty \left([0,T]; \mathbb{R}^d\right).}$$

v certainly looks way too much like $\dot{c}$!! Indeed, if we knew c is C^1, the above formula would indeed tell us $v = \dot{c}$ using integration by parts and the fundamental lemma of calculus of variations.

Unfortunately, we have no way of knowing at this point that c is C^1. As we said, for all we know, c need not even be differentiable. However, since $v \in L^2$, the above formula suggests that probably instead of C^1 curves, we should look for 'curves' with L^2 'derivatives'.

But how can a function which might not be differentiable have a 'derivative'??

The last bit of inspired idea that we need is that we need to outrageously bold and simply call v as a **'derivative'** of c!! This seemingly insane idea is the beginning of modern theory of PDEs and Calculus of Variations.

Definition 1 (weak derivatives)**.** *Let $u \in L^1\left([0,T]; \mathbb{R}^d\right)$. We say u has a **weak derivative** if there exists a function $v \in L^1\left([0,T]; \mathbb{R}^d\right)$ such that*

$$\boxed{\int_0^T \langle v, \psi \rangle = -\int_0^T \left\langle u, \dot{\psi} \right\rangle \qquad \textit{for any } \psi \in C_c^\infty \left([0,T]; \mathbb{R}^d\right).}$$