A BRIEF INTRODUCTION TO CALCULUS OF VARIATION

LONG CHEN

ABSTRACT. In this brief note, we provide an overview of the Calculus of Variations, highlighting three key tools: the chain rule of differentiation, integration by parts, and change of variables. To illustrate these aspects, we present a one-dimensional example.

1. PROBLEM FORMULATION

A typical problem in the calculus of variation is in the form

$$\inf_{u \in \mathcal{M}} I(u), \tag{1}$$

where the integral functional is

$$I(u) = \int_{\Omega} L(x, u(x), \nabla u(x))\, \mathrm{d}x,$$

- $\Omega \subset \mathbb{R}^d$ is an open domain;
- $\mathcal{M}$ is a subset of some function spaces defined on Ω called admissible set;
- L is called *Lagrangian*.

When considering 1-d problems, i.e., $d = 1$, usually we write the independent variable as t, change u to x, and $L = L(t, x(t), \dot{x}(t))$.

So it is a minimization problem or in general optimization problems (finding minimum, maximum, or saddle points) which shares many similarities with the calculus problem

$$\inf_{x \in M} f(x). \tag{2}$$

Using equation (2) can be helpful in understanding key concepts and points related to the calculus of variations problem described in equation (1).

What is the main difference between the calculus problem (2) and the calculus of variation problem (1)?

- The dependence in (2) is two-layer $x \mapsto f(x)$;
- While (1) contains three-layers: $x \mapsto u(x) \mapsto I(u)$;
  In 1-d, the dependence is: $t \mapsto x(t) \mapsto I(x)$.

2. TOOLS

In the calculus of variations, the independent variable in the functional $I(\cdot)$ is a function u, which itself is a function of x in some domain Ω of Euclidean space $\mathbb{R}^d$. A function of functions is called a functional, and functional analysis is the main mathematical tool used in the calculus of variations.

We can gain insight into the role of functional analysis in the calculus of variations by examining the existence and uniqueness of solutions to (1). This involves using various concepts and tools from functional analysis, such as weak convergence, Banach spaces, and Sobolev spaces.

Recall that the infimum of the calculus problem (2) exists if M compact and f is continuous.

The subset $\mathcal{M}$ in equation (1) is a subset of a function space, such as the Sobolev spaces $H^1(\Omega)$, which is typically infinite dimensional. In contrast, M in equation (2) is a subset of the finite dimensional space $\mathbb{R}^d$. This difference is fundamental and requires many results and properties of finite dimensional linear spaces to be re-examined in the infinite dimensional space.

One example is *compactness*. In a finite dimensional normed vector space, a set is compact if and only if it is bounded and closed. However, in infinite dimensional spaces, boundedness and closedness are necessary but not sufficient conditions for compactness. An example is the $\{e_n\} \subset \ell^2$, $e_n = (0, 0, \ldots, 1, 0, \ldots)$ and $\|e_n - e_m\| = \sqrt{2}$. No convergent subsequences. Weaker topologies need to be introduced to enable the unit ball to be compact in that weak topology. In view of open sets, a weaker topology will have less open sets and consequently less open covering which will increase the possibility to satisfy "for every open covering" in the definition of compactness: for every open covering of M, there exists a finite covering.

A function $f : U \to V$ between two topological spaces is continuous if for every open sets in V, its pre-image is continuous. Weaker topology will have less open sets, which is not favorable to the continuity. We can relax the continuity to the lower semi-continuity in the weak topology (w.l.s.c) to balance the less open sets in the weaker topology. Coercivity and convexity of the functional L will be introduced to ensure the existence of a solution to equation (1).

More specifically, we summarize three most used tricks in Calculus of Variation.

- Chain rule of differentiation. (*Easy but tedious*)
- Integration by parts. (*Medium to hard especially in high dimensions*)
- Change of variables. (*Hard and deep*)

When taking derivatives, we have to be careful on the dependence of variables. For example, when write the Lagrangian as $L(x, u, p)$, variables (u, p) are independent. While in the notation $L(x, u(x), \nabla u(x))$, $p = \nabla u$ is substituted and therefore the 2nd and 3rd variables are now related.

The functional $I(\cdot)$ involves the derivative ∇u and integral $\int_\Omega L$. The interplay of these two is: integration by parts. When the domain Ω is less smooth (e.g. a polyhedron), integration by parts should be applied piecewisely and jump conditions at lower geometric objects, such as corners and edges, may appear if the function is not smooth enough.

Change of variables turns out to be the key tool which leads to the deepest results in calculus of variations. Examples include: Legendre transform and Noether's theorem.

Legendre transform changes a Lagrangian to a Hamiltonian and the Euler-Lagrange equation into the Hamiltonian system. The further introduce of a scalar potential gives the Hamilton-Jacobi equation. Different variables and different equations reveal different structures for the same physical system.

If the functional is invariant under some transformations, then there is a conservation law. This is the most beautiful and the deepest result in calculus of variations: Noether's theorem. According to Noether's theorem, the conservation of energy arises from the symmetry of the system in time. In fact, a more accurate term for "invariance" is "symmetry" and "change of variables" is "group actions". Noether's theorem provides examples of the connection between symmetry and conservation laws. Three such examples are:

- Time translation symmetry: conservation of energy;
- Space translation symmetry: conservation of momentum;

- Rotation symmetry: conservation of angular momentum.

This connection between symmetry and conservation laws is a deep, two-way relationship: a symmetry implies a conservation law, and a conserved quantity generates the symmetry itself. This relationship has inspired much of theoretical physics from experimental physics. Scientists first observe a conserved quantity in experiments and then look for the symmetry generated by that quantity.

3. An example

We end this introduction by an example in 1-D. Recall that in 1-D, we use $t \in \mathbb{R}$ instead of x, x to replace u, and $\dot{x}$ instead of ∇u. Although $t \in \mathbb{R}$, the function $x(t)$ could be a curve in space, i.e., x could be a vector. A typical example is

$$\inf_{x\in\mathcal{M}} \int_0^1 L(t, x(t), \dot{x}(t))\,\mathrm{d}t, \tag{3}$$

with

- $\mathcal{M} = \{x \in C^1(0,1), x(0) = x_0, x(1) = x_1\}$
- $L(t,x,v) = \dfrac{1}{2}m|v|^2 - U(x)$ is the difference of kinetic energy $T(v) = \frac{1}{2}m|v|^2$ and potential energy $U(x)$.

Recall that the first order condition for (2) is $f'(x) = 0$. But now $x = x(t)$ itself is a function and may change in every point $t \in (0,1)$. How to take derivative of a function?

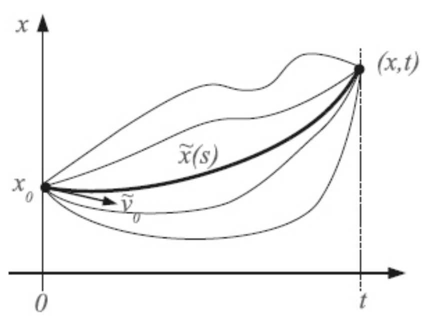

FIGURE 1. Variation of a curve.

The idea is to introduce a variation of the function. Let ϕ be a test function in $\mathcal{M}_0$ satisfies certain boundary conditions so that if $x \in \mathcal{M}$, then $x + \epsilon\phi \in \mathcal{M}$. For the example considered, $\mathcal{M}_0 = \{\phi \in C^1(0,1), \phi(0) = \phi(1) = 0\}$. The term $\epsilon\phi$ is an *variation* of x and the symbol ϵ suggests it is small. Define $f(\epsilon) := I(x + \epsilon\phi)$. Then x is an infimum of (1) if and only if 0 is a minimum of $f(\epsilon)$. So the optimality condition of an extreme curve x of $I(x)$ is characterized as

$$f'(0) = \frac{\mathrm{d}}{\mathrm{d}\epsilon} I(x + \epsilon\phi)|_{\epsilon=0} = 0, \tag{4}$$

We then apply the three trickes.

Chain rule. By the chain rule, we obtain the variational form of Euler-Lagrange equation

$$\int_0^1 L_x(t,x,\dot{x})\phi + L_v(t,x,\dot{x})\dot{\phi} = 0 \quad \forall \phi \in \mathcal{M}_0. \tag{5}$$

Integration by parts. Apply integration by parts, we obtain

$$\int_0^1 \left[L_x(t,x,\dot{x}) - \frac{\mathrm{d}}{\mathrm{d}t}L_v(t,x,\dot{x})\right]\phi = 0 \quad \forall \phi \in \mathcal{M}_0.$$

The boundary terms disappeared as $\phi \in \mathcal{M}_0$. As $\mathcal{M}_0$ is dense in $L^2(0,1)$, we conclude the strong form of Euler-Lagrange equation

$$-\frac{\mathrm{d}}{\mathrm{d}t}L_v(t,x(t),\dot{x}(t)) + L_x(t,x(t),\dot{x}(t)) = 0, \tag{6}$$

which is in general a nonlinear second order ODE.

For $L(t,x,v) = \frac{1}{2}m|v|^2 - U(x)$, (6) becomes the Newton's equation

$$m\ddot{x} = F, \quad \text{with } F = -\nabla_x U.$$

The way to derive Newton's equation from (3) is known as "the principle of least action (of Hamilton's form)".

Change of variables. Consider change of variables

$$\begin{cases} q = x, \\ p = L_v(t,q,\dot{q}). \end{cases}$$

Assume L is convex in v. Then we can solve $\dot{q} = \dot{q}(p,q,t)$. Define Hamiltonian

$$H(p,q,t) := p\dot{q} - L(t,q,\dot{q}). \tag{7}$$

The variable $\dot{q}$ on the right hand side can be expressed by variables (p,q,t) and thus $\dot{q}$ is eliminated. Then the Euler-Lagrange equation becomes the Hamiltonian system

$$\begin{cases} \dot{p} = -H_q, \\ \dot{q} = H_p. \end{cases} \tag{8}$$

To derive the Hamiltonian system, we apply total differentiation to both sides of (7). LHS is

$$\mathrm{d}H = H_p\,\mathrm{d}p + H_q\,\mathrm{d}q + H_t\,\mathrm{d}t$$

and RHS is

$$p\,\mathrm{d}\dot{q} + \dot{q}\,\mathrm{d}p - L_t\,\mathrm{d}t - L_q\,\mathrm{d}q - L_v d\dot{q} = \dot{q}\,\mathrm{d}p - \dot{p}\,\mathrm{d}q - L_t\,\mathrm{d}t.$$

The last step is due to $p = L_v$ by definition and E-L equation $L_q = \dot{p}$ in terms of new variables. Compare the coefficients of $\mathrm{d}p$ and $\mathrm{d}q$, we get (8).

When Hamiltonian $H(p,q)$ is independent of t, for solutions to the Hamiltonian system, we obtain the conservation of Hamiltonian, i.e.

$$\frac{\mathrm{d}}{\mathrm{d}t}H(p(t),q(t)) = H_p\dot{p} + H_q\dot{q} = 0.$$

An important example is $L = T - U$. For $T(v) = \frac{1}{2}m|v|^2$, the variable $p = mv$ is the momentum. Then

$$H = p\dot{q}(p) - L = pv - T(v) - U(q) = \frac{1}{2}m|v|^2 + U(q) = T + U$$

is the total energy.