A BRIEF INTRODUCTION TO CALCULUS OF VARIATION

LONG CHEN

ABSTRACT. In this note, we provide a brief 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 variations has 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 a function space defined on Ω, called the admissible set;
- L is called the *Lagrangian*.

In the case of one-dimensional problems, we usually denote the independent variable by t, write u as x, and express L as $L(t, x(t), x'(t))$ and

$$I(x) = \int_a^b L(t, x(t), x'(t))\,\mathrm{d}t.$$

Problem (1) is a minimization problem and more generally an optimization problem (seeking a minimum, maximum, or saddle point), which shares many similarities with the classical calculus problem

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

Problem (2) helps illustrate key concepts and ideas related to the calculus of variations problem in (1).

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

- The dependence in (2) is two-layer: $x \mapsto f(x)$;
- In contrast, (1) involves three layers: $x \mapsto (u(x), \nabla u(x)) \mapsto I(u)$;
  In one dimension, the dependence is: $t \mapsto (x(t), x'(t)) \mapsto I(x)$.

2. TOOLS AND TRICKS

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

We can better understand the role of functional analysis in the calculus of variations by studying the existence and uniqueness of solutions to (1). This requires concepts and tools from functional analysis, such as weak convergence, Banach spaces, and Sobolev spaces.

The subset $\mathcal{M}$ in (1) is a subset of a function space, such as the Sobolev space $H^1(\Omega)$, which is typically infinite dimensional. In contrast, M in (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 setting.

One example is *compactness*. Recall that the infimum of the calculus problem (2) exists if M is compact and f is continuous on M. 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 sequence $\{e_n\} \subset \ell^2$, where $e_n = (0, 0, \ldots, 1, 0, \ldots)$ and $\|e_n - e_m\| = \sqrt{2}$. No convergent subsequence exists. Weaker topologies need to be introduced to make the unit ball compact in the weak topology.

From the perspective of open sets, a weaker topology has fewer open sets and therefore fewer open coverings, which increases the chance of satisfying the definition of compactness: for every open covering, there exists a finite sub-covering.

A function $f : U \to V$ between two topological spaces is continuous if the pre-image of every open set in V is open in U. A weaker topology, having fewer open sets, is not favorable for continuity. To address this, continuity can be relaxed to lower semi-continuity in the weak topology (w.l.s.c), balancing the reduction in open sets. Coercivity and convexity of the functional L are introduced to ensure the existence of a solution to (1).

2.2. **Three tricks.**

We summarize three most used tricks in the Calculus of Variations:

- 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 about the dependence of variables. For example, when writing the Lagrangian as $L(x, u, p)$, the variables (u, p) are independent. In the notation $L(x, u(x), \nabla u(x))$, however, $p = \nabla u$ is substituted, and therefore the second and third variables are now related and both are functions of the first variable.

The functional $I(\cdot)$ involves the derivative ∇u and the integral $\int_\Omega L$. The interplay between these two is handled using 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-dimensional geometric objects, such as corners and edges, may arise if the function is not smooth enough.

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

The Legendre transform converts a Lagrangian into a Hamiltonian and transforms the Euler–Lagrange equation into a Hamiltonian system. Further introduction of a scalar potential leads to the Hamilton–Jacobi equation. Different variables and different equations reveal different structures of the same physical system.

If a functional is invariant under certain transformations, then there exists a corresponding conservation law. This is the most beautiful and profound result in the calculus of variations: Noether's theorem. According to Noether's theorem, the conservation of energy arises from time symmetry of the system. In fact, a more accurate term for "invariance" is "symmetry", and "change of variables" can be understood as "group actions". Noether's theorem illustrates the connection between symmetries and conservation laws. Three 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, beginning with observations in experimental physics. Scientists often first observe a conserved quantity in experiments and then search for the symmetry that generates it.

3. An example

We end this introduction with an example in one dimension. Recall that in 1-D problems, we use $t \in \mathbb{R}$ instead of x, write x in place of u, and use $\dot{x}$ instead of ∇u. Although $t \in \mathbb{R}$, the function $x(t)$ may be a curve in space; that is, x can be vector-valued. 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

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

Recall that the first-order condition for the finite-dimensional optimization problem (2) is $f'(x) = 0$. Now, however, $x = x(t)$ is itself a function, which may vary at every point $t \in (0,1)$. So, how do we take the derivative of a functional with respect to a function?

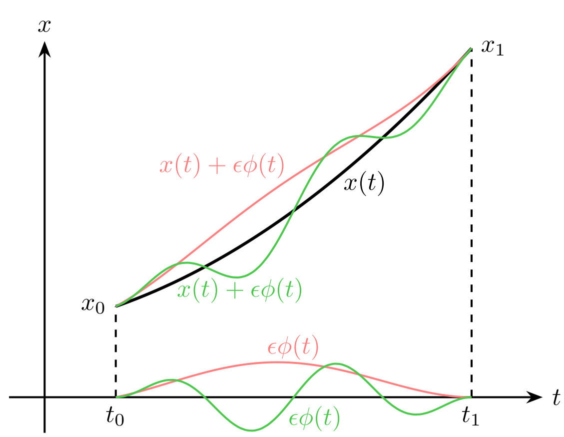

Figure 1. Variations of a curve.

The idea is to introduce a variation of the function. Let ϕ be a test function in $\mathcal{M}_0$ that satisfies certain boundary conditions so that if $x \in \mathcal{M}$, then $x + \epsilon\phi \in \mathcal{M}$. For the example considered,

$$\mathcal{M}_0 = \{\phi \in C^1(0,1) \mid \phi(0) = \phi(1) = 0\}.$$

The term $\epsilon\phi$ is a *variation* of x, and the symbol ϵ suggests it is small. Define

$$f(\epsilon) := I(x + \epsilon\phi).$$

Then x is a minimizer of (3) if and only if 0 is a minimizer of $f(\epsilon)$. So the optimality condition for an extremal curve x of $I(x)$ is characterized by

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

We now apply three standard techniques:

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

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

Integration by parts. Applying integration by parts to the second term and using the fact that $\phi(0) = \phi(1) = 0$, we obtain:

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

The boundary terms vanish due to the definition of $\mathcal{M}_0$. As $\mathcal{M}_0$ is dense in $L^2(0, 1)$, we conclude the strong form of the 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 the Lagrangian

$$L(t, x, v) = \frac{1}{2} m|v|^2 - U(x),$$

(6) becomes Newton's equation:

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

The derivation of Newton's equation from (3) is known as the principle of least action (in Hamilton's formulation).

Change of variables. Consider the change of variables

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

Assume L is strongly convex in v. Then L_v is invertible as a function of v, and we can solve $\dot{q} = \dot{q}(p, q, t)$. Define the Hamiltonian as

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

The variable $\dot{q}$ on the right-hand side can be expressed in terms of (p, q, t), so $\dot{q}$ is eliminated. The Euler–Lagrange equation becomes the Hamiltonian system

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

To derive (8), we apply total differentiation to both sides of (7). The left-hand side is

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

and the right-hand side is

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

The last step follows from the definition $p = L_v$ and the Euler–Lagrange equation $L_q = \dot{p}$ in terms of the new variables. Comparing the coefficients of $\mathrm{d}p$ and $\mathrm{d}q$ yields (8).

When the Hamiltonian $H(p, q)$ is independent of t, solutions to the Hamiltonian system satisfy the conservation of the 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. Conservation of the Hamiltonian corresponds to the conservation of total energy.