LECTURE 1

2. Calculus of variations

2.1. Some classical problems. Three classical problems:

Dido's isoperimetric problem: To build Carthage Dido bought land along the coastline that could be enclosed by the hide of an ox. She sliced the hide into many very thin strips that she tied together to obtain a rope of length l. To get the maximal area the shape of the rope along the coastline should maximize

$$J(y) = \int_a^b y(x)dx,$$

such that $y(a) = y(b) = 0$ and

$$C(y) = \int_a^b \sqrt{1 + (y'(x))^2}dx = l, \tag{2.1}$$

where a and b are start- and end-points on the coastline parameterized by x.

Catenary: When a homogeneous, flexible string of a prescribed length l is hanging in a vertical plane with its endpoints fixed at two points (a, y_0) and (b, y_1) it takes the shape that minimizes the potential energy

$$J(y) = \int_a^b y(x)\sqrt{1 + (y'(x))^2}dx,$$

where $y : [a, b] \to \mathbb{R}$ is the elevation of the chain as a function of the horizontal coordinate. If vertical positions y_0 and y_1 are located high enough above the ground, the constraints on y will be that $y(a) = y_0$, $y(b) = y_1$ and (2.1).

Brachistochrone: The birth of Calculus of variations is often said to be in June of 1696 when Johann Bernoulli posed the brachistochrone problem. The problem is that of finding the shape of a wire such that a frictionless bead slides in minimal time from the point (a, y_0) to (b, y_1), which are taken to be the start- and end-point of the wire ($y_0 \geq y_1$). If we translate the problem so that $y_0 = 0$ and flip the y–axis to point towards the ground, the speed of the beam is given by

$$mgy(x) = \frac{1}{2}mv^2,$$

from which we get

$$v = \sqrt{2gy(x)}.$$

The brachistochrone curve is thus the curve $y : [a, b] \to \mathbb{R}$ that minimizes

$$J(y) = \int_a^b \frac{\sqrt{1 + (y'(x))^2}}{\sqrt{y(x)}}dx,$$

with $y(a) = 0$ and $y(b) = y_0 - y_1$.

2.2. The basic problem and generalizations. The above problems are all examples of infinite-dimensional optimization problems. Instead of finding a vector $x^* \in \mathbb{R}^n$ such that $f(x^*) \leq f(x)$ for all $x \in \mathbb{R}$ we want to find a smooth function y^* such that $J(y^*) \leq J(y)$ for all other functions.

Basic Calculus of Variations Problem: Let $\mathcal{V} = \mathcal{C}^1([a, b] \to \mathbb{R})$ and define $\mathcal{A} = \{y \in \mathcal{V} : y(a) = y_0,\ y(b) = y_1\}$. Among all functions y in $\mathcal{A}$, find the one that minimizes

$$J(y) := \int_a^b L(x, y(x), y'(x))dx, \tag{2.2}$$

where $L : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, called the *Lagrangian* or *running cost*, is a sufficiently smooth function.

Generalizations:

- Variable-endpoint, $y(b) \in \mathbb{R}$.
- Free-time, variable-endpoint $y(t_f) = \varphi(t_f)$, $t_f \geq a$.
- Integral constraints
$$C(y) = \int_a^b M(x, y(x), y'(x))dx = C_0.$$
- Non-integral constraints
$$M(x, y(x), y'(x)) = 0.$$
- y piece-wise $\mathcal{C}^1$.

2.3. **First variation and optimality.** In finite-dimensional optimization a necessary condition for optimality is that
$$\nabla f(x^*) = 0,$$
when f is a smooth function. Can we find a similar condition for our type of problems?

First we note that $\nabla f(x)$ can be defined in the following way. Given a vector $d \in \mathbb{R}^n$, consider the vector
$$x + \alpha d.$$
We have that
$$f(x + \alpha d) = f(x) + \alpha(\nabla f(x))^\top d + o(\alpha).$$

The operator $J : \mathcal{V} \to \mathbb{R}$ is called a *functional.* Consider functions of the form
$$y(x) + \alpha\eta(x),$$
where α is a small number and the function $\eta \in \mathcal{V}$ called a perturbation. The linear functional $\delta J|_y : \mathcal{V} \to \mathbb{R}$ defined by
$$J(y + \alpha\eta) = J(y) + \alpha\ \delta J|_y(\eta) + o(\alpha),$$
for all η and all α is called the *first variation* of J.

Linearity here implies that
$$\delta J|_y(\alpha_1\eta_1 + \alpha_2\eta_2) = \alpha_1\ \delta J|_y(\eta_1) + \alpha_2\ \delta J|_y(\eta_2).$$

A first order necessary condition for y^* to be an optimal solution to the basic calculus of variations problem can be stated as
$$\delta J|_{y^*}(\eta) = 0, \tag{2.3}$$
for all $\eta \in \mathcal{V}$, with $\eta(a) = \eta(b) = 0$.

Example Let $\varphi \in \mathcal{C}^1([a, b] \to \mathbb{R})$ and define $J(y) = \int_a^b \varphi(y(x))dx$. Then
$$\begin{aligned} J(y + \alpha\eta) &= \int_a^b \varphi(y(x) + \alpha\eta(x))dx \\ &= \int_a^b (\varphi(y(x)) + \varphi'(y(x))\alpha\eta(x) + o(\alpha))dx \\ &= \int_a^b \varphi(y(x))dx + \alpha \int_a^b \varphi'(y(x))\eta(x)dx + o(\alpha) \end{aligned}$$
Hence,
$$\delta J|_y(\eta) = \int_a^b \varphi'(y(x))\eta(x)dx.$$

□

Optimality in what sense? We know that y^* is a local optimum if $J(y^*) \leq J(y)$ for all $y \in \mathcal{A}$ such that $\|y^* - y\| < \varepsilon$ for some $\varepsilon > 0$. Clearly (2.3) is a condition for a local extremum, but under what norm?

Two different norms:

0-norm: $\|y\|_0 = \max_{x\in[a,b]} |y(x)|$.

1-norm: $\|y\|_1 = \max_{x\in[a,b]} |y(x)| + \max_{x\in[a,b]} |y'(x)|$.

Let $B_i(y,\varepsilon) = \{z \in \mathcal{V} : \|z - y\|_i < \varepsilon\}$, for $i = 0, 1$. Then $B_1(y,\varepsilon) \subset B_0(y,\varepsilon)$, but given $\varepsilon > 0$, there is no $\varepsilon' > 0$ such that $B_0(y,\varepsilon') \subset B_1(y,\varepsilon)$. Hence, the norms $\|\cdot\|_0$ and $\|\cdot\|_1$ are not equivalent.

The 0-norm and the 1-norm define *strong* and *weak optimality* respectively.

Since all $\eta \in \mathcal{V}$ have a bounded derivative, $y + \alpha\eta$ is close to y in the sense of the 1-norm for small α. Hence, (2.3) is a necessary condition for weak optimality.

2.4. **The Euler-Lagrange equation.** For the cost functional we have that

$$\begin{aligned} J(y+\alpha\eta) &= \int_a^b L(x, y(x) + \alpha\eta(x), y'(x) + \alpha\eta'(x))dx \\ &= \int_a^b L(x, y(x), y'(x))dx + \int_a^b (L_y(x, y(x), y'(x))\alpha\eta(x) + L_{y'}(x, y(x), y'(x))\alpha\eta'(x))dx + o(\alpha), \end{aligned}$$

where L_y is the partial derivative of L with respect to the second variable and $L_{y'}$ is the partial derivative of L with respect to the third variable.

Hence,

$$\delta J\big|_y(\eta) = \int_a^b (L_y(x, y(x), y'(x))\eta(x) + L_{y'}(x, y(x), y'(x))\eta'(x))dx.$$

Using partial integration the last part under the integral sign can be re-written as

$$\int_a^b L_{y'}(x, y(x), y'(x))\eta'(x)dx = L_{y'}(x, y(x), y'(x))\eta(x)\big|_a^b - \int_a^b \frac{d}{dx} L_{y'}(x, y(x), y'(x))\eta(x)dx,$$

where, since $\eta(a) = \eta(b) = 0$ the first part on the RHS equals zero.

A zero first variation is thus equivalent to having

$$\int_a^b (L_y(x, y(x), y'(x)) - \frac{d}{dx} L_{y'}(x, y(x), y'(x)))\eta(x)dx = 0. \tag{2.4}$$

Assuming that $L_y(x, y(x), y'(x)) - \frac{d}{dx} L_{y'}(x, y(x), y'(x))$ is continuous, the fact that η is an arbitrary continuous function with start and end in zero, (2.4) means that the *Euler-Lagrange equation*

$$\boxed{L_y(x, y(x), y'(x)) - \frac{d}{dx} L_{y'}(x, y(x), y'(x)) = 0} \tag{2.5}$$

holds for all $x \in [a, b]$.

This is easily realized by observing that if for some $\bar{x}$, $L_y(\bar{x}, y(\bar{x}), y'(\bar{x})) - \frac{d}{dx} L_{y'}(\bar{x}, y(\bar{x}), y'(\bar{x})) \neq 0$, then $L_y(x, y(x), y'(x)) - \frac{d}{dx} L_{y'}(x, y(x), y'(x)) \neq 0$ on some interval containing $\bar{x}$ (by continuity).

Example We shove that solutions to the brachistochrone problem take the form

$$x(\theta) = a + c(\theta - \sin\theta),$$
$$y(\theta) = c(1 - \cos\theta).$$

The Lagrange equation for the brachistochrone problem is not explicitly a function of x,

$$L(y', y) = \frac{\sqrt{1 + (y'(x))^2}}{\sqrt{y(x)}}.$$

The Euler-Lagrange equation then takes the form

$$L_y(y(x), y'(x)) - \frac{d}{dx} L_{y'}(y(x), y'(x)) = L_y - L_{y'y}y' - L_{y'y'}y'' = 0.$$

Multiplying both sides by y' and changing the sign we get

$$0 = L_{y'y}(y')^2 + L_{y'y'}y''y' - L_y y' = \frac{d}{dx}(L_{y'}y' - L),$$

Thus in the case when L is not a function of x we have that $L_{y'}y' - L = C$ (constant *Hamiltonian*). For the brachistochrone problem this means that

$$\begin{aligned} &\frac{(y'(x))^2}{\sqrt{y(x)(1 + (y'(x))^2)}} - \frac{\sqrt{1 + (y'(x))^2}}{\sqrt{y(x)}} = C \\ \Rightarrow &(y'(x))^2 - (1 + (y'(x))^2) = C\sqrt{y(x)(1 + (y'(x))^2)} \\ \Rightarrow &y(x)(1 + (y'(x))^2) = \frac{1}{C^2}. \end{aligned}$$

We have

$$\frac{dx}{d\theta} = c(1 - \cos\theta) \quad \text{and} \quad \frac{dy}{d\theta} = c\sin\theta.$$

Hence,

$$\frac{dy}{dx} = \frac{dy}{d\theta}\frac{d\theta}{dx} = \frac{\sin\theta}{(1 - \cos\theta)}$$

and

$$\begin{aligned} y(x)(1 + (y'(x))^2) &= c(1 - \cos\theta)\left(1 + \frac{\sin^2\theta}{(1 - \cos\theta)^2}\right) \\ &= c\left(1 - \cos\theta + \frac{\sin^2\theta}{(1 - \cos\theta)}\right) \\ &= c\left(1 - \frac{-cos\theta + \cos^2\theta + \sin^2\theta}{(1 - \cos\theta)}\right) \\ &= c\left(1 - \frac{-cos\theta + 1}{(1 - \cos\theta)}\right) = 2c. \end{aligned}$$

□

2.5. **Variable-endpoint.** If we instead consider $y(a) = y_0$, $y(b) \in \mathbb{R}$ we get

$$\begin{aligned} \delta J|_y(\eta) = &\int_a^b \left(L_y(x, y(x), y'(x)) - \frac{d}{dx}L_{y'}(x, y(x), y'(x))\right)\eta(x)dx \\ &+ L_{y'}(b, y(b), y'(b))\eta(b) = 0. \end{aligned}$$

The same perturbations as before are still valid, hence (2.5) still holds. But $\eta(b)$ is no longer restricted to be zero. For the first-variation to be zero for all allowed perturbations η we must thus add the boundary condition

$$L_{y'}(b, y(b), y'(b)) = 0.$$