Variational Principles

Adam Kelly (`ak2316@cam.ac.uk`)

July 22, 2021, to Lecture 5

From finding geodesics to expressing the fundamental laws of nature, variational principles and the calculus of variations arise naturally in pure & applied mathematics, and theoretical physics. Here, we will study such ideas, with the underlying goal of considering problems involving minimizing (or maximizing) quantities that depend on an entire function.

This article constitutes my notes for the 'Variational Principles' course, held in Easter 2021 at Cambridge. These notes are *not a transcription of the lectures*, and differ significantly in quite a few areas. Still, all lectured material should be covered.

Contents

§1 Introduction

We will begin our study of variational principles by getting a feel for the type of problems that we care about. As a field, the study of variational principles can be traced back to a single problem, posed by Johann Bernoulli in 1696. It is there that we will begin.

§1.1 A Motivating Problem – The Brachistochrone

One of the earliest problems in the calculus of variations is the **Brachistochrone** problem.

Example 1.1 (Brachistochrone)

Consider a particle sliding on a wire between two fixed points A and B, under the influence of gravity. Which shape of the wire will give the shortest travel time, when the particle starts from rest?

It is not immediately obvious what the answer to this problem is. We won't solve it right now, but we can still can take some steps in the right direction. We first setup a coordinate system, setting A to be the origin, and B to be some point (x_2, y_2).
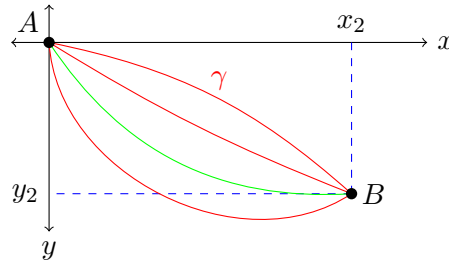
In this coordinate system, for a given curve γ, we can write down the travel time

$$T = \int_\gamma \mathrm{d}t = \int_\gamma \frac{1}{v}\,\mathrm{d}s,$$

where v is the velocity of the particle. Since the particle starts at $y = 0$, conservation of energy implies that

$$\frac{1}{2}mv^2 + mgy = 0$$
$$\implies v = \sqrt{-2gy}.$$

Thus to solve our problem, we really want to minimize

$$T[y] = \frac{1}{\sqrt{2g}} \int_0^{x_2} \frac{\sqrt{1 + (y')^2}}{\sqrt{-y}}\,\mathrm{d}x,$$

subject to y being a smooth function with $y(0) = 0$, and $y(x_2) = y_2$.

§1.2 Problems of Interest

Let's jump to another problem from pure mathematics, finding geodesics.

Example 1.2 (Geodesic Problem)

Find the shortest path γ between two points on a surface Σ (if one exists).

If Σ is $\mathbb{R}^2$, we can again adopt a coordinate system, and try to find the shortest curve γ between the two points $A = (x_1, y_1)$ and $B = (x_2, y_2)$. To solve this problem, we really want to minimize

$$D[y] = \int_{x_1}^{x_2} \sqrt{1 + (y')^2}\,\mathrm{d}x,$$

subject to y being a smooth function with $y(x_1) = y_1$ and $y(x_2) = y_2$.

Both of these problems have a similar flavour: we are trying the minimize or maximize something like

$$F[y] = \int_{x_1}^{x_2} f(x, y(x), y'(x)) \, \mathrm{d}x, \tag{$*$}$$

among all functions such that $y(x_1) = y_1$ and $y(x_2) = y_2$. This quantity $(*)$ is called a **functional**, that is, a function on the space of functions[1].

In this article, we are going to study the *calculus of variations*, in which we will be finding extrema (minima, maxima, and saddle points) of functionals. We will start by looking back at and extending our knowledge of extrema in finite dimensions, before using this knowledge to approach problems involving functionals.

Notation. Throughout this article, we will frequently refer to different spaces of functions. We will let $C(\mathbb{R})$ be the space of continuous functions on $\mathbb{R}$, and $C^k(\mathbb{R})$ be the space of functions with continuous kth derivative. We will also write $C_{(\alpha,\beta)}^k(\mathbb{R})$ for the subset of functions f with $f(\alpha) = f(\beta)$.

§2 Calculus for Functions on $\mathbb{R}^n$

For this section, we are going to consider a function $f \in C^2(\mathbb{R}^n)$, that is, $f : \mathbb{R}^n \to \mathbb{R}$, with continuous second derivative.

§2.1 Stationary Points

Recall the definition of a stationary point for a function $f \in C^2(\mathbb{R}^n)$.

> **Definition 2.1** (Stationary Point)
> We say that $a \in \mathbb{R}^n$ is a **stationary point** of f if $\nabla f(a) = 0$.

If a is a stationary point of f, then we can look at the Taylor expansion of f around a to determine the nature of the stationary point. For x near a, usiing the summation convention we have

$$f(x) = f(a) + \underbrace{(x - a) \cdot \nabla f(a)} + \frac{1}{2}(x_i - a_i)(x_j - a_j) \, \partial_{ij}^2 f\big|_a + O(|x - a|^2),$$

where $\partial_i f = \partial f / \partial x_i$.

Now let $H_{ij} = \partial_{ij}^2 f$ be the Hessian matrix, which we note is symmetric. If we shift the origin such that $a = 0$ and write $H = H(0)$, we can diagonalize this matrix to get

$$H' = R^T H(0) R = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix}.$$

Then we have

$$f(x') - f(0) = \frac{1}{2} \sum_{i=1}^{n} \lambda_i (x_i')^2 + O(|x'|^2).$$

---

[1] In standard calculus, we usually consider functions which map numbers $\to$ numbers. In the calculus of variations, we consider *functional* which map functions $\to$ numbers. An intuitive example is that of arc-length, which takes some curve and gives back a number corresponding to its length.

The behavior then depends on if the eigenvalues of $H(0)$ are positive or negative:

(i) If all $\lambda_i > 0$, then $f(x') > f(0)$ in all directions, and we have a local minimum.

(ii) If all $\lambda_i < 0$, then $f(x') < f(0)$ in all directions, and we have a local maximum.

(iii) If some of the eigenvalues are positive and some are negative, then $f(x')$ is increasing in some directions and is decreasing in others, and we have a saddle point.

(iv) If some (or all) of the eigenvalues are 0, then we need to consider higher order derivatives in the Taylor expansion.

In the special case where $n = 2$, then $\det H = \lambda_1 \lambda_2$ and $\operatorname{tr} H = \lambda_1 + \lambda_2$. From this we obtain the conditions:

- If $\det > 0, \operatorname{tr} > 0$ we have a local minimum.

- If $\det > 0, \operatorname{tr} < 0$ we have a local maximum.

- If $\det < 0$ we have a saddle point.

- If $\det = 0$, then we need to look at higher derivatives.

Remark. If a function $f : D \to \mathbb{R}$ is defined on some domain D, then the minimum or maximum could be obtained on the boundary of D, where we may not have $\nabla f = 0$.

Example 2.2 (Finding and Classifying Stationary Points)

Consider the function $f(x, y) = x^3 + y^3 - 3xy$. Then at a stationary point we have

$$\nabla f = \begin{pmatrix} 3x^2 - 3y \\ 3y^2 - 3x \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

These imply that at a stationary point, $y^4 = y$. Thus we have the points $(0, 0)$ and $(1, 1)$. We can write down the Hessian as

$$H = \begin{pmatrix} 6x & -3 \\ -3 & 6y \end{pmatrix}.$$

At $(0, 0)$, we have $\det H = -9 < 0$, so f has a saddle point there. At $(1, 1)$, we have $\det H = 27 > 0$, and $\operatorname{tr} H = 12 > 0$, so f has a local minimum there.

§2.2 Constraints and Lagrange Multipliers

Consider the following problem, and two natural ways to solve it.

Example 2.3

Find the circle centered at $(0, 0)$ with smallest radius such that the circle intersects the parabola $y = x^2 - 1$.

Direct Solution. One way to solve this problem is to substitute the constraint $y = x^2 - 1$ into the expression we are trying to minimize.

$$r^2 = x^2 + y^2 = x^2 + (x^2 - 1)^2 = x^4 - x^2 + 1.$$

To have a stationary point, the derivative of the RHS must be zero, so $4x^3 - 2x = 0$. This gives us two solutions: $x = \pm 1/\sqrt{2}, y = -1/2$ and $x = 0, y = -1$. These give the two possible radii of $\sqrt{3}/2$ and 1, of which the first is the smallest. $\qquad\square$

While this solution works, it's not hard to imagine that we may not be able to say solve for y. An alternate solution is given below.

Lagrange Multipliers Solution. Define the function $h(x, y, \lambda) = f(x, y) - \lambda g(x, y)$, where $f(x, y) = x^2 + y^2 = r^2$, $g(x, y) = 0$ is the constraint, and λ is the Lagrange multiplier. So $x^2 + y^2 - \lambda(y - x^2 + 1)$. We are going to extremize this over 3 variables with no constraints.

$$\frac{\partial h}{\partial x} = 2x + 2\lambda x = 0,$$
$$\frac{\partial h}{\partial y} = 2y - \lambda = 0,$$
$$\frac{\partial h}{\partial \lambda} = y - x^2 + 1 = 0,$$

where the third condition is our constraint. From these we get either $x = 0, y = -1$ and $f = 1, \lambda = -2$ or $y = -1/2, x = \pm 1/\sqrt{2}$ and $f = 3/4, \lambda = -1$, giving the radius $\sqrt{3}/2$. $\qquad\square$

So why does the second solution work? Well suppose that we wanted to minimize a function f subject to $g = 0$. Then ∇g is perpendicular to $g = 0$. Also, ∇f is perpendicular to $f(x) = c$, for any constant c. So at the extremum, we must have that ∇f and ∇g are parallel, that is,

$$\nabla f - \lambda \nabla g = 0,$$

for some λ. Thus it suffices to consider the extrema of $h = f - \lambda g$.

This method generalizes to multiple variables and multiple constraints. For example, if we wanted to extremize $f : \mathbb{R}^n \to \mathbb{R}$, subject to $g_\alpha(x) = 0$, where $g_\alpha : \mathbb{R}^n \to \mathbb{R}$ and $\alpha = 1, \ldots, k$, we would define

$$h(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_k) = f - \sum_{\alpha=1}^{k} \lambda_\alpha g_\alpha,$$

and would extremize this function with respect to the $n + k$ variables.

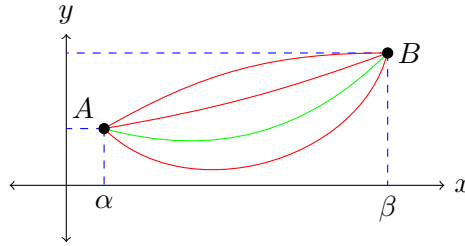## §3   The Euler-Lagrange Equations

We are now going to move on to deriving some necessary conditions for the extrema of a functional.

§3.1   Deriving the Euler-Lagrange Equations

We want to extremize the functional $(*)$,

$$F[y] = \int_\alpha^\beta f(x, y(x), y'(x)) \, \mathrm{d}x, \tag{$*$}$$

where we assume that f is given and depends on y, with fixed endpoints.

Our strategy will be to consider some small perturbation $y \mapsto y + \epsilon\eta(x)$, where $\eta(\alpha) = \eta(\beta) = 0$. From this, we will compute $F[y + \epsilon\eta]$. To proceed with this strategy, we will need an important lemma.

> **Lemma 3.1** (Fundamental Lemma of Calculus of Variations)
>
> If $g : [\alpha, \beta] \to \mathbb{R}$ is continuous on $[\alpha, \beta]$ and is such that
>
> $$\int_\alpha^\beta g(x)\eta(x) \, \mathrm{d}x = 0,$$
>
> for all η continuous on $[\alpha, \beta]$ with $\eta(\alpha) = \eta(\beta) = 0$, then $g(x) = 0$ for all $x \in [\alpha, \beta]$.

> *Proof.* Say (for a contradiction) that there exists some $\bar{x} \in (\alpha, \beta)$ such that $g(\bar{(x)}) = 0$. Say $g(\bar{x}) > 0$. Then by continuity there exists an interval $[x_1, x_2] \subset [\alpha, \beta]$ such that $g(x) > c$ on $[x_1, x_2]$ for some $c > 0$. Then set
>
> $$\eta(x) = \begin{cases} (x - x_1)(x_2 - x) & \text{if } x \in [x_1, x_2], \\ 0 & \text{if } x \notin [x_1, x_2]. \end{cases}$$
>
> Then we have
>
> $$\int_\alpha^\beta g(x)\eta(x)) \, \mathrm{d}x > c \int_{x_1}^{x_2} (x - x_1)(x_2 - x) \, \mathrm{d}x > 0,$$
>
> which is a contradiction. $\qquad\square$

Remark. In the proof above, we construct a *bump function* η. It is possible to make this smoother by taking various powers of the constructed function.

Now let's return to our computation. We have

$$F[y + \epsilon\eta] = \int_\alpha^\beta f(x, y + \epsilon\eta, y' + \epsilon\eta') \, \mathrm{d}x$$

$$= F[y] + \epsilon \int_\alpha^\beta \left(\frac{\partial f}{\partial y}\eta + \frac{\partial f}{\partial y'}\eta' \right) \, \mathrm{d}x + O(\epsilon^2).$$

We will return to the $O(\epsilon^2)$ term later, but for now we note that for an extremum of the functional, we want this first-order term to vanish, so we want something like

$$\left. \frac{dF[y + \epsilon]}{d\epsilon} \right|_{\epsilon=0} = 0.$$

We now make a move which we will return to time and time again in the calculus of

variations. Integrating the ϵ-term by parts,

$$
\begin{aligned}
0 &= \int_\alpha^\beta \frac{\partial f}{\partial y}\eta - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right)\eta \ dx + \left[\frac{\partial f}{\partial y'}\eta\right]_\alpha^\beta \\
&= \int_\alpha^\beta \left(\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right)\right)\eta \ dx.
\end{aligned}
$$

Applying the fundamental lemma of the calculus of variations, we obtain the following necessary condition for an extremum.

Theorem 3.2 (Euler-Lagrange Equation)

Suppose $y \in C^2_{(\alpha,\beta)}(\mathbb{R})$ is a function with fixed endpoints that extremizes the functional

$$
F[y] = \int_\alpha^\beta f(x, y(x), y'(x)) \ dx.
$$

Then y must satisfy the **Euler-Lagrange** equation

$$
\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) = 0. \tag{$\dagger$}
$$

Notation. The LHS of the ($\dagger$) is denoted by $\dfrac{\delta F[y]}{\delta y(x)}$ and is the **functional derivative**.

§3.2 First Integrals of the Euler-Lagrange Equations

Of course, now that we have a necessary condition for externality, our next concern is solving the Euler-Lagrange equation. The Euler-Lagrange equations give us a second-order ODE for y. It is sometimes possible to reduce the order of the ODE to get a first-order ODE, the **first integral**, which is what we will look at in this section.

If f does not explicitly depend on y, then $\partial f/\partial y = 0$. Then we get

$$
\frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) = 0 \implies \frac{\partial f}{\partial y'} = c,
$$

for some constant c.

Example 3.3 (Geodesics on the Euclidean Plane)

Consider again the problem of finding geodesics on the Euclidean plane.

We want to minimize the functional

$$
D[y] = \int_{x_1}^{x_2} \sqrt{1 + (y')^2} \ dx.
$$

Using the Euler-Lagrange equations, we have $f(y') = \sqrt{1 + (y')^2}$, so $\partial f/\partial y = 0$, and thus by the result above

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

---

for some constant c. For this to be the case, we must have y' constant, so y is a straight line.

Example 3.4 (Geodesics on the Sphere)

Consider the sphere embedded in three-dimensional Euclidean space, $S^2 \subset \mathbb{R}^3$, and two points A and B on the sphere. We want to find the shortest path between A and B.
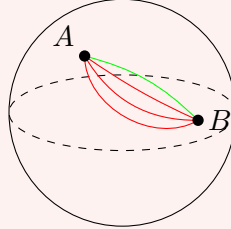
We first parameterize the sphere, with

$$
\begin{aligned}
x &= \sin\theta\sin\phi, \\
y &= \sin\theta\cos\phi, \\
z &= \cos\theta,
\end{aligned}
$$

where $0 < \theta \le \pi$ and $0 \le \phi \le 2\pi$. Then to find lengths of paths on the sphere, we use the line element

$$
\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2.
$$

Now we can parameterize a curve by thinking of ϕ as a function of θ. Then we have

$$
\mathrm{d}s = \sqrt{1 + \sin^2\theta \cdot (\phi')^2}\,\mathrm{d}\theta.
$$

Thus the functional we want to extremize is

$$
F[\phi] = \int_\alpha^\beta \sqrt{1 + \sin^2\theta \cdot (\phi')^2}\,\mathrm{d}\theta.
$$

We note here that $\partial f/\partial\theta = 0$, so using the result we obtained above, $\partial f/\partial\phi' = \kappa$, for some constant κ. That is,

$$
\frac{\sin^2\theta \cdot \phi'}{\sqrt{1 + \sin^2\theta \cdot (\phi')^2}} = \kappa.
$$

Squaring this and solving for $(\phi')^2$, we get[a]

$$
\phi'^2 = \frac{\kappa^2}{\sin^2\theta \cdot (\sin^2\theta - \kappa^2)} \implies \phi = \pm\int \frac{\kappa^2}{\sin^2\theta \cdot (\sin^2\theta - \kappa^2)}\,\mathrm{d}\theta.
$$

From here, we can take a substitution of $\cot\theta = u$ to obtain the equation of a the shortest (and longest) paths.

---

[a]There is a sign ambiguity, corresponding to going either forwards or backwards around the sphere.

Now consider, for general $f(x, y, y')$,

$$\frac{d}{dx}\left(f - y'\frac{\partial f}{\partial y'}\right) = \frac{\partial f}{\partial x} + y'\frac{\partial f}{\partial y} + y''\frac{\partial f}{\partial y'} - y''\frac{\partial f}{\partial y'} - y'\frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right)$$

$$= y'\left(\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'}\right) + \frac{\partial f}{\partial x}$$

$$= \frac{\partial f}{\partial x},$$

where we use the Euler-Lagrange equations. So if f instead does not depend on x, then $\partial f/\partial x = 0$, and

$$f - y'\frac{\partial f}{\partial y'} = c,$$

for some constant c (this is another first integral).

> **Example 3.5** (The Brachistochrone)
>
> Consider again the problem of finding the shape of a wire which gives the shortest travel time for a particle sliding along it, beginning at rest.
>
> We want to minimize the functional
>
> $$T[y] = \frac{1}{\sqrt{2g}}\int_0^{x_2} \frac{\sqrt{1 + (y')^2}}{\sqrt{-y}}\, dx,$$
>
> subject to $y(x_1) = y_1$ and $y(x_2) = y_2$. Here, f does not depend on x and thus we can apply the result above to get
>
> $$\frac{\sqrt{1 + (y')^2}}{\sqrt{-y}} - y'\frac{y'}{\sqrt{1 + (y')^2}\sqrt{-y}} = \kappa,$$
>
> for some constant κ. We can solve this differential equation by simplifying to get
>
> $$\frac{1}{\sqrt{1 + (y')^2}} = \kappa\sqrt{-y},$$
>
> and then we can separate variables to get
>
> $$x = \pm\kappa\int \frac{\sqrt{-y}}{\sqrt{1 + \kappa^2 y}}\, dy.$$
>
> The substitution $y = -\frac{1}{\kappa^2}\sin^2\frac{\theta}{2}$ then gives (after some algebra)
>
> $$x = \pm\frac{1}{\kappa^2}(\theta - \sin\theta) + C,$$
>
> for some constant C. Using the initial conditions, we get $C = 0$, and thus
>
> $$\begin{cases} x = \frac{\theta - \sin\theta}{2\kappa t} \\ y = -\frac{1}{\kappa^2}\sin^2\left(\frac{\theta}{2}\right) \end{cases}$$
>
> This is the parameterized equation of a Cycloid, the curve traced by a point on a wheel moving with constant velocity.

§3.3 Fermat's Principle

We can now look at the first variational principle in this course, which describes the motion of light.

> **Law 3.6** (Fermat's Principle)
>
> Light travels along paths between two points which requires the least time.

Suppose we had a ray with path $y = y(x)$, and suppose the speed of light at (x, y) is $c(x, y)$. Then we want to extremize the functional

$$F[y] = \int \frac{1}{c} \, \mathrm{d}s = \int_\alpha^\beta \frac{\sqrt{1 + (y')^2}}{c(x, y)} \, \mathrm{d}x.$$

If we assume that c depends only on x, then we have $\partial f / \partial y = 0$, and we can use the results of the previous section to obtain

$$\frac{\partial f}{\partial y'} = \frac{y'}{\sqrt{1 + (y')^2} c(x)} = k$$

for some constant k.

If the ray is launched at an angle θ_1 upwards, then $\tan \theta_1 = y'(\alpha)$. So if $\tan \theta = y'$, the above equation implies

$$\frac{\sin \theta_1}{c(x_1)} = \frac{\sin(\theta)}{c(x)},$$

which is known as *Snell's law* in optics.

§4 Extensions of the Euler-Lagrange Equations

We are now going to see how the Euler-Lagrange equations can be used in settings that are not just unconstrained functional extremization.

§4.1 Euler-Lagrange Equations with Constraints

Suppose we wanted to extremimze a functional $F[y]$ with

$$F[y] = \int_\alpha^\beta f(x, y, y') \, \mathrm{d}x,$$

subject to the constraint functional

$$G[y] = \int_\alpha^\beta g(x, y, y') \, \mathrm{d}x = k,$$

for a constant k. We are going to do this by generalizing the method of Lagrange multipliers that we introduced in subsection 2.2. We are going to extremimze the functional
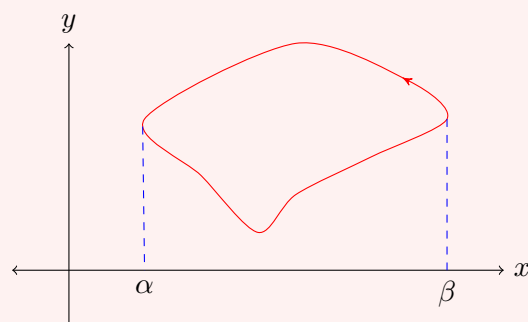
$$\Phi[y, \lambda] = F[y] - \lambda G[y].$$

To do this, we replace f in the Euler-Lagrange equations by $f - \lambda g$, to get

$$\frac{d}{dx} \left(\frac{\partial}{\partial y'} (f - \lambda g) \right) = \frac{\partial}{\partial y} (f - \lambda g) = 0.$$

Example 4.1 (Isopermetric/Dido Problem)

What simple closed plane curve of fixed length L maximizes the area enclosed by this curve?

We can see immediately see that such a curve has to be convex, and we can setup the problem using cartesian coordinates as follows.

On the x-axis, we have minimal and maximal values α and β, and a simple closed curve with the given orientation. The x coordinate monotonically increases from $\alpha \to \beta$, and then decreases from $\beta \to \alpha$.

Given x, there exists y_1, y_2 on the curve with $y_1(x) = y_2(x)$