

IB Variational Principles

Ishan Nath, Easter 2022

Based on Lectures by Dr. Maciej Dunajski

June 15, 2022

Contents

0	Motivation	2
0.1	Notation	3
1	Calculus for functions on \mathbb{R}^n	4
1.1	Constraints and Lagrange Multipliers	5
2	Euler-Lagrange Equations	7
2.1	First integrals of the Euler-Lagrange equations	9
2.2	Fermat's Principle	12
3	Extensions of the Euler-Lagrange equation	13
3.1	Euler-Lagrange with constraints	13
3.2	Sturm-Liouville Problem	14
3.3	Several dependent variables	15
3.4	Several independent variables	16
3.5	Higher derivatives	20
4	Least action principle and Noether's theorem	21
4.1	Noether's theorem	22
5	Convex functions	24
5.1	Conditions for convexity	24
6	Legendre transform	26
6.1	Applications to Thermodynamics	27
7	Hamilton's equations	28
8	The Second Variation	30
8.1	Associated eigenvalue problem	31
8.2	The Jacobian condition	32
	Index	34

0 Motivation

The theory of variational principles is rooted in the many problems that we can solve with it. To motivate it, we will look at such problems.

Example 0.1 (The Brachistochrone Problem). A particle slides on a wire under the influence of gravity, between two fixed points $A = (x_1, y_1)$ and $B = (x_2, y_2)$. Which shape of the wire gives the shortest travel time, starting from rest?

This was considered by Johann Bernoulli, in 1696. To solve, first let $A = (0, 0)$. Now note that the travel time is

$$T = \int_A^B \frac{dl}{v(x, y)}.$$

Now considering the conservation of energy,

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

This gives

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

subject to $y(0) = 0$, and $y(x_2) = y_2$.

Example 0.2 (Geodesics). What is the shortest path γ between two points A, B on a surface?

Take $\Sigma = \mathbb{R}^2$ as our example, and let $A = (x_1, y_1)$ and $B = (x_2, y_2)$. The distance along a curve γ , parametrised by x is

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

We seek to minimise D by varying γ .

In general, variational principles is focused on minimizing (or maximizing) a function

$$F[y] = \int_{x_1}^{x_2} f(x, y, y') dx$$

among all functions such that $y(x_1) = y_1$, $y(x_2) = y_2$. This is an example of a **functional**. The same way a function is a map from numbers to numbers, a functional is a map from functions to numbers.

The calculus of variations is focused on finding the extrema of functionals on spaces of functions.

Variational calculus is also important physically, as we can see from the following examples.

Example 0.3 (Fermat's Principle). Light between two points travels along paths which require the least time.

Example 0.4 (Principle of Least Action). Let T be the kinetic energy ($m|\dot{\mathbf{x}}|^2/2$ in Newtonian mechanics), and V be the potential energy. The **action** is defined as

$$S[\gamma] = \int_{t_1}^{t_2} (T - V) dt.$$

Then the action is minimized along paths of motion.

In this course, we will look at:

- Necessary conditions for extrema of functionals, and the Euler-Lagrange equations.
- Lots of examples, from geometry to physics, and problems with constraints, such as the isoperimetric problem.
- Second variations, which are sufficient conditions for extrema.

0.1 Notation

There is some standard notation that we will be using.

- $C(\mathbb{R})$ is the space of all continuous functions on \mathbb{R} .
- $C^k(\mathbb{R})$ is the space of all functions on \mathbb{R} with continuous k 'th derivatives.
- $C_{(\alpha, \beta)}^k(\mathbb{R})$ is the space of all functions on \mathbb{R} with continuous k 'th derivatives, such that $f(\alpha) = f(\beta)$.

1 Calculus for functions on \mathbb{R}^n

We will begin by looking at functions $f \in C^2(\mathbb{R}^n)$, $f : \mathbb{R}^n \rightarrow \mathbb{R}$. The point $\mathbf{a} \in \mathbb{R}^n$ is **stationary** if

$$\nabla f(\mathbf{a}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right) \Big|_{\mathbf{x}=\mathbf{a}} = \mathbf{0}.$$

Expanding near \mathbf{a} , we get

$$f(\mathbf{x}) = f(\mathbf{a}) + \underbrace{(\mathbf{x} - \mathbf{a})\nabla f(\mathbf{a})}_{0, \text{ as } \mathbf{a} \text{ is stationary}} + \frac{1}{2}(x_i - a_i)(x_j - a_j)\partial_{ij}^2 f(\mathbf{a}) + \mathcal{O}(|\mathbf{x} - \mathbf{a}|^3).$$

Here, the order 2 term is the **Hessian matrix**

$$H_{ij} = \partial_i \partial_j f = H_{ji}.$$

Shift the origin to set $\mathbf{a} = \mathbf{0}$, and since H is symmetric, we can diagonalise $H(\mathbf{0})$ by an orthogonal transformation

$$H' = R^T H(\mathbf{0}) R = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

Then, returning to our expansion, we get that

$$f(\mathbf{x}) - f(\mathbf{0}) = \frac{1}{2} \sum_i \lambda_i x_i^2 + \mathcal{O}(|\mathbf{x}|^3).$$

Depending on the eigenvalues of the Hessian, we can have the following occur:

- (i) If all the eigenvalues are positive, then $f(\mathbf{x}) > f(\mathbf{0})$ in all directions, so we have a **local minimum**.
- (ii) If all the eigenvalues are negative, then similarly we have a **local maximum**.
- (iii) If some eigenvalues are positive, and some are negative, then f increases in some directions, and decreases in others, so we have a **saddle point**.
- (iv) If some eigenvalues are zero, then we must consider higher order terms in the expansion.

In two dimensions, $\det H = \lambda_1 \lambda_2$, and $\text{tr } H = \lambda_1 + \lambda_2$. We can thus efficiently determine which case we are in:

- If $\det > 0$ and $\text{tr} > 0$, then we have a local minimum.
- If $\det > 0$ and $\text{tr} < 0$, then we have a local maximum.
- If $\det < 0$, we have a saddle point.
- If $\det = 0$, then we must look at higher derivatives.

Remark.

- 1) If $f : \mathbb{D} \rightarrow \mathbb{R}$, where \mathbb{D} is bounded, then the global minima or maxima may not be a local maxima or minima, but may be a boundary point.
- 2) If f is harmonic on \mathbb{R}^2 , meaning $f_{xx} + f_{yy} = 0$, then $\text{tr} H = 0$ for all points, meaning the maxima/minima must occur on the boundary.

Example 1.1. If $f(x, y) = x^3 + y^3 - 3xy$, then we can calculate

$$\nabla f = (3x^2 - 3y, 3y^2 - 3x) = (0, 0)$$

at points where $x^2 - y = 0$, and $y^2 - x = 0$, i.e. $y^4 = y$. Thus gives two solution: $(x, y) = (0, 0)$ or $(1, 1)$. We can also compute the Hessian matrix

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

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

Intuitively, this makes sense: near the origin $f \approx -3xy$, which is a saddle point.

1.1 Constraints and Lagrange Multipliers

We begin with an example.

Example 1.2. Find the circle centred at $(0, 0)$ with smallest radius, which intersects the parabola $y = x^2 - 1$.

There are two approaches:

- The direct method: let $f = x^2 + y^2 = x^2 + (x^2 - 1)^2 = x^4 - x^2 + 1$. The derivative is $4x^3 - 2x = 0$, which has roots at $x = 0$ and $x = \pm\sqrt{2}/2$.

In the first case, the radius is 1, and in the second it is $\sqrt{3}/2$, which is the global minimum.

- Lagrange multipliers: Define $h(x, y, \lambda) = f(x, y) - \lambda g(x, y)$, where g is the constraint, and λ is the Lagrange multiplier.

This gives a function $h = x^2 + y^2 - \lambda(y - x^2 + 1)$, which we extremize over all three variables:

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

$$\implies (x, y) = (0, -1) \text{ or } \left(\pm \frac{\sqrt{2}}{2}, -\frac{1}{2} \right).$$

This gives the same minima as before.

The question is: why does this work? The simple answer is geometry.

At each point on the constraint space, we can construct a normal vector ∇g , and similarly a normal vector to the objective function ∇f . At the extrema, we must have $\nabla f \parallel \nabla g$, implying $\nabla f = \lambda \nabla g$, or

$$\nabla(f - \lambda g) = 0.$$

In general, for a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, subject to $g_a(\mathbf{x}) = 0$, where there are k constraints, we can define a function h in $n + k$ variables, such that

$$h(x_1, \dots, x_n, \lambda_1, \dots, \lambda_k) = f(\mathbf{x}) - \sum_a \lambda_a g_a(\mathbf{x}).$$

Now to find extrema, we impose the conditions that

$$\frac{\partial h}{\partial x_i} = 0, \quad \frac{\partial h}{\partial \lambda_j} = 0.$$

2 Euler-Lagrange Equations

As said before, we wish to extremize the functional

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

where $y(\alpha) = y_1$, and $y(\beta) = y_2$.

Assume the extremum $y = y(x)$ exists. Consider a small perturbation

$$y \mapsto y + \varepsilon \eta(x),$$

where $\eta(\alpha) = \eta(\beta) = 0$. Then we wish to compute $F[y + \varepsilon \eta]$.

Lemma 2.1 (Fundamental Lemma of Variational Calculus). *If $g : [\alpha, \beta] \rightarrow \mathbb{R}$ is continuous on $[\alpha, \beta]$, and*

$$\int_{\alpha}^{\beta} g(x) \cdot \eta(x) dx = 0$$

for all continuous η on $[\alpha, \beta]$, with $\eta(\alpha) = \eta(\beta) = 0$, then $g \equiv 0$ on $[\alpha, \beta]$.

Proof: Say there exists $\bar{x} \in (\alpha, \beta)$ such that $g(\bar{x}) \neq 0$. Then without loss of generality, say $g(\bar{x}) > 0$.

Then there exists an interval $[x_1, x_2] \in (\alpha, \beta)$ such that $g(x) > c > 0$ on this interval. Then take

$$\eta(x) = \begin{cases} (x - x_1)(x_2 - x) & \text{if } x \in [x_1, x_2], \\ 0 & \text{if } x \notin [x_1, x_2]. \end{cases}$$

Then

$$\int_{\alpha}^{\beta} g(x) \eta(x) dx > c \int_{x_1}^{x_2} (x - x_1)(x_2 - x) dx > 0.$$

Remark. The function η is an example of a bump function. To find bump functions in C^k , take

$$\eta(x) = \begin{cases} [(x - x_1)(x_2 - x)]^{k+1} & x \in [x_1, x_2], \\ 0 & x \notin [x_1, x_2]. \end{cases}$$

Now we consider $F[y + \varepsilon\eta]$:

$$\begin{aligned} F[y + \varepsilon\eta] &= \int_{\alpha}^{\beta} f(x, y + \varepsilon\eta, y' + \varepsilon\eta') dx \\ &= F[y] + \varepsilon \int_{\alpha}^{\beta} \left[\frac{\partial f}{\partial y} \eta + \frac{\partial f}{\partial y'} \eta' \right] dx + \mathcal{O}(\varepsilon^2) \\ &= F[y] + \varepsilon \left\{ \int_{\alpha}^{\beta} \left[\frac{\partial f}{\partial y} \eta - \frac{d}{dx} \left(\frac{\partial f}{\partial y'} \right) \eta \right] dx + \left[\frac{\partial f}{\partial y'} \eta \right]_{\alpha}^{\beta} \right\} + \mathcal{O}(\varepsilon^2). \end{aligned}$$

Notice the last term is 0, since $\eta(\alpha) = \eta(\beta) = 0$. Therefore, since at extrema we have $F[y + \varepsilon\eta] = F[y] + \mathcal{O}(\varepsilon^2)$, we must have

$$0 = \int_{\alpha}^{\beta} \underbrace{\left[\frac{\partial f}{\partial y} - \frac{d}{dx} \left(\frac{\partial f}{\partial y'} \right) \right]}_g \eta dx.$$

Applying our previous lemma with g , this implies that

$$\frac{d}{dx} \left(\frac{\partial f}{\partial y'} \right) - \frac{\partial f}{\partial y} = 0. \quad (2)$$

This is known as the **Euler-Lagrange equation**, and is a necessary condition for an extrema. (For a bit of history, in 1745 Lagrange writes to Euler citing his method. Euler adopts this approach and (2), and so calculus of variations is born.)

Remark.

- (2) is a second order ODE for $y(x)$, with boundary conditions $y(\alpha) = y_1$, $y(\beta) = y_2$.
- The left hand side of (2) is denoted as

$$\frac{\delta F}{\delta y(x)},$$

and is called the functional derivative.

- We can have other boundary conditions, such as

$$\left. \frac{\partial f}{\partial y'} \right|_{\alpha, \beta} = 0.$$

- We have to be careful with derivatives: in this proof,

$$\frac{\partial f}{\partial y} = \left(\frac{\partial f}{\partial y} \right)_{x, y'},$$

with x, y, y' independent. The total derivative is defined as

$$\frac{d}{dx} = \frac{\partial}{\partial x} + y' \frac{\partial}{\partial y} + y'' \frac{\partial}{\partial y'},$$

provided we are working with functions in x, y, y' only.

Example 2.1. Take the function $f = x \cdot [(y')^2 - y^2]$. Then,

$$\partial_x f = (y')^2 - y^2, \quad \partial_y f = -2xy, \quad \partial_{y'} f = 2xy',$$

$$\frac{df}{dx} = (y')^2 - y^2 - 2xyy' + 2xy'y''.$$

2.1 First integrals of the Euler-Lagrange equations

In some cases the Euler-Lagrange equation (2) can be integrated once to a first order ODE, known as the **first integral**.

- (a) If f does not explicitly depend on y , then the Euler-Lagrange equation becomes

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

which when integrated gives

$$\frac{\partial f}{\partial y'} = \text{constant}. \quad (2.1)$$

With this, we can solve the problem of geodesics:

Example 2.2 (Geodesics on \mathbb{R}^2). In this case, our function is

$$F[y] = \int_{\alpha}^{\beta} \sqrt{dx^2 + dy^2} = \int_{\alpha}^{\beta} \underbrace{\sqrt{1 + (y')^2}}_{f(y')} dx.$$

Then applying (2.1),

$$\frac{\partial f}{\partial y} = 0 \implies \frac{\partial f}{\partial y'} = \frac{y'}{\sqrt{1 + (y')^2}} = \text{constant}.$$

Therefore y' is equal to a constant m , so $y = mx + c$ is an extrema.

Example 2.3 (Geodesics on a sphere). Consider $S^2 \subset \mathbb{R}^3$. In spherical polars θ and ϕ , the line element is

$$ds^2 = dx^2 + dy^2 + dz^2 = d\theta^2 + \sin^2 \theta d\phi^2.$$

Parametrise the curve by $\phi = \phi(\theta)$, then

$$ds = \sqrt{1 + \sin^2 \theta (\phi')^2} d\theta.$$

Then, the required functional is

$$F[\phi] = \int_{\theta_1=\alpha}^{\theta_2=\beta} \sqrt{1 + \sin^2 \theta (\phi')^2} d\theta.$$

We can thus apply (2.1):

$$\frac{\partial f}{\partial \theta} = 0 \implies \frac{\partial f}{\partial \phi'} = \frac{\sin^2 \theta \cdot \phi'}{\sqrt{1 + \sin^2 \theta (\phi')^2}} = \kappa.$$

Solving this gives

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

We can integrate by substituting $\cot \theta = u$, to give

$$\int \frac{\kappa d\theta}{\sin \theta \sqrt{\sin^2 \theta - \kappa^2}} = \int \frac{-\kappa}{\sqrt{1 - \kappa^2(u^2 + 1)}} du = \cos^{-1} \left(\frac{\kappa}{\sqrt{1 - \kappa^2}} u \right) + C.$$

Therefore,

$$\pm \frac{\sqrt{1 - \kappa^2}}{\kappa} \cos(\phi - \phi_0) = \cot \theta.$$

This is the intersection of a plane through the origin with the unit sphere; therefore, we have proven that the extrema are segments of great circles.

(b) Consider, for general $f(x, y, y')$,

$$\begin{aligned} \frac{d}{dx} \left(f - y' \frac{\partial f}{\partial y'} \right) &= \frac{\partial f}{\partial x} + y' \frac{\partial f}{\partial y} + y'' \frac{\partial f}{\partial y'} - y'' \frac{\partial f}{\partial y'} - y' \frac{d}{dx} \left(\frac{\partial f}{\partial y'} \right) \\ &= y' \underbrace{\left(\frac{\partial f}{\partial y} - \frac{d}{dx} \frac{\partial f}{\partial y'} \right)}_{0 \text{ by (2)}} + \frac{\partial f}{\partial x} = \frac{\partial f}{\partial x}. \end{aligned}$$

Hence, if $\partial f / \partial x = 0$, so f does not depend explicitly on x , then

$$f - y' \frac{\partial f}{\partial y'} = \text{constant}. \quad (2.2)$$

Example 2.4 (Brachistochrone). We return to example (0.1): recall that

$$F[y] = \frac{1}{\sqrt{2g}} \int_0^\beta \frac{\sqrt{1 + (y')^2}}{\sqrt{-y}} dx.$$

We can apply the first integral (2.2) to get

$$\begin{aligned} \frac{\sqrt{1 + (y')^2}}{\sqrt{-y}} - y' \frac{y'}{\sqrt{1 + (y')^2} \sqrt{-y}} &= \kappa, \\ \implies \frac{dy}{dx} = y' &= \pm \frac{\sqrt{1 + \kappa^2 y}}{\kappa \sqrt{-y}}. \end{aligned}$$

Writing this as an integral in x ,

$$x = \pm \kappa \int \frac{\sqrt{-y}}{\sqrt{1 + \kappa^2 y}} dy.$$

Now we can use the substitution

$$y = -\frac{1}{\kappa^2} \sin^2 \frac{\theta}{2}$$

to get

$$x = \mp \frac{1}{2\kappa^2} \int (1 - \cos \theta) d\theta = \mp \frac{1}{2\kappa^2} (\theta - \sin \theta) + c.$$

Since the curve goes through $(0, 0)$, we must have $c = 0$. Moreover we will take the positive root.

Therefore we get the parametrisation

$$\begin{cases} x = \frac{\theta - \sin \theta}{2\kappa^2}, \\ y = -\frac{1}{\kappa^2} \sin^2 \frac{\theta}{2}. \end{cases}$$

This is the parametrised form of a **cycloid**. This is the curve traced by a point on the boundary of a circle, as the circle rolls along an integer line.

Proving that the brachistochrone is a cycloid was first done by Bernoulli, in 1697.

2.2 Fermat's Principle

Recall Fermat's principle:

Proposition 2.1 (Fermat's principle). *Light and sound travels along paths between two points which require the least time.*

If we parametrise the ray as a path $y = y(x)$, and the speed of light as a function $c(x, y)$, then the appropriate functional is

$$F[y] = \int_A^B \frac{dl}{c} = \int_\alpha^\beta \frac{\sqrt{1 + (y')^2}}{c(x, y)} dx.$$

We assume that $c = c(x)$, then by the Euler-Lagrange equation, specifically (2.1), we get that

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

However, notice that at a point (x, y) , the quantity

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

the sine of the tangent line. Therefore, this equation can be written as

$$\frac{\sin \theta_1}{c(x_1)} = \frac{\sin \theta_2}{c(x_2)},$$

which is **Snell's law**.

3 Extensions of the Euler-Lagrange equation

3.1 Euler-Lagrange with constraints

We again wish to extremise $F[y]$, subject to constraints

$$B[y] = \int_{\alpha}^{\beta} g(x, y, y') dx = K,$$

where K is constant. The idea is to use Lagrange multipliers on a new functional

$$\phi[y, \lambda] = F[y] - \lambda G[y].$$

Replacing f in the Euler-Lagrange equation with $f - \lambda g$ gives

$$\frac{d}{dx} \left(\frac{\partial}{\partial y'} (f - \lambda g) \right) = \frac{\partial}{\partial y} (f - \lambda g). \quad (3.1)$$

Example 3.1 (Dido problem/Isoperimetric problem). What simple and closed planar curve γ of fixed length maximizes the enclosed area?

Here, a curve is **simple** if it is contractible to a point, and non-self-intersecting. We also assume the curve is convex, otherwise we can form a curve with bigger area by taking the convex hull.

Assuming convexity, we can pick coordinate axes such that there exists a leftmost and rightmost points on the curve, with x -coordinates α and β . Then given $x \in (\alpha, \beta)$, there exists (y_1, y_2) on the curve such that $y_1(x) < y_2(x)$. Then y_1, y_2 are functions on $[\alpha, \beta]$. The area is then

$$A[y] = \int_{\alpha}^{\beta} (y_2(x) - y_1(x)) dx = \oint_{\gamma} y(x) dx.$$

The constraint that we have a constant length is given by

$$L[y] = \oint_{\gamma} \sqrt{1 + (y')^2} dx.$$

Since γ has no boundary, we do not need to worry about boundary conditions. From (3.1), we apply Euler-Lagrange to

$$h = f - \lambda g = y - \lambda \sqrt{1 + (y')^2}.$$

Since $\partial h / \partial x = 0$, from (2.2), we get that

$$h - y' \frac{\partial h}{\partial y'} = y - \lambda \sqrt{1 + (y')^2} + y' \lambda \frac{y'}{\sqrt{1 + (y')^2}} = y - \frac{\lambda}{\sqrt{1 + (y')^2}} = K,$$

$$\implies (y')^2 = \frac{\lambda^2}{(y - K)^2} - 1.$$

The solutions to this are simply circles of radius λ , given by

$$(x - x_0)^2 + (y - y_0)^2 = \lambda^2.$$

In this case, since the circumference is $2\pi\lambda = L$, the value of the Lagrange multiplier is $\lambda = L/2\pi$.

3.2 Sturm-Liouville Problem

Consider a function $\rho(x) > 0$ for $x \in [\alpha, \beta]$, and $\sigma(x)$ a function of x . Define

$$F[y] = \int_{\alpha}^{\beta} [\rho \cdot (y')^2 + \sigma \cdot y^2] dx, \quad G[y] = \int_{\alpha}^{\beta} y^2 dx.$$

The **Sturm-Liouville problem** asks to minimize F subject to $G = 1$.

We begin by defining the functional $\phi[y, \lambda] = F[y] - \lambda(G[y] - 1)$. The corresponding function h is

$$h = \rho(y')^2 + \sigma y^2 - \lambda \left(y^2 - \frac{1}{\beta - \alpha} \right).$$

Then the partials are

$$\frac{\partial h}{\partial y'} = 2\rho y', \quad \frac{\partial h}{\partial y} = 2\sigma y - 2\lambda y.$$

Therefore the Euler-Lagrange equation (2) gives

$$-\frac{d}{dx}(\rho y') + \sigma y = \lambda y. \tag{3.2}$$

The left-hand side is a linear operator acting on y , therefore we can write this as

$$\mathcal{L}(y) = \lambda y.$$

Here \mathcal{L} is called the **Sturm-Liouville operator**.

If $\rho = 1$, the equation becomes

$$-y'' + \sigma y = \lambda y.$$

This is the Schrödinger equation.

If $\sigma(x) > 0$, and $y(\alpha) = y(\beta) = 0$, then the positive minimum of F is the smallest eigenvalue of (3.2).

Proof: Take equation (3.2) and multiply by y :

$$-\frac{d}{dx}(\rho y')y + \sigma y^2 = \lambda y^2.$$

Integrating by parts gives

$$F[y] - \underbrace{[yy'\rho]_\beta}_0 = \lambda \underbrace{G[y]}_1.$$

Hence the smallest eigenvalue λ is the minimum of F .

3.3 Several dependent variables

Let $\mathbf{y}(x) = (y_1(x), y_2(x), \dots, y_n(x))$ be a vector of variables. Then we can define a functional

$$F[\mathbf{y}] = \int_\alpha^\beta f(x, y_1, \dots, y_n, y'_1, \dots, y'_n) dx.$$

Assume that there exists an extremum \mathbf{y} . We perturb by

$$y_i \rightarrow y_i(x) + \eta_i(x),$$

where $\eta(\alpha) = \eta(\beta) = 0$. The same derivation as the Euler-Lagrange equation gives

$$F[\mathbf{y} + \varepsilon \boldsymbol{\eta}] - F[\boldsymbol{\eta}] = \varepsilon \int_\alpha^\beta \sum_{i=1}^n \eta_i \left[\frac{d}{dx} \left(\frac{\partial f}{\partial y_i} \right) - \frac{\partial f}{\partial y_i} \right] dx + \varepsilon \sum_{i=1}^n \left[\frac{\partial f}{\partial y'_i} \eta_i \right]_\alpha^\beta + \mathcal{O}(\varepsilon^2).$$

Using well-chosen η_i and the fundamental theorem of variational calculus, we get

$$\frac{d}{dx} \left(\frac{\partial f}{\partial y_i} \right) = \frac{\partial f}{\partial y_i}, \quad (3.3)$$

for all i . This is a system of second order ODEs.

Similar to the Euler-Lagrange equations, we can take first integrals under certain conditions:

- If $\partial f / \partial y_j = 0$ for some $1 \leq j \leq n$, then

$$\frac{\partial f}{\partial y_j} = \text{constant}.$$

- If $\partial f / \partial x = 0$, then

$$f - \sum_{i=1}^n y_i' \frac{\partial f}{\partial y_i} = \text{constant}.$$

Example 3.2 (Geodesics on surfaces). Take $\Sigma \subset \mathbb{R}^3$, a surface given by $g(x, y, z) = 0$. Then a geodesic is the shortest path γ on the surface between $A, B \in \Sigma$.

We parametrise γ by $\mathbf{x}(0) = A$, $\mathbf{x}(1) = B$. Then we can take a functional

$$\phi[\mathbf{x}, \lambda] = \int_0^1 \left[\sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2} - \lambda(t)g(x, y, z) \right] dt = \int_0^1 h(t, x, y, z, \lambda, \dot{x}, \dot{y}, \dot{z}) dt.$$

Note that the Lagrange multiplier λ is now a function of t , as we require the curve to lie on Σ .

Taking the Euler-Lagrange equations with h gives:

$$\frac{d}{dt} \frac{\partial h}{\partial \dot{\lambda}} - \frac{\partial h}{\partial \lambda} = 0 \implies g(x, y, z) = 0,$$

$$\frac{d}{dt} \left(\frac{\dot{x}}{\sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}} \right) + \lambda \frac{\partial g}{\partial x} = 0,$$

and similarly for y and z . Alternatively, we can solve the constraint $g = 0$, avoiding Lagrange multipliers.

3.4 Several independent variables

In general, give $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^m$, we can define

$$F[\phi] = \int_D f(x_1, x_2, \dots, x_n, \phi, \phi_{x_1}, \phi_{x_2}, \dots, \phi_{x_n}) dx_1 dx_2 \dots dx_n,$$

where we evaluate the integral over a domain $D \subset \mathbb{R}^n$.

Assuming we have an extremum ϕ , consider perturbations

$$\phi \rightarrow \phi(x_1, \dots, x_n) + \varepsilon \eta(x_1, \dots, x_n),$$

where we require $\eta = 0$ on ∂D . Then

$$\begin{aligned} F[\phi + \varepsilon \eta] - F[\phi] &= \varepsilon \int_D \left[\eta \frac{\partial f}{\partial \phi} + \sum_{i=1}^n \eta_{x_i} \frac{\partial f}{\partial \phi_{x_i}} \right] dx_1 \dots dx_n + \mathcal{O}(\varepsilon^2) \\ &= \varepsilon \int_D \left[\eta \frac{\partial f}{\partial \phi} + \nabla \cdot \left(\eta \frac{\partial f}{\partial \phi_{\mathbf{x}}} \right) - \eta \nabla \cdot \frac{\partial f}{\partial \phi_{\mathbf{x}}} \right] dx_1 \dots dx_n + \mathcal{O}(\varepsilon^2), \end{aligned}$$

where

$$\frac{\partial f}{\partial \phi_{\mathbf{x}}} = \left(\frac{\partial f}{\partial \phi_{x_1}}, \frac{\partial f}{\partial \phi_{x_2}}, \dots, \frac{\partial f}{\partial \phi_{x_n}} \right).$$

Then, applying divergence theorem on $\eta \partial f / \partial \phi_{\mathbf{x}}$,

$$\int_D \nabla \cdot \left(\eta \frac{\partial f}{\partial \phi_{\mathbf{x}}} \right) dx_1 \dots dx_n = \int_{\partial D} \eta \frac{\partial f}{\partial \phi_{\mathbf{x}}} \cdot d\mathbf{S} = 0.$$

Therefore, we have

$$F[\phi + \varepsilon \eta] - F[\phi] = \varepsilon \int_D \eta \left(\frac{\partial f}{\partial \phi} - \nabla \cdot \frac{\partial f}{\partial \phi_{\mathbf{x}}} \right) dx_1 \dots dx_n + \mathcal{O}(\varepsilon^2).$$

Thus, for several independent variables, the Euler-Lagrange equation becomes a single second order PDE for one function ϕ :

$$\frac{\partial f}{\partial \phi} - \sum_{i=1}^n \frac{\partial}{\partial x_i} \left(\frac{\partial f}{\partial \phi_{x_i}} \right) = 0. \quad (3.4)$$

Example 3.3. Consider $n = 2$. We wish to extremize the potential energy function

$$F[\phi] = \int_D \frac{1}{2}(\phi_x^2 + \phi_y^2) dx dy.$$

Since the partial derivatives of f are

$$\frac{\partial f}{\partial \phi} = 0, \quad \frac{\partial f}{\partial \phi_x} = \phi_x, \quad \frac{\partial f}{\partial \phi_y} = \phi_y,$$

then equation (3.4) gives

$$\frac{\partial}{\partial x} \phi_x + \frac{\partial}{\partial y} \phi_y = 0,$$

or $\phi_{xx} + \phi_{yy} = 0$. This is simply the Laplace equation. Therefore harmonic functions extremize potential functions.

Example 3.4 (Minimal surfaces). We wish to minimise the area of $\Sigma \subset \mathbb{R}^3$, subject to boundary conditions

$$\Sigma = \{\mathbf{x} \in \mathbb{R}^3 \mid K(\mathbf{x}) = 0\}.$$

By the implicit function theorem, we can solve $K = 0$ to give $z = \phi(x, y)$. Then,

$$\begin{aligned} ds^2 &= dx^2 + dy^2 + dz^2, \quad dz = \phi_x dx + \phi_y dy, \\ \implies ds^2 &= (1 + \phi_x^2) dx^2 + 2\phi_x \phi_y dx dy + (1 + \phi_y^2) dy^2. \end{aligned}$$

This is called the **Riemannian metric**. Thus we can write

$$ds^2 = \begin{pmatrix} dx & dy \end{pmatrix} \begin{pmatrix} 1 + \phi_x^2 & \phi_x \phi_y \\ \phi_x \phi_y & 1 + \phi_y^2 \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} = \sum_{i,j=1}^2 g_{ij}(x, y) dx_i dx_j.$$

Then the area element on Σ is $\sqrt{\det g} dx dy$, and the area functional is

$$A[\phi] = \int_{\Sigma} \sqrt{1 + \phi_x^2 + \phi_y^2} dx dy.$$

Applying (3.4), we get

$$\frac{\partial}{\partial x} \left(\frac{\phi_x}{\sqrt{1 + \phi_x^2 + \phi_y^2}} \right) + \frac{\partial}{\partial y} \left(\frac{\phi_y}{\sqrt{1 + \phi_x^2 + \phi_y^2}} \right) = 0.$$

Expanding, this becomes

$$(1 + \phi_y^2)\phi_{xx} + (1 + \phi_x^2)\phi_{yy} - 2\phi_x\phi_y\phi_{xy} = 0. \quad (3.5)$$

This is known as the **minimal surface equation**.

If we assume circular symmetry, then $z = \phi(r)$, where $r = \sqrt{x^2 + y^2}$. Then writing $z = z(r)$, the partial derivatives becomes

$$\phi_x = \frac{dz}{dr} \frac{\partial r}{\partial x} = z' \frac{x}{r}, \quad \phi_y = z' \frac{y}{r}.$$

Thus equation (3.5) becomes

$$rz'' + z' + (z')^3 = 0.$$

Substituting $z' = w$, we get

$$\frac{1}{2}r \frac{d}{dr}(w^2) + w^4 + w^2 = 0.$$

We can solve this to get

$$r = r_0 \cosh \left(\frac{z - z_0}{r_0} \right).$$

This gives a **catenoid**, which is the smallest surface of revolution. The boundary of a catenoid is two circles of the same radius. Assuming $r_0 = 0$, if at $z = R$, we have $r = L$, then we get

$$\frac{L}{r_0} \frac{R}{L} = \cosh \left(\frac{L}{r_0} \right).$$

This gives 2 minimal surfaces for large R . The large value of R gives a **stable** catenoid, while the smaller value is **unstable**. This refers to their behaviour as the bounding circles are drawn apart.

3.5 Higher derivatives

Consider a functional

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

To solve, we proceed as we did in section (2). Assume that y extremizing $F[y]$ exists. Then perturb $y \rightarrow y + \varepsilon\eta$, where $\eta = \eta' = \dots = \eta^{(n-1)} = 0$ at α, β . This gives

$$F[y + \varepsilon\eta] - F[y] = \varepsilon \int_{\alpha}^{\beta} \left(\frac{\partial f}{\partial y} \eta + \frac{\partial f}{\partial y'} \eta' + \dots + \frac{\partial f}{\partial y^{(n)}} \eta^{(n)} \right) dx + \mathcal{O}(\varepsilon^2).$$

We can integrate by parts multiple times and use lemma (2.1) to get

$$0 = \frac{\partial f}{\partial y} - \frac{d}{dx} \left(\frac{\partial f}{\partial y'} \right) + \frac{d^2}{dx^2} \left(\frac{\partial f}{\partial y''} \right) + \dots + (-1)^n \frac{d^n}{dx^n} \left(\frac{\partial f}{\partial y^{(n)}} \right). \quad (3.6)$$

Example 3.5. Take the case $n = 2$, and assume $\partial f / \partial y = 0$. Then equation (3.6) gives

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

Thus if we wish to extremize

$$F[y] = \int_0^1 (y'')^2 dx,$$

where $y(0) = y'(0) = y(1) = 0$, and $y'(1) = 1$, then the above gives

$$\frac{d}{dx}(2y'') = \text{constant} \implies y''' = k.$$

If we impose boundary constraints, this leaves us with $y = x^3 - x^2$. We can prove this is an absolute minimum as well: let $y_0 = x^3 - x^2$, and perturb $y_0 \rightarrow y_0 + \eta$, where $\eta(0) = \eta'(0) = \eta(1) = \eta'(1) = 0$. Then

$$\begin{aligned} F[y_0 + \eta] - F[y_0] &= \int_0^1 (\eta'')^2 dx + 2 \int_0^1 (y_0'' \eta'') dx > 4 \int_0^1 (3x - 1) \eta'' dx \\ &= 4 \left([-\eta']_0^1 + \int_0^1 \left(\frac{d}{dx} (3x \eta') - 3\eta' \right) dx \right) \\ &= 4 \left([3x\eta']_0^1 - [3\eta]_0^1 \right) = 0. \end{aligned}$$

4 Least action principle and Noether's theorem

Consider a particle in \mathbb{R}^3 , with kinetic energy T and potential energy V . Then we can define

$$L(\mathbf{x}, \dot{\mathbf{x}}, t) = T - V, \quad (4.1)$$

the **Lagrangian**. Here L is a function of three dependent variables, $\mathbf{x} = (x_1, x_2, x_3)$, and one independent variable, t . We can define the **action functional**

$$S[\mathbf{x}] = \int_{t_1}^{t_2} L dt. \quad (4.2)$$

Using these, we can state Hamilton's principle.

Proposition 4.1 (Hamilton's principle/Least action principle). *The motion of a particle is such that $S[\mathbf{x}]$ is stationary, that is, L satisfies the Euler-Lagrange equations.*

Example 4.1. If we take

$$T = \frac{1}{2}m|\dot{\mathbf{x}}|^2, \quad V = V(\mathbf{x}),$$

then using Euler-Lagrange gives

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}_i} \right) = \frac{\partial L}{\partial x_i} \implies m\ddot{x}_i = -\frac{\partial V}{\partial x_i}.$$

This can be rewritten as

$$m\ddot{\mathbf{x}} = -\nabla V,$$

which is Newton's second law.

Example 4.2 (Central force in 2 dimensions). Consider the action

$$L = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) - V(r).$$

The Euler-Lagrange equations gives

$$\begin{aligned} \frac{d}{dt} \frac{\partial L}{\partial \dot{r}} - \frac{\partial L}{\partial r} &= 0, \\ \frac{d}{dt} \frac{\partial L}{\partial \dot{\theta}} - \frac{\partial L}{\partial \theta} &= 0. \end{aligned}$$

Since the action doesn't depend on θ , we get that

$$\frac{\partial L}{\partial \dot{\theta}} = mr^2 \dot{\theta} = \text{constant}.$$

This shows that angular momentum is conserved. Moreover, since L is independent of t , we get

$$\begin{aligned} \dot{r} \frac{\partial L}{\partial \dot{r}} + \dot{\theta} \frac{\partial L}{\partial \dot{\theta}} - L &= \text{constant}, \\ \implies \dot{r} m \dot{r} + \dot{\theta} m r^2 \dot{\theta} - L &= \frac{1}{2} m \dot{r}^2 + \frac{1}{2} m r^2 \dot{\theta}^2 + V(r) = E. \end{aligned}$$

Therefore energy is conserved as well.

Example 4.3 (Configuration space, and generalised coordinates). In Newtonian dynamics, we think of a system of particles as N independent trajectories in \mathbb{R}^3 .

The Lagrangian approach is to instead view these trajectories as a single particle moving in **configuration space**. This is of dimension $3N$, and can encode the behaviour of all particles.

To find the trajectory q in the configuration space, we must find the Lagrangian $L(q_i, \dot{q}_i, t)$ and solve the Euler-Lagrange equations.

4.1 Noether's theorem

We consider functionals of the form

$$F[\mathbf{y}] = \int_{\alpha}^{\beta} (y_1, \dots, y_n, y'_1, \dots, y'_n, x) dx.$$

Suppose there exists a 1-parameter family of transformations

$$\begin{aligned} y_i(x) &\rightarrow Y_i(x, s), \\ Y_i(x, 0) &= y_i(x). \end{aligned}$$

This is a continuous symmetry of a Lagrangian f , if

$$\frac{d}{ds} f(Y_1(x, s), \dots, Y_n(x, s), Y'_1(x, s), \dots, Y'_n(x, s), x) = 0.$$

Theorem 4.1 (Noether's theorem). *Given a continuous symmetry $Y_i(x, s)$ of f , the quantity*

$$\sum_{i=1}^n \frac{\partial f}{\partial y'_i} \frac{\partial Y_i}{\partial s} \Big|_{s=0} \quad (4.3)$$

is a first integral of the Euler-Lagrange equation with $Y_i(x, 0) = y_i(x)$.

Proof: Summing over i , we get

$$\begin{aligned} 0 &= \frac{\partial f}{\partial s} \Big|_{s=0} = \frac{\partial f}{\partial y_i} \frac{\partial Y_i}{\partial s} \Big|_{s=0} + \frac{\partial f}{\partial y'_i} \frac{\partial Y'_i}{\partial s} \Big|_{s=0} \\ &= \left[\frac{d}{dx} \left(\frac{\partial f}{\partial y'_i} \right) \frac{\partial Y_i}{\partial s} + \frac{\partial f}{\partial y'_i} \frac{d}{dx} \frac{\partial Y_i}{\partial s} \right] \Big|_{s=0} \\ &= \frac{d}{dx} \left[\frac{\partial f}{\partial y'_i} \frac{\partial Y_i}{\partial s} \right] \Big|_{s=0}. \end{aligned}$$

Example 4.4. Consider

$$f = \frac{1}{2}(y')^2 + \frac{1}{2}(z')^2 - V(y - z).$$

Taking $Y = y + s$, $Z = z + s$, then this is a symmetry of f . Hence, from (4.3),

$$\left(\frac{\partial f}{\partial y'} \frac{\partial Y}{\partial s} + \frac{\partial f}{\partial z'} \frac{\partial Z}{\partial s} \right) \Big|_{s=0} = y' + z'$$

is conserved. Hence we have conserved momentum in the $y + z$ direction.

Example 4.5 (Central forces again). Take the action

$$L = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) - V(r).$$

Taking $\Theta = \theta + s$, $R = r$, then this is a symmetry, so (4.3) gives

$$\left(\frac{\partial L}{\partial \dot{\theta}} \frac{\partial \Theta}{\partial s} + \frac{\partial L}{\partial \dot{r}} \frac{\partial R}{\partial s} \right) = mr^2\dot{\theta}$$

is conserved. So due to the rotational invariance of L , angular momentum is conserved.

5 Convex functions

We will go back to calculus in \mathbb{R}^n , and look at a class of functions for which it is easy to classify stationary points.

Definition 5.1 (Convexity). A set $S \subset \mathbb{R}^n$ is **convex** if for all $x, y \in S$, $(1-t)x + ty \in S$, for all $0 \leq t \leq 1$.

A graph of a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a surface $\{z - f(\mathbf{x}) = 0\}$ in \mathbb{R}^{n+1} . A chord of f is a line segment in \mathbb{R}^{n+1} joining two points on the graph.

A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex if

- (i) the domain of f is a convex set,
- (ii) $f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$ for all $x, y \in S$, $0 \leq t \leq 1$.

Hence f is convex if the graph of f lies below, or on its chords.

Remark.

- (i) f is concave by replacing \leq by \geq in the above definitions.
- (ii) f is convex if and only if $-f$ is concave.
- (iii) f is strictly convex if $f((1-t)x + ty) < (1-t)f(x) + tf(y)$ for $0 < t < 1$.

Functions such as x^2 are convex, and $1/x$ is convex on the positive reals.

5.1 Conditions for convexity

There are 3 different tests for f to be convex.

- a) If f is once differentiable, then f is convex if and only if

$$f(y) \geq f(x) + (y - x) \cdot \nabla f(x).$$

Proof: Assume the above holds. Then applying it twice,

$$\begin{aligned} f(x) &\geq f(z) + (x - z) \cdot \nabla f(z), \\ f(y) &\geq f(z) + (y - z) \cdot \nabla f(z). \end{aligned}$$

Take $z = (1-t)x + ty$. Then adding $(1-t)$ times the first equation, and t times the second gives

$$(1-t)f(x) + tf(y) \geq f(z) = f((1-t)x + ty).$$

Now assume f is convex. For fixed x and y , define

$$h(t) = (1-t)f(x) + tf(y) - f((1-t)x + ty) \geq 0.$$

Then

$$h'(0) = -f(x) + f(y) - (y-x) \cdot \nabla f(x).$$

However since h is positive on $(0,1)$, this gives $h'(0) \geq 0$, as desired.

Corollary 5.1. *If f is convex and has a stationary point, then this point is a global minimum.*

This follows as if $\nabla f(x_0) = 0$, then

$$f(y) \geq f(x_0) + (y-x_0) \cdot \nabla f(x_0) = f(x_0).$$

There are more tests that can be used:

- b) If $(\nabla f(y) - \nabla f(x)) \cdot (y-x) \geq 0$, then f is convex.
- c) Assume f is twice differentiable. Then f is convex if and only if the Hessian has all eigenvalues non-negative (this is known as the second order conditions).

Proof: Assume convexity. Then applying test b) on $y = x + h$, we get

$$h \cdot (\nabla f(x+h) - \nabla f(x)) \geq 0.$$

For small h , since

$$\partial_i f(x+h) = \partial_i f(x) + h_j H_{ij}(x) + \mathcal{O}(|h|^2),$$

by dotting with h , we get

$$h_i h_j H_{ij}(x) + \mathcal{O}(|h|^2) \geq 0.$$

This implies that the Hessian is non-negative definite, implying the eigenvalues are non-negative. The converse is an exercise.

6 Legendre transform

Definition 6.1 (Legendre transform). The Legendre transform of $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is

$$f^*(\mathbf{p}) = \sup_{\mathbf{x}} (\mathbf{p} \cdot \mathbf{x} - f(\mathbf{x})). \quad (6.1)$$

The domain of f^* consists of all vectors \mathbf{p} such that the supremum is finite.

Example 6.1. Consider the Legendre transform of $f(x) = ax^2$, $a > 0$. Then

$$f^*(p) = \sup_x (px - ax^2).$$

The maximum occurs at $p = 2xa$, so $x = p/2a$. Hence

$$f^*(p) = \frac{p^2}{4a}.$$

Computing

$$(f^*)^*(s) = \sup_p \left(sp - \frac{p^2}{4a} \right).$$

This occurs when $p = 2as$, so

$$f^{**}(s) = as^2 = f(s).$$

In fact $f^{**} = f$ for convex functions.

Proposition 6.1. *The domain of f^* is a convex set, and f^* is convex.*

Proof:

$$\begin{aligned} f^*((1-t)p + tq) &= \sup_x [(1-t)p \cdot x + tq \cdot x - f(x)] \\ &= \sup_x [(1-t)[p \cdot x - f(x)] + t[q \cdot x - f(x)]] \\ &\leq (1-t)f^*(p) + tf^*(q). \end{aligned}$$

Hence if $f^*(p)$, $f^*(q)$ exist, then $f^*[(1-t)p + tq]$ does as well, and f^* satisfies the convexity definition.

In practice, if f is convex and differentiable, then $f^*(p)$ is the global minimum over all x . Hence

$$\nabla(p \cdot x - f(x)) = 0 \implies p = \nabla f.$$

If f is strictly convex, then there exists a unique inversion of $p = \nabla f$, so that

$$f^*(p) = p \cdot x(p) - f(x(p)).$$

6.1 Applications to Thermodynamics

Consider a system of many particles. In thermodynamics, we have an order of 10^{27} particles. Using Newtonian or Lagrangian mechanics, this gives 10^{27} ODE's to solve, which simply cannot be done.

Instead, we define a few macroscopic variables, such as pressure P , volume V , temperature T and entropy S .

One such quantity we can define is the internal energy $U(S, V)$. Then we can define the Hermholtz free energy

$$F(T, V) = \min_S (U(S, V) - TS) = -\max_S (TS - U(S, V)) = -U^*(T, V).$$

This is the Legendre transform of U with respect to S , with V held fixed. Then

$$\left. \frac{\partial}{\partial S} (TS - U(S, V)) \right|_{T, V} = 0 \implies T = \left. \frac{\partial U}{\partial S} \right|_V.$$

We can also define the enthalpy as

$$H(S, P) = \min_V (U(S, V) + PV) = -U^*(P, S).$$

Again, this has a minimum when

$$P = - \left. \frac{\partial U}{\partial V} \right|_S.$$

Thus the Legendre transform is a way to swap from dependence on variables (S, V) , to other variables (T, V) and (S, P) .

7 Hamilton's equations

Recall in §(4.1) that the Lagrangian

$$L = T - V = L(\mathbf{q}, \dot{\mathbf{q}}, t)$$

is a function on the configuration space.

The **Hamiltonian** is the Legendre transform of L with respect to $\dot{\mathbf{q}} = \mathbf{v}$, where \mathbf{v} is the solution to

$$p_i = \frac{\partial L}{\partial \dot{q}_i}.$$

Here we assume the convexity of L in \mathbf{v} . We say that \mathbf{p} is the generalised momentum.

Example 7.1. Take

$$T = \frac{1}{2}m|\dot{\mathbf{q}}|^2, \quad V = V(\mathbf{q}).$$

Then we have

$$\mathbf{p} = \frac{\partial L}{\partial \dot{\mathbf{q}}} = m\dot{\mathbf{q}}.$$

Therefore \mathbf{p} is precisely the momentum. Hence the Hamiltonian is

$$H(\mathbf{q}, \mathbf{p}, t) = \mathbf{p} \cdot \frac{\mathbf{p}}{m} - \frac{1}{2}m\frac{|\mathbf{p}|^2}{m^2} - V(\mathbf{q}) = \frac{1}{2m}|\mathbf{p}|^2 + V(\mathbf{q}).$$

Consider taking the total derivative of H . Since

$$H = \mathbf{p} \cdot \dot{\mathbf{q}} - L(\mathbf{q}, \dot{\mathbf{q}}, t),$$

we have

$$\begin{aligned} dH &= \frac{\partial H}{\partial q_i} dq_i + \frac{\partial H}{\partial p_i} dp_i + \frac{\partial H}{\partial t} dt \\ &= p_i d\dot{q}_i + \dot{q}_i dp_i - \underbrace{\frac{\partial L}{\partial q_i}}_{\dot{p}_i \text{ by E-L}} dq_i - \underbrace{\frac{\partial L}{\partial \dot{q}_i}}_{p_i} d\dot{q}_i - \frac{\partial L}{\partial t} dt \\ &= \dot{q}_i dp_i - \dot{p}_i dq_i - \frac{\partial L}{\partial t} dt. \end{aligned}$$

Comparing differentials, we get that

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q_i}, \quad \frac{\partial H}{\partial t} = -\frac{\partial L}{\partial t}. \quad (7.1)$$

This is known as Hamilton's equations. Assuming there is no explicit t -dependence in L , then (7.1) is a system of $2n$ first order ODEs. The solution curves form a trajectory in $2n$ -dimensional phase space.

Remark. Hamilton's equations also arise from extremizing a functional in phase space

$$S[\mathbf{q}, \mathbf{p}] = \int_{t_1}^{t_2} [\dot{q}_i p_i - H(\mathbf{q}, \mathbf{p}, t)] dt.$$

Using Euler-Lagrange,

- Variations with respect to p_i give

$$\frac{\partial f}{\partial p_i} - \frac{d}{dt} \left(\frac{\partial f}{\partial \dot{p}_i} \right) = 0 \implies \dot{q}_i = \frac{\partial H}{\partial p_i}.$$

- Variations with respect to q_i gives

$$\frac{\partial f}{\partial q_i} - \frac{d}{dt} \left(\frac{\partial f}{\partial \dot{q}_i} \right) = 0 \implies \dot{p}_i = -\frac{\partial H}{\partial q_i}.$$

8 The Second Variation

The Euler-Lagrange equation can lead to a minimum, maximum or a saddle point. We will now look at the nature of the stationary points of

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

Assume y is a critical point, and expand $F[y + \varepsilon\eta]$ around a solution to the Euler-Lagrange equation:

$$\begin{aligned} F[y + \varepsilon\eta] - F[y] &= \int_{\alpha}^{\beta} [f(x, y + \varepsilon\eta, y' + \varepsilon\eta') - f] dx \\ &= \varepsilon \int_{\alpha}^{\beta} \eta \left(\frac{\partial f}{\partial y} - \frac{d}{dx} \frac{\partial f}{\partial y'} \right) dx \\ &\quad + \frac{\varepsilon^2}{2} \int_{\alpha}^{\beta} \left[\eta^2 \frac{\partial^2 f}{\partial y^2} + 2\eta\eta' \frac{\partial^2 f}{\partial y \partial y'} + (\eta')^2 \frac{\partial^2 f}{\partial y'^2} \right] dx + \mathcal{O}(\varepsilon^3). \end{aligned}$$

It is the latter term which is the second variation:

$$\begin{aligned} \delta^2 F[y] &= \frac{1}{2} \int_{\alpha}^{\beta} \left[\eta^2 \frac{\partial^2 f}{\partial y^2} + \frac{d}{dx} (\eta^2) \frac{\partial^2 f}{\partial y \partial y'} + (\eta')^2 \frac{\partial^2 f}{\partial y'^2} \right] dx \\ &= \frac{1}{2} \int_{\alpha}^{\beta} [Q\eta^2 + P(\eta')^2] dx, \end{aligned} \tag{8.1}$$

where

$$\begin{aligned} P &= \frac{\partial^2 f}{\partial y'^2}, \\ Q &= \frac{\partial^2 f}{\partial y^2} - \frac{d}{dx} \left(\frac{\partial^2 f}{\partial y \partial y'} \right). \end{aligned}$$

Therefore, if $y_0(x)$ is a solution to the Euler-Lagrange equation (2) and $Q\eta^2 + P(\eta')^2 > 0$, for all η vanishing at α and β , then $y_0(x)$ is a local minimum of $F[y]$.

Example 8.1 (Geodesics on \mathbb{R}^2). Take the function $f = \sqrt{1 + (y')^2}$. Then

$$Q = 0, \quad P = \frac{\partial}{\partial y'} \frac{y'}{\sqrt{1 + (y')^2}} = \frac{1}{[1 + (y')^2]^{3/2}}.$$

Hence if $\eta' = 0$, then $\eta = 0$, so there is no variation, and if $\eta' \neq 0$, then $P(\eta')^2 > 0$. Hence straight lines are local length minimizers.

Proposition 8.1. *If $y_0(x)$ is a local minimum, then*

$$P = \frac{\partial^2 f}{\partial y'^2} \Big|_{y_0(x)} \geq 0. \quad (8.2)$$

This is known as the Legendre condition.

To see why this is true, note that if η' is small, then η can't be too large. However the converse is not true. Indeed, η can be small, but η' arbitrarily large, if it oscillates very rapidly.

Proof: (Sketch) Assume there exists $x_0 \in (\alpha, \beta)$ such that $P(x_0, y_0(x_0), y'_0(x_0)) < 0$. Then taking a small, oscillating function around x_0 , $P(\eta')^2 + Q\eta^2 < 0$.

Remark. Note that the Legendre condition is not sufficient. However $P > 0, Q \geq 0$ is sufficient, as if $\eta' = 0$, then $\eta = 0$, otherwise $P(\eta')^2 > 0$.

8.1 Associated eigenvalue problem

Looking at (8.1),

$$Q\eta^2 + P(\eta')^2 = Q\eta^2 + \frac{d}{dx}(P\eta\eta') - \eta(P\eta')'.$$

Now integrating and dropping boundary terms,

$$\delta^2 F[y_0] = \frac{1}{2} \int_{\alpha}^{\beta} \eta [-(P\eta')' + Q\eta] dx.$$

Here $\mathcal{L}(\eta) = -(P\eta')' + Q\eta$ is a Sturm-Liouville operator. Thus if there exists η such that

$$\mathcal{L}(\eta) = -\omega^2 \eta, \quad \eta(\alpha) = \eta(\beta) = 0, \quad (8.2)$$

then y_0 is not a local minimizer.

Example 8.2. Take

$$F[y] = \int_0^{\beta} [(y')^2 - y^2] dx,$$

with $y(0) = y(\beta) = 0$, and $\beta \neq N\pi$. Then by Euler-Lagrange, $y'' + y = 0$, so $y_0 = 0$ is a stationary point. Then

$$\delta^2 F[y_0] = \int_0^\beta [(\eta')^2 - \eta^2] dx.$$

Here $P = 2 > 0$. We wish to extremize (8.2):

$$\eta'' - \eta = -\omega^2 \eta, \quad \eta(0) = \eta(\beta) = 0.$$

This has solutions

$$\eta = A \sin\left(\frac{\pi x}{\beta}\right), \quad \left(\frac{\pi}{\beta}\right)^2 = 1 - \omega^2.$$

Thus this is possible for $\beta > \pi$. So if $P > 0$, a problem with the positivity of $\delta^2 F$ may arise if the interval is too large.

8.2 The Jacobian condition

Legendre tried to prove that $P > 0$ is sufficient for $y = y_0$ to be a local minimum. From our last condition, this could not have worked, however the idea used was good: Let $\phi = \phi(x)$ be any differentiable function of x on $[\alpha, \beta]$. Then

$$0 = \int_\alpha^\beta (\phi \eta^2)' dx = \int_\alpha^\beta [\phi' \eta^2 + 2\eta \eta' \phi] dx.$$

Adding this to (8.1),

$$\delta^2 F[y] = \frac{1}{2} \int_\alpha^\beta [P(\eta')^2 + 2\eta \eta' \phi + (Q + \phi')\eta^2] dx.$$

Assume that $P > 0$ at this solution y , and complete the square:

$$\delta^2 F[y] = \frac{1}{2} \int_\alpha^\beta \left[P \left(\eta' + \frac{\phi}{P} \eta \right)^2 + \left(Q + \phi' - \frac{\phi^2}{P} \right) \eta^2 \right] dx,$$

which is positive if we can choose ϕ such that

$$\phi^2 = P(Q + \phi'). \tag{8.3}$$

If (8.3) holds then $\delta^2 F > 0$ unless

$$\eta' + \frac{\phi}{P} \eta =$$

on $[\alpha, \beta]$. But $\eta = 0$ at α, β , so $\eta'(\alpha) = 0$, meaning $\eta = 0$ on $[\alpha, \beta]$.

The equation (8.3) is of Ricatti type, so we can transform it into a linear second order equation. Set $\phi = -pu'/u$, where $u \neq 0$ on $[\alpha, \beta]$. Then

$$P \left(\frac{u'}{u} \right)^2 = Q - \left(\frac{(Pu')}{u} \right)' = Q - \frac{(Pu')'}{u} + P \left(\frac{u'}{u} \right)^2,$$

or

$$-(Pu')' + Qu = 0. \quad (8.4)$$

This is the Jacobi accessory condition. Thus we require a solution to (8.4) such that $u \neq 0$ on $[\alpha, \beta]$, which may not exist on a large interval.

Example 8.3. Take

$$F[y] = \frac{1}{2} \int_{\alpha}^{\beta} [(y')^2 - y^2] dx.$$

Then

$$\delta^2 F[y] = \frac{1}{2} \int_{\alpha}^{\beta} [(\eta')^2 - \eta^2] dx.$$

The Jacobi accessory condition is $u'' + u = 0$. The general solution is $u = A \sin x + B \cos x$, so we want to know when this is non-zero on $[\alpha, \beta]$. This occurs if

$$\tan x \neq \frac{B}{A}.$$

It is possible to avoid B/A on an interval smaller than π .

Example 8.4 (Geodesics on a sphere). On a geodesic, the line element is

$$\sqrt{d\theta^2 + \sin^2 \theta d\phi^2} = \sqrt{(\theta')^2 + \sin^2 \theta} d\phi.$$

We found earlier that the critical points are segments of great circles. The second variation at the stationary points are

$$\frac{\partial^2 f}{(\partial \theta')^2} = 1 = P, \quad Q = -1.$$

Hence

$$\delta^2 F[\theta_0 = \pi/2, \eta] = \frac{1}{2} \int_{\phi_1}^{\phi_2} [(\eta')^2 - \eta^2] d\phi.$$

This is positive if $\phi_2 - \phi_1 < \pi$.

Index

- action functional, 21
- angular momentum, 22
- brachistochrone, 11
- brachistochrone problem, 2
- catenoid, 19
- central force, 21
- configuration space, 22
- constraint function, 6
- continuous symmetry, 22
- convex, 24
- convex function, 24
- cycloid, 12
- Dido problem, 13
- energy, 22
- enthalpy, 27
- Euler-Lagrange equation, 8
- Fermat's principle, 3, 12
- first integral, 9
- functional, 2
- functional derivative, 8
- fundamental lemma of variational calculus, 7
- generalised momentum, 28
- geodesics, 2, 9, 10, 16
- Hamilton's equations, 29
- Hamilton's principle, 21
- Hamiltonian, 28
- harmonic function, 5
- Hermholtz free energy, 27
- Hessian matrix, 4
- isoperimetric problem, 13
- Jacobi accessory condition, 33
- Jacobian condition, 33
- Lagrange multipliers, 6, 13
- Lagrangian, 21
- Laplace equation, 18
- least action principle, 21
- Legendre condition, 31
- Legendre transform, 26
- local maximum, 4
- local minimum, 4
- minimal surface equation, 19
- minimal surfaces, 18
- Newton's second law, 21
- Noether's theorem, 23
- principle of least action, 3
- Riemannian metric, 18
- saddle point, 4
- Schrödinger equation, 15
- second order conditions, 25
- second variation, 30
- simple curve, 13
- Snell's law, 12
- stationary point, 4
- Sturm-Liouville operator, 14
- Sturm-Liouville problem, 14
- symmetry, 22
- thermodynamics, 27
- total derivative, 9