Variational Principles

Cambridge University Mathematical Tripos: Part IB

4th May 2024

Contents

1 History and motivation

1.1 The brachistochrone problem

Consider a particle sliding on a wire under the influence of gravity between two fixed points in the plane. What is the shape of the wire that produces the shortest travel time between the end points, given that the particle starts at rest? This problem is known as the brachistochrone problem, an archetypical variational problem. Suppose the end points are labelled A and B, where A is the origin, i.e. $(x_1, y_1) = (0, 0)$, and where B has coordinates (x_2, y_2). Note that $y_2 < 0$ in order that the particle has sufficient energy to reach the destination. The travel time T is given by

$$T = \int \mathrm{d}t = \int_A^B \frac{\mathrm{d}\ell}{v(x, y)}$$

Note that the kinetic energy and the potential energy sum to a constant.

$$\frac{1}{2}mv^2 + mgy = mgy_1 = 0 \implies v = \sqrt{2g}\sqrt{-y}$$

So we must find the function y that minimises

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

subject to $y_0 = 0$, $y(x_2) = y_2$. This problem's solution will be explored in a later lecture.

1.2 Geodesics

A geodesic is the shortest path γ between two points on a surface Σ, assuming such a path exists. Initially, let $\Sigma = \mathbb{R}^2$. On this plane, the Pythagorean theorem for measuring distances holds. Using a Cartesian coordinate system, we can say that a point A has coordinates (x_1, y_1), and a point B has coordinates (x_2, y_2). The distance from A to B along any path γ can be computed using a line integral.

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

In this case, we have defined y as a function of x, and we seek to minimise D by varying the path γ on which we are moving.

1.3 Calculus of variations

A variational problem involves minimising an object of the form

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

subject to fixed values of y at the end points. We call such an F a *functional*; it is a function on the space of functions. Calculus applied to functionals is called the calculus of variations; we would like to find minima and maxima of functionals. In order to talk about functionals rigorously, we must define first the space of functions we are operating on; analogously to how we must define the domain of a function we are analysing when dealing with real or complex analysis. We write $C(\mathbb{R})$ for the space of continuous functions on $\mathbb{R}$, and $C^k(\mathbb{R})$ for the space of functions with continuous kth derivatives on $\mathbb{R}$. Sometimes, the notation $C^k_{(\alpha,\beta)}(\mathbb{R})$ is used to denote $C^k(\mathbb{R})$ such that $f(\alpha)$ and $f(\beta)$ are fixed, typically fixed to zero.

1.4 Variational principles

We can now define what variational principles are: they are such principles where laws follow from finding the minima or maxima of functionals. An introductory example is Fermat's principle, which states that light that travels between two points takes the path which requires the least travel time. There is also the principle of least action. Consider a particle moving under some potential $V(\mathbf{x})$, and let $T = \frac{1}{2}m|\dot{\mathbf{x}}|^2$ be its kinetic energy. We can define

$$S[\gamma] = \int_{t_1}^{t_2} (T - V)\,\mathrm{d}t$$

where γ represents the path along which the particle travels. The left hand side $S[\gamma]$ is called the *action*, and the principle of least action states that the action is minimised along paths of motion. Then, Newton's laws of motion should follow from this principle by minimising action.

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

2.1 Introduction

Let $f \in C^2(\mathbb{R}^n)$, so $f : \mathbb{R}^n \to \mathbb{R}$ with all continuous second partial derivatives. We say that the point $\mathbf{a} \in \mathbb{R}^n$ is stationary if

$$\nabla f(\mathbf{a}) = \mathbf{0}$$

Consider a Taylor series expansion near a stationary point.

$$f(\mathbf{x}) = f(\mathbf{a}) + \frac{1}{2}(x_i - a_i)(x_j - a_j)\, \partial^2_{ij} f\Big|_{\mathbf{a}} + O(\|\mathbf{x} - \mathbf{a}\|^2)$$

The Hessian matrix is defined as $H_{ij} = \partial_i \partial_j f = H_{ji}$, where $\partial_i \equiv \frac{\partial}{\partial x_i}$. For convenience, we will shift the origin to let $\mathbf{a} = \mathbf{0}$. The Hessian, evaluated at $\mathbf{0}$, written $H(\mathbf{0})$, is a real symmetric matrix and hence can be diagonalised using an orthogonal transformation.

$$H' = R^\intercal H(\mathbf{0}) R = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

Then

$$f(\mathbf{x}') - f(\mathbf{0}) = \frac{1}{2}\sum \lambda_i (x_i')^2 + O(\|\mathbf{x}\|^2)$$

We can characterise the stationary point using the eigenvalues of the Hessian.

(i) If all $\lambda_i > 0$, then $f(\mathbf{x}') > f(\mathbf{0})$ so $f(\mathbf{x}')$ is a local minimum.

(ii) If all $\lambda_i < 0$, then $f(\mathbf{x}') < f(\mathbf{0})$ so $f(\mathbf{x}')$ is a local maxmimum.

(iii) If the eigenvalues have mixed signs, this is a saddle point. $f(\mathbf{x}')$ increases in some directions, but decreases in other directions.

(iv) If some eigenvalues are zero, we must consider higher-order terms of the Taylor expansion.

When $n = 2$, this is a special case. We can compute properties of the eigenvalues using the trace and determinant of the matrix.

$$\det H = \lambda_1\lambda_2;\ \operatorname{tr} H = \lambda_1 + \lambda_2$$

(i) If $\det H > 0,\ \operatorname{tr} H > 0$ then we have a local minimum.

(ii) If $\det H > 0,\ \operatorname{tr} H < 0$ then we have a local maximum.

(iii) If $\det H < 0$ then we have a saddle point.

(iv) If $\det H = 0$ we need to consider higher-order terms.

Note that if $f : D \to \mathbb{R}$ where $D \subset \mathbb{R}^n$, it is possible that we have a local maximum which is not the global maximum, if such a global maximum actually lies on the boundary and is not a stationary point.

Now, let us suppose that f is harmonic, i.e. $\nabla^2 f(\mathbf{x}) = 0$ on $D \subset \mathbb{R}^2$. Hence, $\operatorname{tr} H = 0$ which implies that if there exists a turning point it is a saddle point. The minimum or maximum of a harmonic function must therefore occur on the boundary.

Example. Let

$$f(x, y) = x^3 + y^3 - 3xy$$

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 3x^2 - 3y \\ 3y^2 - 3x \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \implies \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ or } \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

The Hessian is

$$H = \begin{pmatrix} 6x & -3 \\ -3 & 6y \end{pmatrix} \implies H(\mathbf{0}) = \begin{pmatrix} 0 & -3 \\ -3 & 0 \end{pmatrix};\quad H\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 6 & -3 \\ -3 & 6 \end{pmatrix}$$

The determinant is negative at zero, giving us a saddle point. At the other point, the determinant is positive and the trace is positive, giving a local minimum.

2.2 Constraints and Lagrange multipliers

Example. Find the circle centered at $(0, 0)$ with smallest radius that intersects the parabola $y = x^2 - 1$. There are essentially two approaches.

- First, we consider the 'direct' method. We solve the constraints directly, which in this case means solving the equations

  $$f = x^2 + y^2$$
  $$y = x^2 - 1$$

  for minimal f. This gives

  $$f = x^2 + (x^2 - 1)^2 = x^4 - x^2 + 1$$

  Then by setting $\partial_x f = 0$ we have

  $$4x^3 - 2x = 0 \implies x \in \left\{0, \frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right\}$$

  which gives

  $$x = \frac{\pm 1}{\sqrt{2}} \implies y = \frac{-1}{2};\ r = \frac{\sqrt{3}}{2}$$

The other solution for x yields a larger radius. This method works fine for simple problems like this where the constraints are solvable. Therefore, we present an alternative method that works in the more general case.

- This method uses 'Lagrange multipliers'. We define a new function

$$h(x, y, \lambda) = f(x, y) - \lambda g(x, y)$$

where $g(x, y)$ is defined such that $g = 0$ is the constraint. λ is called the Lagrange multiplier. In this example,

$$h(x, y, \lambda) = x^2 + y^2 - \lambda(y - x^2 + 1)$$

We now extremise h over all free variables without constraints.

$$\nabla h = \begin{pmatrix} \partial h / \partial x \\ \partial h / \partial y \\ \partial h / \partial \lambda \end{pmatrix} = \begin{pmatrix} 2x + 2\lambda x \\ 2y - \lambda \\ y - x^2 + 1 \end{pmatrix}$$

Solving $\nabla h = 0$, we have

$$2x + 4xy = 0 \implies x = 0 \text{ or } y = \frac{-1}{2}$$

and the same results follow as before by substitution.

2.3 Geometric justification of Lagrange multipliers

Consider a curve given by $g = 0$. At each point on this curve, there is a normal to the curve of gradient ∇g. In particular, ∇g is perpendicular to $g = 0$. The function f has gradient perpendicular to the function $f = c$ for some constant c. So at the extremum, $\nabla f \propto \nabla g$, so $\nabla f - \lambda g = 0$ for some λ. This guides the creation of the new function h, for which we can optimise without constraints. This same reasoning generalises to functions in higher dimensions and with multiple constraints.

3 Euler–Lagrange equation

3.1 Fundamental lemma of calculus of variations

Consider again the functional

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

where f is given, and $f(\alpha, \cdot, \cdot)$ and $f(\beta, \cdot, \cdot)$ are fixed. Consider a small perturbation

$$y \mapsto y + \varepsilon\eta(x); \quad \eta(\alpha) = \eta(\beta) = 0$$

In order to compute the functional for this new function, we first need an additional lemma.

Lemma (Fundamental lemma of calculus of variations)**.** If $g \colon [\alpha, \beta] \to \mathbb{R}$ is continuous on this interval, and is such that

$$\forall \eta \text{ continuous}, \eta(\alpha) = \eta(\beta) = 0, \quad \int_\alpha^\beta g(x)\eta(x)\,\mathrm{d}x = 0$$

Then

$$\forall x \in (\alpha, \beta),\ \ g(x) \equiv 0$$

Proof. Suppose that there exists a value $\overline{x} \in (\alpha, \beta)$ such that $g(\overline{x}) \neq 0$. Without loss of generality suppose that this value is positive. Then, by continuity, there exists a sub-interval $[x_1, x_2] \subset (\alpha, \beta)$ where $g(x) > c$ for some positive real c in this sub-interval. So we will construct an η such that $\eta > 0$ in $[x_1, x_2]$ and $\eta = 0$ outside this interval, for example

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

Then the integrand is non-negative everywhere, and is lower bounded by a positive number:

$$\int_\alpha^\beta g(x)\eta(x) > c \int_{x_1}^{x_2} (x - x_1)(x_2 - x)\,\mathrm{d}x > 0$$

So this leads to a contradiction. □

Remark. We call such an η function a 'bump function'. In general it is possible to construct a C^k bump function, e.g.

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

3.2 Euler–Lagrange equation

Now, we can evaluate the original functional. Using a Taylor expansion,

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

For an extremum,

$$\left.\frac{\mathrm{d}F}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} = 0$$

So we want the first order term to vanish, so

$$\varepsilon \int_\alpha^\beta \left(\frac{\partial f}{\partial y}\eta + \frac{\partial f}{\partial y'}\eta' \right) \mathrm{d}x = 0$$

Integrating by parts, we have

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

We can apply the lemma above, showing that a necessary condition for the optimum is

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

This is the Euler–Lagrange equation.

Remark. Note that

- This can be seen as a second-order differential equation for $y(x)$ with boundary conditions at α and β.
- The left hand side of the Euler–Lagrange equation is called a 'functional derivative' of y, and is written

  $$\frac{\delta F[y]}{\delta y(x)}$$

  Sometimes, the notation

  $$\delta y = \varepsilon\eta(x)$$

  is used, but is not used in this course. Note that in this notation,

  $$F[y + \delta y] = F[y] + \delta F[y]; \quad \delta F[y] = \int_\alpha^\beta \left[\frac{\delta F[y]}{\delta y(x)}\delta y(x)\right]\mathrm{d}x$$

- Other boundary conditions, such as $\left.\frac{\partial f}{\partial y'}\right|_{\alpha,\beta}$ can be used.
- Note that when computing the derivatives, we regard x, y, y' as independent;

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

  We can also compute a total derivative, for instance

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

  Note that these give different results. As an example, let $f(x, y, y') = x[(y')^2 - y^2]$. Then

  $$\frac{\partial f}{\partial x} = (y')^2 - y^2; \quad \frac{\partial f}{\partial y} = -2xy; \quad \frac{\partial f}{\partial y'} = 2xy'$$

  Hence

  $$\frac{\mathrm{d}f}{\mathrm{d}x} = (y')^2 - y^2 - 2xyy' + 2xy'y''$$

3.3 First integral of Euler–Lagrange equation (eliminating y)

In some cases, we can integrate the Euler–Lagrange equation to give a first-order ordinary differential equation. Suppose f does not explicitly depend on y. Then

$$\frac{\partial f}{\partial y} = 0 \implies \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right) = 0$$

Hence,

$$\frac{\partial f}{\partial y'} = c; \quad c \in \mathbb{R}$$

Example. Consider geodesics on $\mathbb{R}^2$; we want to find curves on which the length is minimised.

$$F[y] = \int_\alpha^\beta \sqrt{\mathrm{d}x^2 + \mathrm{d}y^2} = \int_\alpha^\beta \underbrace{\sqrt{1 + \frac{\mathrm{d}y}{\mathrm{d}x}^2}}_{f(y')}\,\mathrm{d}x$$

We can apply this 'first integral' form of the Euler–Lagrange equation to get

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

Hence y' is a constant, so let $y' = m$ for $m \in \mathbb{R}$. Hence $y = mx + c$.

3.4 Geodesics on a sphere

Consider the unit sphere $S^2 \subset \mathbb{R}^3$, and two points $A, B \in S^2$ which we wish to connect by a path of minimal length, where the path is constrained to the sphere. We will parametrise the sphere with spherical polar coordinates:

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

where $\theta \in [0, \pi]; \phi \in [0, 2\pi]$. We can calculate the length of a path using the Pythagorean theorem:

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

We will parametrise the path by thinking of ϕ as a function of θ. This gives

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

We wish to extremise the functional F, given by

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

The integrand does not depend on ϕ but only on its derivative; so $\frac{\mathrm{d}f}{\mathrm{d}\phi} = 0$. Using the first integral form of the Euler–Lagrange equation, we have

$$\frac{\partial f}{\partial \phi'} = k$$

Now, we have

$$
\begin{aligned}
\frac{\sin^2\theta\phi'}{\sqrt{1+\sin^2\theta(\phi')^2}} &= k \\
\sin^4\theta(\phi')^2 &= k^2(1+\sin^2\theta(\phi')^2) \\
(\phi')^2 &= \frac{k^2}{\sin^2\theta(\sin^2\theta-k^2)} \\
\frac{\mathrm{d}\phi}{\mathrm{d}\theta} &= \pm\sqrt{\frac{k^2}{\sin^2\theta(\sin^2\theta-k^2)}} \\
\phi &= \pm\int\frac{k\,\mathrm{d}\theta}{\sin\theta\sqrt{\sin^2\theta-k^2}}
\end{aligned}
$$

The two solutions correspond to the two directions in which we can trace the path. We then can arrive at

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

We will be able to see that this corresponds to a great circle; that is, the intersection of a plane through the origin with the sphere. We will show later that geodesics on a sphere are *only* segments of a great circle.

3.5 First integral of Euler–Lagrange equation (eliminating x)

For any $f(x,y,y')$, consider the quantity

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

This is exactly

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\left(f-y'\frac{\partial f}{\partial y'}\right) &= \frac{\partial f}{\partial x}+y'\frac{\partial f}{\partial y}+y''\frac{\partial f}{\partial y'}-y''\frac{\partial f}{\partial y'}-y'\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right) \\
&= \frac{\partial f}{\partial x}+y'\frac{\partial f}{\partial y}-y'\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right) \\
&= y'\underbrace{\left(\frac{\partial f}{\partial y}-\frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial f}{\partial y'}\right)}_{\text{zero by Euler–Lagrange}}+\frac{\partial f}{\partial x} \\
&= \frac{\partial f}{\partial x}
\end{aligned}
$$

So, in the case that f does not depend explicitly on x (that is, $\frac{\partial f}{\partial x}\equiv 0$), then we have another first integral condition from the Euler–Lagrange equation:

$$f-y'\frac{\partial f}{\partial y'} = \text{constant}$$

3.6 Solving the brachistochrone problem

Consider a curve in the plane with a fixed endpoint at the origin and another fixed endpoint at $x = \beta$. We want to find a path such that the time taken for a particle to travel along this curve is minimised. We previously computed that the travel time is given by

$$F[y] = \frac{1}{\sqrt{2g}} \int_0^\beta \frac{\sqrt{1 + (y')^2}}{\sqrt{-y}} \,\mathrm{d}x$$

This does not depend on x, so we can write (ignoring the $\frac{1}{\sqrt{2g}}$ factor)

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

This gives

$$\begin{aligned} \frac{1}{\sqrt{1 + (y')^2}} &= k\sqrt{-y} \\ y' &= \pm \frac{\sqrt{1 + k^2 y^2}}{k\sqrt{-y}} \\ x &= \pm k \int \frac{\sqrt{-y}}{\sqrt{1 + k^2 y}} \,\mathrm{d}y \end{aligned}$$

We will parametrise further:

$$y = \frac{-1}{k^2} \sin^2 \frac{\theta}{2} \implies \mathrm{d}y = \frac{-1}{k^2} \sin \frac{\theta}{2} \cos \frac{\theta}{2}$$

Hence,

$$\begin{aligned} x &= \pm k \int \frac{-1}{k^2} \frac{\sin^2 \frac{\theta}{2} \cos \frac{\theta}{2}}{\sqrt{1 - \sin^2 \frac{\theta}{2}}} \,\mathrm{d}\theta \\ &= \mp \frac{1}{2k^2} \int (1 - \cos\theta) \,\mathrm{d}\theta \\ &= \mp \frac{1}{2k^2} (\theta - \sin\theta) + c \end{aligned}$$

The initial condition at $(0, 0)$ gives

$$\theta_0 = 0 \implies c = 0$$

Taking the positive solution, we have

$$\begin{aligned} x &= \frac{\theta - \sin\theta}{2k^2} \\ y &= \frac{-1}{k^2} \sin^2 \frac{\theta}{2} \end{aligned}$$

This can be shown to be a parametrised equation of a cycloid.

3.7 Fermat's principle

Fermat's principle states that as light travels between two points, it takes the path of least time. Let a ray of light be represented by a path $y(x)$. The speed of light is given by a function $c(x, y)$ since it depends on the material it is in. Then the time taken is

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

In this general form, f depends on x, y, y'. Now, let us assume c depends only on x and not on y. Then we can use a first integral form to get

$$\frac{\partial f}{\partial y'} = \text{constant}$$

This gives

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

Suppose that at α, the light ray's path has an angle θ_1 with the x-axis, and at β the angle is θ_2. Note that $\theta_1 = \arctan y'|_\alpha$ and the corresponding result for β. Then,

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

This is known as Snell's law.

Suppose we have a material in which c increases with x. In such a material, we then have that θ increases with x. In a material in which c decreases as x increases, θ naturally decreases.

Now, suppose we have a slow material with $c = c_S$ and a fast material with $c = c_F$ adjacent to each other. We might like to find the path that light takes in its path between points that cross the material boundary. Snell's law can be used to determine that the ratio between the sine of the angle and the speed of light remains constant along the light ray's path.

4 Extensions to the Euler–Lagrange equation

4.1 Euler–Lagrange equation with constraints

Given a functional $F[y] = \int_\alpha^\beta f(x, y, y')\,\mathrm{d}x$, we would like to extremise F subject to $G[y] = \int_\alpha^\beta g(x, y, y')\,\mathrm{d}x = k$ for some constant k. We can use the method of Lagrange multipliers. Instead of extremising F, we will extremise

$$\Phi[y; \lambda] = F[y] - \lambda G[y]$$

Thus, we replace f in the Euler–Lagrange equation with $f - \lambda g$, giving

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

4.2 Dido's isoparametric problem

Given a fixed perimeter, we wish to find the simple and closed plane curve which maximises the enclosed area. We can restrict ourselves to convex curves. This is because any concave curve can be transformed into a convex curve with greater area and equal perimeter, by reflecting the non-convex region. We will parametrise the curve in $\mathbb{R}^2$ by letting the minimal and maximal values of x be α, β. Then, as we trace out the curve, x monotonically increases from α to β, and then monotonically decreases as we return from β to α. This induces two functions y_1, y_2 on (α, β) where $y_2 > y_1$. The infinitesimal area is given by $\mathrm{d}A = (y_2 - y_1)\,\mathrm{d}x$. Thus, the area functional is given by

$$A[y] = \int_\alpha^\beta (y_2(x) - y_1(x))\,\mathrm{d}x = \oint_C y(x)\,\mathrm{d}y$$

The constraint functional is

$$L[y] = \oint_C \mathrm{d}\ell = \oint_C \sqrt{1 + (y')^2}\,\mathrm{d}x = L$$

where L is the fixed perimeter. Using Lagrange multipliers, we can define

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

Note that we do not need to consider a boundary term in the derivation of the Euler–Lagrange equation, since the curve has no boundary. Using a first integral form of the Euler–Lagrange equation on h, we have

$$k = h - y'\frac{\mathrm{d}h}{\mathrm{d}y'} = y - \lambda\sqrt{1 + (y')^2} + y'\lambda\frac{y'}{\sqrt{1 + (y')^2}} = y - \frac{\lambda}{\sqrt{1 + (y')^2}}$$

for some constant k. Hence,

$$(y')^2 = \frac{\lambda^2}{(y - k)^2} - 1$$

A solution here is the circle of radius λ:

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

Here, $L = 2\pi\lambda$ so we can write the solution in terms of L instead, giving

$$(x - x_0)^2 + (y - y_0)^2 = \frac{L^2}{4\pi^2}$$

4.3 The Sturm–Liouville problem

Let $\rho(x), \sigma(x)$ be defined for $x \in [\alpha, \beta]$, and let $\rho(x) > 0$ on this interval. Consider the functional

$$F[y] = \int_\alpha^\beta \left[\rho(y')^2 + \sigma y^2\right]\mathrm{d}x$$

Let us extremise F subject to the constraint

$$G[y] = \int_\alpha^\beta y^2\,\mathrm{d}x = 1$$

We have

$$\Phi[y;\lambda]F[y] - \lambda(G[y] - 1)$$

This induces the integrand

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

We consider the derivatives for the Euler–Lagrange equation:

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

Hence,

$$-\frac{\mathrm{d}}{\mathrm{d}x}(\rho y') + \sigma y = \lambda y$$

We can write this as $\mathcal{L}(y) = \lambda y$, where the $\mathcal{L}$ is known as the Sturm–Liouville operator. This is essentially an eigenvalue problem, since $\mathcal{L}$ is a linear operator. For example, if $\rho = 1$, this eigenvalue problem is exactly the time-independent Schrödinger equation where σ is the quantum-mechanical potential.

Suppose $\sigma > 0$. Then the functional $F[y]$ is also greater than zero. Then, the positive minimum of F (if it exists) is the lowest eigenvalue.

Proof. Using the result from the Euler–Lagrange equation, we can multiply by y and integrate by parts giving

$$\begin{aligned} -y\frac{\mathrm{d}}{\mathrm{d}x}(\rho y') + \sigma y^2 &= \lambda y^2 \\ F[y] - \underbrace{[yy'\rho]_\alpha^\beta}_{\text{zero}} &= \lambda \underbrace{G[y]}_{\text{one}} \end{aligned}$$

Thus, the lowest eigenvalue is the minimum of $F[y]/G[y]$. □

4.4 Multiple dependent variables

Suppose we have some vector

$$\mathbf{y}(x) = (y_1(x), y_2(x), \dots, y_n(x))$$

Suppose we want to extremise the functional

$$F[\mathbf{y}] = \int_\alpha^\beta f(x, y_1, \dots, y_n, y_1', \dots, y_n')\,\mathrm{d}x$$

If there is some critical point $\mathbf{y}$, we perturb by a small amount $\varepsilon\boldsymbol{\eta} = \varepsilon(\eta_1(x), \dots, \eta_n(x))$, where $\boldsymbol{\eta}(\alpha) = \boldsymbol{\eta}(\beta) = \mathbf{0}$. Following the derivation of the one-dimensional Euler–Lagrange equation, we can deduce that

$$F[\mathbf{y} + \varepsilon\boldsymbol{\eta}] - F[\mathbf{y}] = \int_\alpha^\beta \sum_{i=1}^n \eta_i \left(\frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial f}{\partial y_i'} - \frac{\partial f}{\partial y_i} \right) \mathrm{d}x + \text{boundary term} + O(\varepsilon^2)$$

We can apply the fundamental lemma, choosing η_i in a useful way, we can show that a necessary condition for a critical point is

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial f}{\partial y_i'} - \frac{\partial f}{\partial y_i} = 0$$

for all i. This is a second-order *system* of n ODEs that we can solve. If f does not depend on one of the y_i, then we have a first integral form for this particular equation. In particular, if $\frac{\partial f}{\partial y_j} \equiv 0$ then $\frac{\partial f}{\partial y_j'} = \text{constant}$. If f does not depend on x, then we have $f - \sum_i y_i' \frac{\partial f}{\partial y_i'} = \text{constant}$.

4.5 Geodesics on surfaces

Consider a surface Σ in $\mathbb{R}^3$, given by

$$\Sigma = \{\mathbf{x} : g(\mathbf{x}) = 0\}$$

Consider two points A, B on Σ. What are the geodesics (the shortest paths on the surface) between the two points, if one exists at all? Consider a parametrisation of such a path given by $t \in [0, 1]$ where $A = \mathbf{x}(0), B = \mathbf{x}(1)$. We wish to extremise

$$\Phi[\mathbf{x}, \lambda] = \int_0^1 \left\{ \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2} - \lambda(t) g(\mathbf{x}) \right\} \mathrm{d}t$$

The Lagrange multiplier, a function of t, since we want the entire curve (for all t) to lie on Σ. We substitute the integrand h in the Euler–Lagrange equation. Considering the variation with respect to λ, we have

$$\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial h}{\partial \dot{\lambda}} - \frac{\partial h}{\partial \lambda} = 0$$

But h does not depend on $\dot{\lambda}$, hence $\frac{\partial h}{\partial \lambda} = 0$, giving $g(\mathbf{x}) = 0$ for all $\mathbf{x}$. Considering the variation with respect to x_i, we have

$$\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial h}{\partial \dot{x}_i} - \frac{\partial h}{\partial x_i} = 0$$

Hence

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\dot{x}_i}{\sqrt{\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2}} \right) + \lambda \frac{\partial g}{\partial x_i} = 0$$

We could alternatively solve the constraint $g = 0$, and parametrise the surface according to this solution.

4.6 Multiple independent variables

In the most general case, we may have multiple independent variables in a variational problem. This converts the Euler–Lagrange equation into a partial differential equation. Suppose $\phi : \mathbb{R}^n \to \mathbb{R}^m$. If $n = 3$, for example, we have

$$F[\phi] = \iiint_{\mathcal{D}} f(\underbrace{x, y, z}_{\text{independent}}, \phi, \phi_x, \phi_y, \phi_z) \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z$$

where $\mathcal{D} \subset \mathbb{R}^3$, and $\phi_{x_i} := \partial\phi/\partial x_i$. Suppose there exists some extremum ϕ, and consider a small variation $\phi \mapsto \phi(x, y, z) + \varepsilon\eta(x, y, z)$ where $\eta = 0$ on $\partial\mathcal{D}$. Evaluating the functional on this perturbed ϕ gives

$$F[\phi + \varepsilon\eta] - F[\phi] = \varepsilon \iiint_{\mathcal{D}} \left\{ \eta \frac{\partial f}{\partial \phi} + \eta_x \frac{\partial f}{\partial \phi_x} + \eta_y \frac{\partial f}{\partial \phi_y} + \eta_z \frac{\partial f}{\partial \phi_z} \right\} \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z + O(\varepsilon^2)$$

$$= \varepsilon \iiint_{\mathcal{D}} \left\{ \eta \frac{\partial f}{\partial \phi} + \underbrace{\nabla \cdot \left(\eta \left(\frac{\partial f}{\partial \phi_x}, \frac{\partial f}{\partial \phi_y}, \frac{\partial f}{\partial \phi_z} \right) \right)}_{\text{apply divergence theorem since } \eta \text{ vanishes on } \partial\mathcal{D}} - \eta \nabla \cdot \left(\frac{\partial f}{\partial \phi_x}, \frac{\partial f}{\partial \phi_y}, \frac{\partial f}{\partial \phi_z} \right) \right\} \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z + O(\varepsilon^2)$$

$$= \varepsilon \iiint_{\mathcal{D}} \eta \left\{ \frac{\partial f}{\partial \phi} - \nabla \cdot \left(\frac{\partial f}{\partial \phi_x}, \frac{\partial f}{\partial \phi_y}, \frac{\partial f}{\partial \phi_z} \right) \right\} \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z + O(\varepsilon^2)$$

Now, we can apply the fundamental lemma to give the Euler–Lagrange equation for multiple independent variables.

$$\frac{\partial f}{\partial \phi} - \nabla \cdot \left(\frac{\partial f}{\partial \phi_x}, \frac{\partial f}{\partial \phi_y}, \frac{\partial f}{\partial \phi_z} \right) = 0$$

Or, in suffix notation (with the summation convention),

$$\frac{\partial f}{\partial \phi} - \partial_i \frac{\partial f}{\partial(\partial_i \phi)} = 0$$

This result applies for any n. Note that this is now a partial differential equation for ϕ, instead of an ordinary differential equation.

4.7 Potential energy and the Laplace equation

Consider the functional

$$F[\phi] = \iint_{\mathcal{D} \subset \mathbb{R}^2} \frac{1}{2} [\phi_x^2 + \phi_y^2] \,\mathrm{d}x\,\mathrm{d}y$$

Note that $\frac{\partial f}{\partial \phi} = 0$ and $\frac{\partial f}{\partial \phi_x} = \phi_x$; $\frac{\partial f}{\partial \phi_y} = \phi_y$. The Euler–Lagrange equation becomes

$$\frac{\partial}{\partial x} \phi_x + \frac{\partial}{\partial y} \phi_y = 0 \implies \phi_{xx} + \phi_{yy} = 0$$

This produces the Laplace equation.

4.8 Minimal surfaces

Consider minimising the area of a surface $\Sigma \subset \mathbb{R}^3$, where we want the surface to have two boundaries defined by fixed closed curves. This is sometimes known as Plateau's problem. We will let $\Sigma = \{\mathbf{x} = \mathbb{R}^3 : k(x, y, z) = 0\}$, and assume there exists a parametrisation of Σ given by $z = \phi(x, y)$. The line element is given by

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2$$

We have $\mathrm{d}z = \phi_x \,\mathrm{d}x + \phi_y \,\mathrm{d}y$ hence

$$\mathrm{d}s^2 = (1 + \phi_x^2)\,\mathrm{d}x^2 + (1 + \phi_y^2)\,\mathrm{d}y^2 + 2\phi_x\phi_y \,\mathrm{d}x\,\mathrm{d}y$$

This is a quadratic form in the differentials $\mathrm{d}x, \mathrm{d}y$, known as the first fundamental form (also the Riemannian metric). Alternatively,

$$\mathrm{d}s^2 = g_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j$$

where

$$g = \begin{pmatrix} 1 + \phi_x^2 & \phi_x\phi_y \\ \phi_x\phi_y & 1 + \phi_y^2 \end{pmatrix}$$

From this, we can compute the area element, which is defined as

$$\mathrm{d}A = \sqrt{\det g}\,\mathrm{d}x\,\mathrm{d}y$$

We will extremise the area functional

$$A[\phi] = \int_{\mathcal{D}} \sqrt{1 + \phi_x^2 + \phi_y^2}\,\mathrm{d}x\,\mathrm{d}y$$

Let the integrand be h, and apply the Euler–Lagrange equation.

$$\frac{\partial h}{\partial \phi_x} = \frac{\phi_x}{\sqrt{1 + \phi_x^2 + \phi_y^2}}; \quad \frac{\partial h}{\partial \phi_y} = \frac{\phi_y}{\sqrt{1 + \phi_x^2 + \phi_y^2}}$$

Hence

$$\partial_x\left(\frac{\phi_x}{\sqrt{1 + \phi_x^2 + \phi_y^2}}\right) + \partial_y\left(\frac{\phi_x}{\sqrt{1 + \phi_x^2 + \phi_y^2}}\right) = 0$$

which can be expanded to give

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

This is known as the minimal surface equation. We will solve a special case, where there is circular (cylindrical) symmetry, so $z = \phi(r)$. Since $r = \sqrt{x^2 + y^2}$, we can find that

$$\phi_x = z'\frac{x}{r}; \quad \phi_y = z'\frac{y}{r}$$

and we can analogously compute $\phi_{xx}, \phi_{yy}, \phi_{xy}$. This gives

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

We can integrate this by first setting $z' = w$ and multiplying through by w.

$$\frac{1}{2}r\frac{\mathrm{d}}{\mathrm{d}r}w^2 + w^2 + w^4 = 0$$

Now let $w^2 = u$ to make this a separable equation for u. Solving this, we can find that the solution surface is given by

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

This is known as the *catenoid*. At the maximal and minimal values of z, we have the circular boundaries with radii R. At $z = z_0$, the radius is minimal, and the circle here has radius r_0. Supposing $z_0 = 0$ and that the maximal value of z is L, we have

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

Let $L = 1$ without loss of generality. This essentially chooses a scale for the coordinate system. This gives

$$R = r_0 \cosh\frac{1}{r_0}$$

Plotting R as a function of r_0, there exists a minimum point $r_0 = \mu \approx 0.833$ which gives $R \approx 1.5$. So if $R > 1.5$, there exist two distinct minimal surfaces, one with $r_0 > \mu$ and one with $r_0 < \mu$. The 'tighter' minimal surface (with $r_0 < \mu$) is unstable, but the 'looser' surface is stable (however this cannot be shown from our current understanding of variational principles).

4.9 Higher derivatives

Consider the functional

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

We can find an analogous Euler–Lagrange equation to extremise this functional. Let η be a variation where $\eta^{(k)} = 0$ for $k \in \{1, \dots, n-1\}$ at the endpoints α, β. Now,

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

We can repeatedly integrate each term by parts, integrating the $\eta^{(k)}$ term k times. Many of these terms will vanish due to the boundary conditions we specified for η. This then gives

$$F[y + \varepsilon\eta] - F[y] = \varepsilon \int_\alpha^\beta \left(\frac{\partial f}{\partial y}\eta - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial f}{\partial y'}\eta + \dots + (-1)^n \frac{\mathrm{d}^n}{\mathrm{d}x^n}\frac{\partial f}{\partial y^{(n)}}\eta \right) \mathrm{d}x + O(\varepsilon^2)$$

Applying the fundamental lemma of calculus of variations, we have

$$\frac{\partial f}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial f}{\partial y'} + \dots + (-1)^n \frac{\mathrm{d}^n}{\mathrm{d}x^n}\frac{\partial f}{\partial y^{(n)}} = 0$$

This is the Euler–Lagrange equation in the context of a function with higher derivatives. The alternating signs come from the negative signs produced in the iterated integration by parts.

4.10 First integral for $n = 2$

Suppose $n = 2$. If $\frac{\partial f}{\partial y} = 0$, we have

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial f}{\partial y'} - \frac{\mathrm{d}^2}{\mathrm{d}x^2}\frac{\partial f}{\partial y''} = 0$$

Hence

$$\frac{\partial f}{\partial y'} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial f}{\partial y''} = \text{constant}$$

Example. Extremise the functional

$$F[y] = \int_0^1 (y'')^2\,\mathrm{d}x$$

subject to the conditions

$$y(0) = y'(0) = 0; \quad y(1) = 0; \quad y'(1) = 1$$

Using the above first integral form, we have

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

for some $k \in \mathbb{R}$. Imposing the boundary conditions on this cubic gives

$$y = x^3 - x^2$$

Now, we are going to show that this is an absolute *minimum* of the functional, not just a stationary point. Let $y_0 = x^2 - x^2$. Consider a variation η of y_0, where all relevant endpoints of η are zero. In this case, we are *not* going to assume that η is small; we will simply look at all possible variations.

$$F[y_0 + \eta] - F[y_0] = \underbrace{\int_0^1 (\eta'')^2 \,\mathrm{d}x}_{>0} + 2\int_0^1 y_0''\eta'' \,\mathrm{d}x$$

Substituting for y_0, given that $\eta \not\equiv 0$,

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

Hence y_0 is an absolute minimum of F. This method of showing y_0 is an absolute minimum is easier than calculating second variations, where we know the solution y_0.

4.11 Principle of least action

Consider a particle moving in $\mathbb{R}^3$ with kinetic energy T and potential energy V. We define the *Lagrangian* to be

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

We now define the *action* to be

$$S[\mathbf{x}] = \int_{t_1}^{t_2} L \,\mathrm{d}t$$

We can now formulate the principle of least (or stationary) action: on the path of motion of a particle,

$$\frac{\delta S}{\delta \mathbf{x}} = 0$$

Equivalently, L satisfies the Euler–Lagrange equations:

$$\frac{\partial L}{\partial x_i} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{x}_i} = 0$$

Consider

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

The Euler–Lagrange equations are now

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{x}_i} &= \frac{\partial L}{\partial x_i} \\
m\ddot{x}_i &= -\frac{\partial V}{\partial x_i} \\
\implies m\ddot{\mathbf{x}} &= -\nabla V
\end{aligned}$$

This is exactly Newton's second law, derived from the principle of stationary action.

4.12 Central forces

Example. Consider a central force in the plane. The Lagrangian is

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

The Euler–Lagrange equation gives

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{r}} - \frac{\partial L}{\partial r} &= 0 \\ \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{\theta}} - \frac{\partial L}{\partial \theta} &= 0 \end{aligned}$$

Since $\frac{\partial L}{\partial \theta} = 0$, we have a first integral form:

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

This can be interpreted physically as the law of conservation of angular momentum. Further, we have $\frac{\partial L}{\partial t} = 0$ so we have another first integral:

$$\begin{aligned} \dot{r}\frac{\partial L}{\partial \dot{r}} + \dot{\theta}\frac{\partial L}{\partial \dot{\theta}} - L &= \text{constant} \\ m\dot{r}^2 + mr^2\dot{\theta}^2 - \frac{1}{2}m\dot{r}^2 - \frac{1}{2}mr^2\dot{\theta}^2 + V(r) &= \text{constant} \\ \frac{1}{2}m\left(\dot{r}^2 + r^2\dot{\theta}^2\right) + V(r) &= \text{constant} \end{aligned}$$

The left hand side is the total energy of the system, denoted E. This is the law of conservation of energy.

4.13 Configuration space and generalised coordinates

Example. Consider N particles moving in $\mathbb{R}^3$. Typically we represent each point as a distinct vector in $\mathbb{R}^3$ that changes over time. We can alternatively consider a point in $\mathbb{R}^{3N}$, which contains the information about every point. This is called the configuration space. The Lagrangian in configuration space is

$$L = L(q_i, \dot{q}_i, t)$$

where $\mathbf{q}$ is the combined position vector of all N points, and likewise $\dot{\mathbf{q}}$ is the combined velocity.

5 Noether's theorem

5.1 Statement and proof

Consider a functional

$$F[\mathbf{y}] = \int_\alpha^\beta f(y_i, y_i', x)\,\mathrm{d}x\,; \quad i = 1, \dots, n$$

Suppose there exists a one-parameter family of transformations

$$y_i(x) \mapsto Y_i(x, s); \quad Y_i(x, 0) = y_i(x)$$

This can be thought of as a change of variables parametrised by $s \in \mathbb{R}$, where $s = 0$ implies no change of variables. This family is called a *continuous symmetry* of the Lagrangian f if

$$\frac{\mathrm{d}}{\mathrm{d}s} f(Y_i(x,s), Y_i'(x,s), x) = 0$$

In this course, we only consider continuous symmetries, so they may be abbreviated as just 'symmetries'.

Theorem (Noether's Theorem). Given a continuous symmetry $Y_i(x,s)$ of f,

$$\left.\frac{\partial f}{\partial y_i'} \frac{\partial Y_i}{\partial s}\right|_{s=0}$$

is a first integral of the Euler–Lagrange equation (where the summation convention applies).

Proof.

$$\begin{aligned}
0 &= \left.\frac{\mathrm{d}}{\mathrm{d}s} f\right|_{s=0} \\
&= \left.\frac{\partial f}{\partial y_i} \frac{\mathrm{d}Y_i}{\mathrm{d}s}\right|_{s=0} + \left.\frac{\partial f}{\partial y_i'} \frac{\partial Y_i'}{\partial s}\right|_{s=0} \\
&= \left.\left[\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y_i'}\right)\frac{\mathrm{d}Y_i}{\mathrm{d}s} + \frac{\partial f}{\partial y_i'} \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}Y_i}{\mathrm{d}s}\right)\right]\right|_{s=0} \\
&= \left.\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{\partial f}{\partial y_i'} \frac{\partial Y_i}{\partial s}\right]\right|_{s=0} \\
\therefore \text{constant} &= \frac{\partial f}{\partial y_i'} \frac{\partial Y_i}{\partial s}
\end{aligned}$$

□

5.2 Conservation of momentum

Example. Consider a vector $\mathbf{y} = (y, z)$ and the function

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

Consider the symmetry

$$\begin{aligned}
Y = y + s &\implies Y' = y' \\
Z = z + s &\implies Z = z' \\
\therefore V(Y - Z) = V(y - z) &\implies \frac{\mathrm{d}}{\mathrm{d}s} f = 0
\end{aligned}$$

Then from Noether's theorem,

$$\text{constant} = \left.\left[\frac{\partial f}{\partial y'} \frac{\mathrm{d}Y}{\mathrm{d}s} + \frac{\partial f}{\partial z'} \frac{\mathrm{d}Z}{\mathrm{d}s}\right]\right|_{s=0} = y' + z'$$

This can be thought of as a conserved momentum in the $y + z$ direction.

5.3 Conservation of angular momentum under central force

Example. Suppose $\Theta = \theta + s, R = r$. Our space is isotropic, so $\frac{\mathrm{d}L}{\mathrm{d}s} = 0$, hence

$$\left[\frac{\partial L}{\partial \dot\theta}\frac{\partial \Theta}{\partial s} + \frac{\partial L}{\partial \dot r}\frac{\partial R}{\partial s}\right]\Bigg|_{s=0} = mr^2\dot\theta$$

which shows that angular momentum is conserved.

6 Convexity and the Legendre transform

6.1 Convex functions

This subsection is covered by Lecture 1 of the IB Optimisation course.

Definition. A set $S \subset \mathbb{R}^n$ is convex if $\forall \mathbf{x}, \mathbf{y} \in S, \forall t \in [0,1], (1-t)\mathbf{x} + t\mathbf{y} \in S$.

Definition. The graph of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the surface $\left\{(\mathbf{x}, z) \in \mathbb{R}^{n+1} : z - f(\mathbf{x}) = 0\right\}$.

Definition. A chord of a function $f : \mathbb{R}^n \to \mathbb{R}$ is a line segment connecting two points on the graph of f.

Definition. A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if
(i) the domain of f is a convex set; and
(ii) $\forall \mathbf{x}, \mathbf{y} \in S, \forall t \in (0,1), f((1-t)\mathbf{x} + t\mathbf{y}) \leq (1-t)f(\mathbf{x}) + tf(\mathbf{y})$
Equivalently, f is convex if the graph of f lies below (or on) all of its chords. We say that f is concave if f lies above (or on) all of its chords. Clearly, f is convex if and only if $-f$ is concave. We say f is *strictly* convex (or concave) if the inequality in (ii) becomes strict.

Example. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = x^2$$

The domain is clearly convex. To show convexity, we need

$$f((1-t)x + ty) - (1-t)f(x) - tf(y) \leq 0$$

We have

$$[(1-t)x + ty]^2 - (1-t)x^2 - ty^2 = x^2(1-t)(-t) + ty^2(1-t) + 2(1-t)txy = -(1-t)t(x-y)^2 < 0$$

as required. Hence $f(x) = x^2$ is a strictly convex function.

Example. Consider

$$f(x) = \frac{1}{x}$$

where the domain is $\mathbb{R} \setminus \{0\}$. This domain is not convex, so f is not convex. However, restricted to the domain $\{x \in \mathbb{R} : x > 0\}$, f can be shown to be convex.

6.2 Conditions for convexity

Proofs for these conditions, where appropriate, are given in Lecture 1 of the IB Optimisation course.

Theorem. If f is a once-differentiable function, then f is convex if and only if

$$f(\mathbf{y}) \geq f(\mathbf{x}) + (\mathbf{y} - \mathbf{x}) \cdot \nabla f(\mathbf{x})$$

Corollary. If f is convex, and has a stationary point, then it is a global minimum.

Proof. Suppose the stationary point is at $\mathbf{x}_0$, so $\nabla f(\mathbf{x}_0) = \mathbf{0}$. We then have

$$f(\mathbf{y}) \geq f(\mathbf{x}_0) + (\mathbf{y} - \mathbf{x}_0) \cdot \mathbf{0}$$

which is larger than $f(\mathbf{x}_0)$ as required. □

Theorem. If f is a once-differentiable function, then f is convex if

$$(\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})) \cdot (\mathbf{y} - \mathbf{x}) \geq 0$$

This can be thought of as stating that f' is monotonically increasing.

Theorem. If f is a twice-differentiable function, then f is convex if and only if

$$\nabla^2 f \geq 0$$

i.e. all eigenvalues of the Hessian matrix are non-negative. Note that $\nabla^2 f > 0$ implies strict convexity.

Example. Consider the function

$$f(x, y) = \frac{1}{xy}$$

for $x > 0, y > 0$. Then the Hessian is

$$H = \frac{1}{xy} \begin{pmatrix} \frac{2}{x^2} & \frac{1}{xy} \\ \frac{1}{xy} & \frac{2}{y^2} \end{pmatrix}$$

Then,

$$\det H = \frac{3}{x^3 y^3} > 0$$

$$\operatorname{tr} H > 0$$

Hence the eigenvalues are both positive. So f is strictly convex.

6.3 Legendre transform

Definition. The Legendre transform of a function $f : \mathbb{R}^n \to \mathbb{R}$ is a function $f^\star$ given by

$$f^\star(\mathbf{p}) = \sup_{\mathbf{x}}(\mathbf{p} \cdot \mathbf{x} - f(\mathbf{x}))$$

The domain of $f^\star$ is such that the supremum provided is finite. In one dimension, we can consider $f^\star(p)$ to be the maximum vertical distance between the graphs of $y = f(x)$ and $y = px$.

Example. Consider the function $f(x) = ax^2$, which is convex where $a > 0$. Computing the derivative of the right hand side and setting it to zero,

$$\begin{aligned} f^\star(p) &= \sup_x(px - ax^2) \\ &= p\Big(\frac{p}{2a}\Big) - a\Big(\frac{p}{2a}\Big)^2 \\ &= \frac{p^2}{4a} \end{aligned}$$

We can apply the Legendre transform twice:

$$f^{\star\star}(s) = \sup_p(sp - f^\star(p)) = as^2 = f(s)$$

In fact, if f is convex, then we always have $f^{\star\star} = f$. If $a < 0$, the supremum does not exist so $f^\star$ has an empty domain, and thus $f^{\star\star} \neq f$.

Proposition. If the domain of $f^\star$ is non-empty, it is a convex set, and $f^\star$ is convex.

Proof. Given $\mathbf{p}, \mathbf{q}$ in the domain of $f^\star$,

$$\begin{aligned} f^\star((1-t)\mathbf{p} + t\mathbf{q}) &= \sup_{\mathbf{x}}\,[(1-t)\mathbf{p} \cdot \mathbf{x} + t\mathbf{q} \cdot \mathbf{x} - f(\mathbf{x})] \\ &= \sup_{\mathbf{x}}\,[(1-t)(\mathbf{p} \cdot \mathbf{x} - f(\mathbf{x})) + t(\mathbf{q} \cdot \mathbf{x} - f(\mathbf{x}))] \\ &\leq \sup_{\mathbf{x}}\,[(1-t)(\mathbf{p} \cdot \mathbf{x} - f(\mathbf{x}))] + \sup_{\mathbf{x}}\,[t(\mathbf{q} \cdot \mathbf{x} - f(\mathbf{x}))] \\ &< \infty \end{aligned}$$

as required. □

In practice, if f is convex and differentiable, we compute $f^\star(\mathbf{p})$ by considering the derivative:

$$\nabla(\mathbf{p} \cdot \mathbf{x} - f(\mathbf{x})) = 0 \implies \mathbf{p} = \nabla f$$

If f is strictly convex, the condition $\mathbf{p} = \nabla f$ has a unique inverse to give $\mathbf{x}$ as a function of $\mathbf{p}$, so $f^\star(\mathbf{p}) = \mathbf{p} \cdot \mathbf{x}(\mathbf{p}) - f(\mathbf{x}(\mathbf{p}))$. This eliminates the supremum condition.

6.4 Applications to thermodynamics

If we consider the particles in a gas, we could theoretically solve the Euler–Lagrange equations for a system of around 10^{23} particles. However, solving such a complicated system is difficult. Instead of solving for each particle, we instead consider macroscopic quantities such as pressure P, volume V, temperature T, and entropy S. A system has *internal energy* $U(S, V)$. The *Helmholtz free energy* is

$$\begin{aligned} F(T, V) &= \min_S (U(S, V) - TS) \\ &= -\max_S (TS - U(S, V)) \\ &= -U^\star(T, V) \end{aligned}$$

where $U^\star$ is the Legendre transform of U with respect to S, fixing V constant. Assuming U is convex,

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

There are other thermodynamical quantities that can be represented using a Legendre transform, for instance enthalpy $H(S, P)$.

$$\begin{aligned} H(S, P) &= \min_V (U(S, V) + PV) \\ &= -U^\star(-P, S) \end{aligned}$$

At this minimum, $P = -\frac{\partial U}{\partial V}\big|_S$. We can think of the Legendre transform in this context as a way of swapping from dependence on entropy and volume to dependence on other variables.

6.5 Legendre transform of the Lagrangian

Recall that the Lagrangian in mechanics was defined as

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

This is a function on the configuration space. We define the *Hamiltonian* to be the Legendre transform of L with respect to $\dot{\mathbf{q}}$. We find, assuming that L is convex,

$$\begin{aligned} H(\mathbf{q}, \mathbf{p}, \mathbf{t}) &= \sup_{\mathbf{v}} (\mathbf{p} \cdot \mathbf{v} - L) \\ &= \mathbf{p} \cdot \mathbf{v}(\mathbf{p}) - L(\mathbf{q}, \mathbf{v}(\mathbf{p}), t) \end{aligned}$$

where $\mathbf{v}(\mathbf{p})$ is the solution to $p_i = \frac{\partial L}{\partial \dot{q}_i}$. The $\mathbf{p}$ are referred to as *generalised momenta* or *conjugate momenta*. Consider

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

Then,

$$\mathbf{p} = \frac{\partial L}{\partial \dot{\mathbf{q}}} = m\dot{\mathbf{q}} \implies \dot{\mathbf{q}} = \frac{1}{m}\mathbf{p}$$

The Hamiltonian is therefore

$$\begin{aligned}
H(\mathbf{q}, \mathbf{p}, t) &= \mathbf{p} \cdot \frac{1}{m}\mathbf{p} - L \\
&= \mathbf{p} \cdot \frac{1}{m}\mathbf{p} - \left(\frac{1}{2}m\frac{|\mathbf{p}|^2}{m^2} - V(\mathbf{q})\right) \\
&= \frac{1}{2m}|\mathbf{p}|^2 + V(\mathbf{q}) \\
&= T + V
\end{aligned}$$

6.6 Hamilton's equations from Euler–Lagrange equation

Given that the Lagrangian satisfies the Euler–Lagrange equation, we can deduce analogous equations for the Hamiltonian. We often write the indices of the generalised coordinates in superscript, as follows, where the summation convention applies:

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

Using this equation, we can compute two expressions for the differential of the Hamiltonian:

$$\begin{aligned}
\mathrm{d}H &= \frac{\partial H}{\partial q^i}\,\mathrm{d}q^i + \frac{\partial H}{\partial p_i}\,\mathrm{d}p_i + \frac{\partial H}{\partial t}\,\mathrm{d}t \\
&= p_i\,\mathrm{d}\dot{q}^i + \dot{q}^i\,\mathrm{d}p_i - \frac{\partial L}{\partial q^i}\,\mathrm{d}q^i - \frac{\partial L}{\partial \dot{q}^i}\,\mathrm{d}\dot{q}^i - \frac{\partial L}{\partial t}\,\mathrm{d}t
\end{aligned}$$

Now, note that $\frac{\partial L}{\partial \dot{q}^i} = p_i$. This cancels some terms. Making use of the Euler–Lagrange equation,

$$\frac{\partial L}{\partial q^i} = \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}^i} = \frac{\mathrm{d}}{\mathrm{d}t}p_i = \dot{p}_i$$

This gives

$$\mathrm{d}H = \frac{\partial H}{\partial q^i}\,\mathrm{d}q^i + \frac{\partial H}{\partial p_i}\,\mathrm{d}p_i + \frac{\partial H}{\partial t}\,\mathrm{d}t = \dot{q}^i\,\mathrm{d}p_i - \dot{p}_i\,\mathrm{d}q^i - \frac{\partial L}{\partial t}\,\mathrm{d}t$$

Comparing the differentials, we can see that

$$\dot{q}^i = \frac{\partial H}{\partial p_i}; \quad \dot{p}_i = -\frac{\partial H}{\partial q^i}; \quad \frac{\partial L}{\partial t} = -\frac{\partial H}{\partial t}$$

This system of equations is known as Hamilton's equations. Note that in the last equation, $\left.\frac{\partial}{\partial t}\right|_{q,\dot{q}} \neq \left.\frac{\partial}{\partial t}\right|_{p,q}$. For now, we will assume that there is no explicit t dependence in the Lagrangian. Then, Hamilton's equations are a system of $2n$ first-order ordinary differential equations. (Note, for comparison, that the Euler–Lagrange equations were a system of n second-order differential equations, which gives the same amount of initial conditions.) The initial conditions are typically a configuration of $\mathbf{p}, \mathbf{q}$ at some fixed t_0. The solutions to Hamilton's equations are called the *trajectories* in $2n$-dimensional phase space.

6.7 Hamilton's equations from extremising a functional

Note that we can also arrive at Hamilton's equations by extremising a functional in phase space.

$$S[\mathbf{q},\mathbf{p}] = \int_{t_1}^{t_2} \big(\dot{q}^i p_i - H(\mathbf{q},\mathbf{p},t)\big)\,\mathrm{d}t$$

The integrand, denoted f, is a function of $\mathbf{q},\mathbf{p},\dot{\mathbf{q}},t$. Writing the Euler–Lagrange equations for S, varying first with respect to p_i,

$$\frac{\partial f}{\partial p_i} - \underbrace{\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial f}{\partial \dot{p}_i}}_{0} = 0 \implies \dot{q}^i = \frac{\partial H}{\partial p_i}$$

Now varying with respect to q^i,

$$\frac{\partial f}{\partial q^i} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial f}{\partial \dot{q}^i} = 0 \implies \dot{p}_i = -\frac{\partial H}{\partial q^i}$$

These results are exactly Hamilton's equations.

7 Second variations

7.1 Conditions for local minimisers

The Euler–Lagrange equation gives a necessary condition for a stationary point. We cannot tell whether this leads to a minimum, a maximum, or a saddle point, just from the Euler–Lagrange equation. We can analyse the nature of the stationary points by considering the second variation. Consider the functional

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

where y is perturbed by a perturbation $\varepsilon\eta$. Let us assume that y is a solution to the Euler–Lagrange equation, so has no first variation. We will then expand $F[y+\varepsilon\eta]$ to second order.

$$\begin{aligned}
F[y+\varepsilon\eta] &= \int_\alpha^\beta [f(x,y+\varepsilon\eta,y'+\varepsilon\eta')]\,\mathrm{d}x \\
F[y+\varepsilon\eta] - F[y] &= \int_\alpha^\beta [f(x,y+\varepsilon\eta,y'+\varepsilon\eta') - f(x,y,y')]\,\mathrm{d}x \\
&= 0 + \varepsilon \underbrace{\int_\alpha^\beta \eta\left(\frac{\partial f}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial f}{\partial y'}\right)\mathrm{d}x}_{\text{zero by Euler–Lagrange equation}} \\
&\quad + \frac{1}{2}\varepsilon^2 \int_\alpha^\beta \left(\eta^2\frac{\partial^2 f}{\partial y^2} + \eta'^2\frac{\partial^2 f}{\partial (y')^2} + 2\eta\eta'\frac{\partial^2 f}{\partial y\partial y'}\right)\mathrm{d}x + O(\varepsilon^3)
\end{aligned}$$

The last term (excluding the ε^2 component) is called the second variation. We write

$$\delta^2 F[y] \equiv \frac{1}{2}\int_\alpha^\beta \left(\eta^2\frac{\partial^2 f}{\partial y^2} + \eta'^2\frac{\partial^2 f}{\partial (y')^2} + \frac{\mathrm{d}}{\mathrm{d}x}(\eta^2)\frac{\partial^2 f}{\partial y\partial y'}\right)\mathrm{d}x$$

Integrating the last term by parts, using $\eta = 0$ at α, β, we have

$$\delta^2 F[y] = \frac{1}{2}\int_\alpha^\beta \left(Q\eta^2 + P(\eta')^2\right) \mathrm{d}x$$

where

$$P = \frac{\partial^2 f}{\partial (y')^2}; \quad Q = \frac{\partial^2 f}{\partial y^2} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial^2 f}{\partial y \partial y'}\right)$$

Thus, if y is a solution to the Euler–Lagrange equation, and also $Q\eta^2 + P(\eta')^2 > 0$ for all η vanishing at α, β, then y is a local minimiser of F.

Example. We will prove that the geodesic on a plane is a local minimiser of path length. The functional we will analyse is given by

$$f = \sqrt{1 + (y')^2}$$

Hence,

$$P = \frac{\partial^2 f}{\partial (y')^2} = \frac{\partial}{\partial y'}\left(\frac{y'}{\sqrt{1+(y')^2}}\right) = \frac{1}{(1+(y')^2)^{\frac{3}{2}}} > 0$$

$$Q = 0$$

Therefore the second variation is positive, so any y that satisfies the Euler–Lagrange equation minimises path length. In particular, straight lines minimise path length on the plane.

7.2 Legendre condition for minimisers

Proposition (Legendre condition)**.** If $y_0(x)$ is a local minimiser, then $P|_{y=y_0} \geq 0$.

We can say that the Legendre condition is a necessary condition for a minimiser. In less formal terms, P is 'more important' than Q when determining if a stationary point is a minimiser.

Proof. This condition is not proven rigorously. However, the general idea of the proof is to construct a function η which is small everywhere (giving a small Q contribution), but oscillates very rapidly near some point x_0, at which $P < 0$. This gives a large P contribution which can overpower the Q contribution. Then this gives $Q\eta^2 + P(\eta')^2 < 0$ if there exists some x_0 where $P|_{y=y_0} < 0$. □

Note that the Legendre condition is not a sufficient condition for local minima, but $P > 0$ and $Q \geq 0$ is sufficient.

Example. Consider again the brachistochrone problem.

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

We have

$$\frac{\partial f}{\partial y} = -\frac{1}{2y} f$$

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

Hence

$$P = \frac{1}{(1+(y')^2)^{\frac{3}{2}}\sqrt{-y}} > 0$$

$$Q = \frac{1}{2\sqrt{1+(y^2)^2}y^2\sqrt{-y}} > 0$$

Hence the cycloid is a local minimiser of the time taken to travel between the two points.

7.3 Associated eigenvalue problem

When deriving the minimiser condition, we had the integrand

$$Q\eta^2 + P(\eta')^2$$

We can integrate this by parts:

$$Q\eta^2 + \frac{\mathrm{d}}{\mathrm{d}x}(P\eta\eta') - \eta\frac{\mathrm{d}}{\mathrm{d}x}(P\eta')$$

giving

$$\delta^2 F[y] = \frac{1}{2}\int_\alpha^\beta \eta[-(P\eta')' + Q\eta]\,\mathrm{d}x$$

The bracketed term $-(P\eta')' + Q\eta$ is known as the Sturm–Liouville operator acting on η, denoted $\mathcal{L}(\eta)$. If there exists η such that $\mathcal{L}(\eta) = -\omega^2\eta$, $\omega \in \mathbb{R}$, and $\eta(\alpha) = \eta(\beta) = 0$, then y is not a minimiser, since the integrand will be $-\omega^2\eta^2 < 0$.

Example. Consider

$$F[y] = \int_0^\beta \left((y')^2 - y^2\right)\mathrm{d}x$$

such that

$$y(0) = y(\beta) = 0; \quad \beta \neq k\pi, k \in \mathbb{N}$$

The Euler–Lagrange equation gives

$$y'' + y = 0$$

Thus, constrained to the boundary conditions, the only stationary point of F is

$$y \equiv 0$$

Analysing the second variation,

$$\delta^2 F[0] = \frac{1}{2}\int_0^\beta \left[\eta'^2 - \eta^2\right]\mathrm{d}x$$

giving

$$P = 1 > 0; \quad Q < 0$$

Let us now examine the eigenvalue problem, since we cannot find whether $y \equiv 0$ is a minimiser from what we know already. Consider the eigenvalue problem

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

Let us take

$$\eta = A\sin\left(\frac{\pi x}{\beta}\right)$$

to give

$$\left(\frac{\pi}{\beta}\right)^2 = 1 - \omega^2$$

So this has a solution $\omega > 0$ if and only if $\beta > \pi$. If $P > 0$, a problem may arise if the interval of integration is 'too large' (in this case $\beta > \pi$). Next lecture we will make this notion precise.

7.4 Jacobi accessory condition

Legendre tried to prove that $P > 0$ implied local minimality; obviously this was impossible due to the counterexample shown above. However, the method he used is still useful to analyse, since we can find an actual sufficient condition using the same idea. Let $\phi(x)$ be any differentiable function of x on $[\alpha, \beta]$. Then note that

$$\int_\alpha^\beta \frac{\mathrm{d}}{\mathrm{d}x}(\phi\eta^2)\,\mathrm{d}x = 0$$

since $\eta(\alpha) = \eta(\beta) = 0$. We can expand the integrand to give

$$\int_\alpha^\beta (\phi'\eta^2 + 2\eta\eta'\phi)\,\mathrm{d}x = 0$$

We can add this new zero to both sides of the second variation equation.

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

Now, suppose that $P > 0$ at a particular y. Then, we can complete the square on the integrand, giving

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

If we could choose a ϕ such that the second bracket vanishes, then the integrand would be $P\left(\eta' + \frac{\phi}{P}\eta\right)^2$. The only way the integral can be zero is if $\eta' + \frac{\phi}{P}\eta \equiv 0$. Since $\eta = 0$ at α, we have $\eta'(\alpha) = 0$. Hence, $\eta \equiv 0$ by the uniqueness of solutions to first order differential equations. Therefore, by contradiction, the integrand is not identically zero, and the second variation is positive. Now, such a ϕ function is given by

$$\phi^2 = P(Q + \phi')$$

If a solution to this differential equation exists, then $\delta^2 F[y] > 0$. We can transform this non-linear equation into a second order equation by the substitution $\phi = -P\frac{u'}{u}$ for some function $u \neq 0$. We have

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

Hence,

$$-(Pu')' + Qu = 0$$

This is known as the Jacobi accessory condition. Note that the left hand side is just $\mathcal{L}(u)$, where $\mathcal{L}$ is the Sturm–Liouville operator.

7.5 Solving the Jacobi condition

We need to find a solution to $\mathcal{L}(u) = 0$, where $u \neq 0$ on $[\alpha, \beta]$. The solution we find may not be nonzero on a large enough interval, in which case we would not have a local minimum.

Example. Consider

$$F[y] = \frac{1}{2}\int_\alpha^\beta \left((y')^2 - y^2\right) \mathrm{d}x$$

The second variation is

$$\delta^2 F[y] = \frac{1}{2}\int_\alpha^\beta \left((\eta')^2 - \eta^2\right) \mathrm{d}x$$

In this case, $P = 1, Q = -1$. The Jacobi accessory equation is

$$u'' + u = 0$$

We can solve this to find

$$u = A\sin x - B\cos x; \quad A, B \in \mathbb{R}$$

We want this to be nonzero on the interval $[\alpha, \beta]$. In particular,

$$\tan x \neq \frac{B}{A}; \quad \forall x \in [\alpha, \beta]$$

Note that $\tan x$ repeats every π, so if $|\beta - \alpha| < \pi$ we have a positive second variation for any stationary y.

Example. Consider again the geodesic on a sphere.

$$F[\theta] = \int \sqrt{\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\phi^2} = \int \sqrt{(\theta')^2 + \sin^2\theta}\, \mathrm{d}\phi$$

We have already proven that critical points of this functional are segments of great circles. Considering an equatorial great circle (since all great circles are equatorial under a change of perspective),

$$\theta = \frac{\pi}{2}$$

Consider ϕ_1, ϕ_2 on this great circle. The minor arc is clearly the shortest path, but the major arc is also a stationary point and must still be analysed.

$$P = 1; \quad Q = -1$$

Thus,

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

which is exactly the example from above. This is a minimiser if $|\phi_2 - \phi_1| < \pi$, which is exactly the condition of being a minor arc. If $\phi_2 - \phi_1 = \pi$, we have an infinite amount of geodesics, since these represent antipodal points. The set of geodesics exhibit rotational symmetry.