Variational calculus

Kukka-Emilia Huhtinen

Contents

1 Introduction

Variational calculus is used to find extrema of functionals, which are mappings from a function space to scalars. The problem is therefore not to simply find a point at which a function is an extremum, but to solve for functions which extremize an expression. Some famous problems, like the Brachistochrone problem or finding the shortest line between two points on a surface, can be solved using calculus of variations. In physics, the method is typically used in classical mechanics, but also has applications in, for example, field theory, general relativity, and quantum mechanics.

2 The Euler-Lagrange equation

The simplest problem of variational calculus is to determine a function $y(x)$ that extremizes the functional

$$I = \int_{x_A}^{x_B} f(x, y, y')\mathrm{d}x, \tag{1}$$

given boundary conditions $y(x_A) = y_A$ and $y(x_B) = y_B$ at the fixed endpoints x_A and x_B. The function f is assumed to be continuous and twice differentiable. To solve the problem, we will derive the differential equation y must obey to make I a stationary point. Two ways to obtain the equation are presented here : the first involves using test functions, and the second uses functional differentiation.

2.1 Derivation with test functions

Let us denote the function which extremizes I as $y(x)$. We now define a family of test functions $\overline{y}(x,\epsilon) = y(x) + \epsilon\eta(x)$, where η is an arbitrary twice differentiable function that satisfies $\eta(x_A) = \eta(x_B) = 0$ (see Fig. 1). The following conditions then hold for the test functions:

(i) $\overline{y}(x_A,\epsilon) = y_A$ and $\overline{y}(x_B,\epsilon) = y_B$ for all ϵ,

(ii) $\overline{y}(x,0) = y(x)$,

(iii) $\overline{y}(x,\epsilon)$ is twice differentiable for all ϵ.

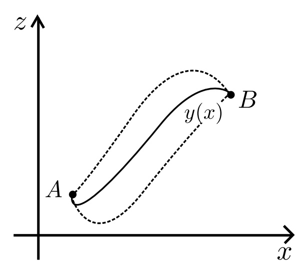


Figure 1: Example of solution function $y(x)$ (full line), and test functions (dashed lines).

We now define the function

$$I(\epsilon) = \int_{x_A}^{x_B} f(x,\overline{y},\overline{y}')\mathrm{d}x.$$

Clearly, when $\epsilon = 0$, the test functions are replaced by y, the extremizing function. By the definition of y, $I(\epsilon)$ therefore has an extremum at $\epsilon = 0$, and

$$\left.\frac{\mathrm{d}I(\epsilon)}{\mathrm{d}\epsilon}\right|_{\epsilon=0} = 0.$$

Using the chain rule, the derivative can be written as

$$\frac{\mathrm{d}I(\epsilon)}{\mathrm{d}\epsilon} = \int_{x_A}^{x_B} \frac{\partial f}{\partial \overline{y}}\frac{\mathrm{d}\overline{y}}{\mathrm{d}\epsilon} + \frac{\partial f}{\partial \overline{y}'}\frac{\mathrm{d}\overline{y}'}{\mathrm{d}\epsilon}\ \mathrm{d}x.$$

As the test functions are continuously differentiable (condition (iii)), the order of differentiation in the second term can be changed:

$$\begin{aligned}\frac{\mathrm{d}I(\epsilon)}{\mathrm{d}\epsilon} &= \int_{x_A}^{x_B} \frac{\partial f}{\partial \overline{y}}\frac{\mathrm{d}\overline{y}}{\mathrm{d}\epsilon} + \frac{\partial f}{\partial \overline{y}'}\frac{\mathrm{d}}{\mathrm{d}x}\frac{\mathrm{d}\overline{y}}{\mathrm{d}\epsilon}\ \mathrm{d}x \\ &= \int_{x_A}^{x_B} \frac{\partial f}{\partial \overline{y}}\frac{\mathrm{d}\overline{y}}{\mathrm{d}\epsilon}\ \mathrm{d}x + \left.\frac{\mathrm{d}\overline{y}}{\mathrm{d}\epsilon}\frac{\partial f}{\partial \overline{y}'}\right|_{x_A}^{x_B} - \int_{x_A}^{x_B} \frac{\mathrm{d}\overline{y}}{\mathrm{d}\epsilon}\frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial f}{\partial \overline{y}'}\ \mathrm{d}x.\end{aligned}$$

By condition (i), the ϵ–derivative of $\overline{y}$ at x_A and x_B is zero, so the second term is zero. The expression then simplifies to

$$\frac{\mathrm{d}I(\epsilon)}{\mathrm{d}\epsilon} = \int_{x_A}^{x_B} \frac{\mathrm{d}\overline{y}}{\mathrm{d}\epsilon}\left[\frac{\partial f}{\partial \overline{y}} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial f}{\partial \overline{y}'}\right]\ \mathrm{d}x.$$

Setting $\epsilon = 0$ is equivalent to setting $\overline{y}(x,\epsilon) = y(x)$, $\overline{y}'(x,\epsilon) = y'(x)$, and $\mathrm{d}\overline{y}/\mathrm{d}\epsilon = \eta(x)$. Then

$$\left.\frac{\mathrm{d}I(\epsilon)}{\mathrm{d}\epsilon}\right|_{\epsilon=0} = \int_{x_A}^{x_B} \left[\frac{\partial f}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial f}{\partial y'}\right]\eta(x)\ \mathrm{d}x = 0.$$

As this integral is zero for arbitrary $\eta(x)$,

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

which is the Euler-Lagrange equation.

When f does not depend explicitly on x, the Euler-Lagrange equation simplifies considerably. Assuming $\partial f/\partial x = 0$,

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial f}{\partial y'} - \frac{\partial f}{\partial y} = 0 \\
\Rightarrow &y'\frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial f}{\partial y'} - y'\frac{\partial f}{\partial y} - \frac{\partial f}{\partial x} = y'\frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial f}{\partial y'} + y''\frac{\partial f}{\partial y'} - y''\frac{\partial f}{\partial y'} - y'\frac{\partial f}{\partial y} - \frac{\partial f}{\partial x} = 0 \\
\Rightarrow &\frac{\mathrm{d}}{\mathrm{d}x}\left[y'\frac{\partial f}{\partial y'} - f\right] = 0 \\
\Rightarrow &y'\frac{\partial f}{\partial y'} - f = \text{constant}. \qquad (3)
\end{aligned}$$

This result is known as the Beltrami identity.

It should be noted that the assumptions made in the derivation limit the applicability of the Euler-Lagrange equation. All functions were assumed to be continuous and differentiable, so if the solution is actually discontinuous (which can be the case in e.g. control theory), a different method should be used to solve the problem. Moreover, the condition $\partial I(\epsilon)/\partial\epsilon = 0$ at $\epsilon = 0$ is a necessary but not sufficient condition for an extremum: the solution of the Euler-Lagrange equation can be a minimum, maximum, or inflection point. The nature of the solution is usually clear from the nature of the problem, but if it is not, one has to study higher derivatives of $I(\epsilon)$.

2.2 Derivation with functional differentiation

Recall the definition of the derivative of a function

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \lim_{\epsilon\to 0}\frac{f(x+\epsilon) - f(x)}{\epsilon}.$$

The derivative is the rate of change of the function as an infinitesimal change is applied on x. The functional derivative is defined in a similar way, where an infinitesimal change is applied on the function taken as a parameter by the functional (see e.g. [3]) :

$$\frac{\delta F}{\delta y(x)} = \lim_{\epsilon\to 0}\frac{F[y(x') + \epsilon\delta(x-x')] - F[y(x')]}{\epsilon}, \qquad (4)$$

where x' is a dummy variable integrated out in the functional.

The functional derivative of

$$I = \int_{x_A}^{x_B} f[x', y(x'), y'(x')]\ \mathrm{d}x'$$

would contain derivatives of the Dirac delta. This is fine in our case because this only occurs inside integrals, but we can interpret the Dirac delta function as the limit of a function, $\delta(x) = \lim_{h\to 0^+} g_h(x)$. For example, we can choose $g_h(x) = \exp[-x^2/(4h)]/(2\sqrt{\pi h})$. Then the functional derivative of I can be written as

$$\frac{\delta I}{\delta y(x)} = \lim_{\epsilon\to 0}\lim_{h\to 0^+}\frac{1}{\epsilon}\int_{x_A}^{x_B} f[x', y(x') + \epsilon g_h(x-x'), y'(x') + \epsilon g_h'(x-x')] - f[x', y, y']\ \mathrm{d}x'.$$

The Taylor expansion of the first term around $\epsilon = 0$ is

$$f[x', y+\epsilon g_h(x-x'), y'+\epsilon g_h'(x-x')] = f[x', y, y'] + \epsilon\frac{\partial f[x', y, y']}{\partial y(x')}g_h(x-x') + \epsilon\frac{\partial f[x', y, y']}{\partial y'(x')}g_h'(x-x') + \mathcal{O}(\epsilon^2).$$

When taking the limit $\epsilon \to 0$, the terms of higher than first order in ϵ tend to zero, so

$$\frac{\delta I}{\delta y(x)} = \lim_{h\to 0^+} \int_{x_A}^{x_B} \frac{\partial f[x', y(x'), y'(x')]}{\partial y(x')} g_h(x-x') + \frac{\partial f[x', y(x'), y'(x')]}{\partial y'(x')} g_h'(x-x')\ \mathrm{d}x'.$$

Integrating by parts and assuming x is in the interval (x_A, x_B), we obtain

$$\begin{aligned}\frac{\delta I}{\delta y(x)} &= \frac{\partial f[x, y(x), y'(x)]}{\partial y(x)} + \lim_{h\to 0^+} \left(\frac{\partial f[x', y(x'), y'(x')]}{\partial y'(x')} g_h(x-x') \Big|_{x'=x_A}^{x_B} - \int_{x_A}^{x_B} \frac{\mathrm{d}}{\mathrm{d}x'} \frac{\partial f[x', y(x'), y'(x')]}{\partial y'(x')} g_h(x-x')\ \mathrm{d}x' \right) \\ &= \frac{\partial f[x, y, y']}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial f[x, y, y']}{\partial y'}.\end{aligned}$$

At stationary points, the functional derivative is zero, so we again obtain the Euler-Lagrange equation:

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

2.3 Generalizations

2.3.1 Several independent functions

The Euler-Lagrange equation can easily be generalized for a functional of the form

$$I = \int_{x_A}^{x_B} f(x, y_1, y_2, \ldots, y_n, y_1', y_2', \ldots, y_n')\ \mathrm{d}x,$$

for which we want to determine the independent functions $y_1(x), y_2(x), \ldots, y_n(x)$ that extremize the functional. When we assume that $y_i(x)$ are in the extremizing form for all $i \neq j$, the problem is the same as in the case with only one function, which means y_j solves the Euler-Lagrange equation. Repeating the argument for all the other functions, we find that

$$\frac{\partial f}{\partial y_i} - \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial f}{\partial y_i'} = 0, \qquad i = 1, \ldots, n. \tag{5}$$

2.3.2 Higher derivatives

We consider a functional

$$I = \int_{x_A}^{x_B} f(x, y, y', \ldots, y^{(n)})\ \mathrm{d}x,$$

where $y^{(n)}$ denotes the n:th derivative of y. Here, we use functional differentiation to derive the Euler-Lagrange equation is this case, but the same result is obtained by using test functions.

As in Sec. 2.2, we Taylor expand $f[x', y+\epsilon g_h(x-x'), \ldots, y^{(n)} + \epsilon g_h^{(n)}(x-x')]$, where g_h are function such that $\delta(x) = \lim_{h\to 0^+} g_h(x)$, around $\epsilon = 0$ to compute the limit

$$\lim_{\epsilon\to 0} \lim_{h\to 0^+} \frac{1}{\epsilon} \int_{x_A}^{x_B} f[x', y+\epsilon g_h(x-x'), \ldots, y^{(n)} + \epsilon g_h^{(n)}(x-x')] - f[x', y, \ldots, y^{(n)}]\ \mathrm{d}x.$$

The term of first order in ϵ corresponding to the derivative of order $m \leq n$ is

$$\epsilon \frac{\partial f}{\partial y^{(m)}} g_h^{(m)}(x-x').$$

When computing the functional derivative of I, this term is repeatedly integrated by parts to yield

$$\begin{aligned}
\lim_{h\to 0^+}\int_{x_A}^{x_B}\frac{\partial f}{\partial y^{(m)}}g_h^{(m)}(x-x')\,\mathrm{d}x' &= \lim_{h\to 0^+}\left(\frac{\partial f}{\partial y^{(m)}}g_h^{(m-1)}(x-x')\Big|_{x_A}^{x_B}-\int_{x_A}^{x_B}\frac{\mathrm{d}}{\mathrm{d}x'}\frac{\partial f}{\partial y^{(m)}}g_h^{(m-1)}(x-x')\,\mathrm{d}x'\right)\\
&= \lim_{h\to 0^+}\left(-\frac{\mathrm{d}}{\mathrm{d}x'}\frac{\partial f}{\partial y^{(m)}}g_h^{(m-2)}(x-x')\Big|_{x_A}^{x_B}+\int_{x_A}^{x_B}\frac{\mathrm{d}^2}{\mathrm{d}x'^2}\frac{\partial f}{\partial y^{(m)}}g_h^{(m-2)}(x-x')\,\mathrm{d}x'\right)\\
&= \ldots = \lim_{h\to 0^+}\left((-1)^{(m-1)}\frac{\mathrm{d}^{(m-1)}}{\mathrm{d}x'^{(m-1)}}\frac{\partial f}{\partial y^{(m)}}g_h(x-x')\Big|_{x_A}^{x_B}\right.\\
&\left.+(-1)^m\int_{x_A}^{x_B}\frac{\mathrm{d}^m}{\mathrm{d}x'^m}\frac{\partial f}{\partial y^{(m)}}g_h(x-x')\,\mathrm{d}x'\right)\\
&= (-1)^m\frac{\mathrm{d}^m}{\mathrm{d}x^m}\frac{\partial f}{\partial y^{(m)}}.
\end{aligned}$$

A similar term is obtained for each $y^{(m)}$, $1\le m\le n$, and the Euler-Lagrange equation is

$$\frac{\partial f}{\partial y}+\sum_{k=1}^{n}(-1)^k\frac{\mathrm{d}^k}{\mathrm{d}x^k}\frac{\partial f}{\partial y^{(k)}}=0. \tag{6}$$

2.3.3 Multiple integrals

The Euler-Lagrange equation can be generalized to extremize functionals

$$I=\int_S f(x_i,y_j,\frac{\partial y_j}{\partial x_i})\,\mathrm{d}x_1\mathrm{d}x_2\ldots\mathrm{d}x_n,$$

where S is an n-dimensional region and y_j, $j=1,2,\ldots,m$, are functions with a given value on the $(n-1)$-dimensional boundary of the region.

The set of m Euler-Lagrange equations to solve in this case is

$$\frac{\partial f}{\partial y_j}-\sum_{k=1}^{n}\frac{\partial}{\partial x_k}\frac{\partial f}{\partial\left(\frac{\partial y_j}{\partial x_k}\right)}=0,\qquad j=1,2,\ldots,m. \tag{7}$$

The derivation of the equation for $n=2$ is given in [1]. Note that $\partial f/\partial(\partial y_j/\partial x_k)$ is considered an explicit function of only $x_1,\ldots,x_n$, meaning the functions y_i and their derivatives should be considered explicit functions of $x_1,\ldots,x_n$ when taking the partial derivative with reference to x_k. Keeping this in mind, it is easy to see Eq. (7) becomes the usual Euler-Lagrange equation in one fimension.

2.4 Examples

2.4.1 Shortest line

Suppose we are given two points $A=(x_A,y_A)$ and $B=(x_B,y_B)$, $x_A<x_B$, in a plane, and we want to determine the shortest path between the two. The length of a curve given by a function $y(x)$ is

$$I=\int_A^B\sqrt{1+y'^2}\,\mathrm{d}x.$$

Applying the Euler-Lagrange equation to $f = \sqrt{1+y'^2}$, we get

$$\begin{aligned}
\frac{\partial f}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial f}{\partial y'} &= -\frac{\mathrm{d}}{\mathrm{d}x}\frac{y'}{\sqrt{1+y'^2}} = 0 \\
\Rightarrow \frac{y'}{\sqrt{1+y'^2}} &= \text{constant.}
\end{aligned}$$

This implies that y' is a constant, so

$$y = ax + b,$$

where $a = (y_A - y_B)/(x_A - x_B)$ and $b = (x_A y_B - x_B y_A)/(x_A - x_B)$. From the nature of the problem, it is clear this function minimizes the functional I : the path between the points can be made arbitrarily long. The shortest line between two points is therefore the straight line.

Note that since f does not depend explicitly on x, we could also have used the Beltrami identity to get the same result.

2.4.2 Brachistochrone

The Brachistochrone problem was proposed by Johann Bernoulli in 1696 [1]. The problem is to find the path between two points A and B (A being the highest point) that a free-falling particle M will travel in the shortest time (see Fig. 2).

We define the origin of the coordinate system to be A, and take the y-axis pointing downward. The time taken to travel from A to B along a curve s is

$$I = \int_A^B \frac{\mathrm{d}s}{v} = \int_A^B \frac{\sqrt{1+y'^2}}{v}\,\mathrm{d}x.$$

As the particle is free-falling,

$$\frac{1}{2}mv^2 = mgy \quad \Rightarrow \quad v = \sqrt{2gy}.$$

The functional to minimize is thus

$$I = \frac{1}{\sqrt{2g}}\int_0^{x_B} \sqrt{\frac{1+y'^2}{y}}\,\mathrm{d}x.$$

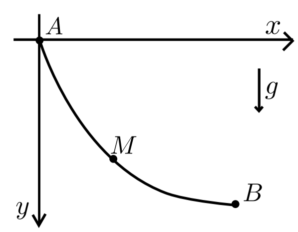


Figure 2: The Brachistochrone.

The function $f = \sqrt{(1+y'^2)/y}$ does not depend explicitly on x, so instead of applying the Euler-Lagrange equation, we can apply the Beltrami identity.

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

Defining $1/C^2 = 2a$, the equation becomes

$$\begin{aligned}
y' &= \sqrt{\frac{2a-y}{y}} \\
x - x_0 &= \int \frac{1}{y'}\,\mathrm{d}y = \int \sqrt{\frac{y}{2a-y}}\,\mathrm{d}y.
\end{aligned}$$

We change variables $y = a(1 - \cos(\phi))$, and

$$x - x_0 = 2a \int \sin^2 \frac{\phi}{2} \, \mathrm{d}\phi = a(\phi - \sin(\phi)).$$

The solution to the problem is therefore the cycloid

$$x = a(\phi - \sin(\phi)) + x_0, \qquad y = a(1 - \sin\phi),$$

where a and x_0 are constants determined by the coordinates of A and B.

3 Introducing constraints

3.1 Global constraints

In some problems, we want to extremize the integral

$$I = \int_{x_A}^{x_B} f(x, y, y') \, \mathrm{d}x,$$

while satisfying n global constraints of the form

$$J_i = \int_{x_A}^{x_B} g_i(x, y, y') \, \mathrm{d}x = C_i, \qquad i = 1, 2, \ldots, n,$$

where C_i are constants. This problem is solved with the use of test functions in [1]. A derivation using functional differentiation is presented here.

To minimize I while satisfying the constraints J_i, we introduce Lagrange multipliers λ_i, and define

$$K = I - \sum_{i=1}^{n} \lambda_i J_i = \int_{x_A}^{x_B} h(x, y, y') \, \mathrm{d}x,$$

where $h = f - \sum_i \lambda_i g_i$.

Once again, the functional is extremized when its derivative is zero. The derivation of $\delta K/\delta y$ is the same as was presented in Sec. 2.2, so the equation to be solved is

$$\frac{\delta K}{\delta y} = \frac{\partial h}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial h}{\partial y'} = 0, \tag{8}$$

with the constraints $J_i = C_i$.

If the functional depends on m independent functions y_i, the Euler-Lagrange equation is

$$\frac{\partial h}{\partial y_i} - \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial h}{\partial y_i'} = 0, \qquad i = 1, 2, \ldots, m, \tag{9}$$

where h is defined the same way as above. The full system of m Euler-Lagrange equations and n constraints, with the boundary conditions, determines the coordinates y_i and the Lagrange multipliers λ_i.

3.2 Local constraints

In this section, we extremize a functional of the form

$$I = \int_{x_A}^{x_B} f[x, y_1, \ldots, y_n, y_1', \ldots, y_n'] \; \mathrm{d}x,$$

this time while satisfying m local constraints of the form

$$g_i(x, y_1, \ldots, y_n) = 0, \qquad i = 1, \ldots, m.$$

As in the derivation in Sec. 2.1, we introduce a collection of test functions $\overline{y}_i(x, \epsilon_1, \ldots, \epsilon_{m+1})$ such that

(a) $\overline{y}_i(x_A, \epsilon_1, \ldots, \epsilon_{m+1}) = y_{i,A}$ and $\overline{y}_i(x_B, \epsilon_1, \ldots, \epsilon_{m+1}) = y_{i,B}$ for all $\epsilon_1, \ldots, \epsilon_{m+1}$,

(b) $\overline{y}_i(x, 0, \ldots, 0) = y_i(x)$, where y_i are the solutions,

(c) $\overline{y}_i$ are twice differentiable.

Note that we introduce $m+1$ parameters, where m is the number of constraints, so that the parameters are not determined by the constraints.

To get the constraints in integral form, we introduce a set of arbitrary functions $\phi_i(x), i = 1, \ldots, m$. Then

$$g_i(x, y_1, \ldots, y_n) = 0 \;\;\Leftrightarrow\;\; J_i = \int_{x_A}^{x_B} \phi_i(x) g_i(x, y_1, \ldots, y_n) \; \mathrm{d}x = 0$$

Now we proceed the same way as in the previous part, and introduce Lagrange multipliers λ_i to define the functional

$$K = I + \sum_{i=1}^{m} \lambda_i J_i = \int_{x_A}^{x_B} f + \sum_{i=1}^{m} \lambda_i(x) g_i \; \mathrm{d}x,$$

and the function

$$K(\epsilon_1, \ldots, \epsilon_{m+1}) = \int_{x_A}^{x_B} h(x, \overline{y}_1, \ldots, \overline{y}_n, \overline{y}_1', \ldots, \overline{y}_n') \; \mathrm{d}x, \tag{10}$$

where $\lambda_i(x) = \lambda_i \phi(x)$ and $h = f + \sum_{i=1}^{m} \lambda_i(x) g_i$. .

We compute the derivative of K with reference to each ϵ_i, and integrate by parts :

$$\frac{\partial K}{\partial \epsilon_j} = \sum_{i=1}^{n} \int_{x_A}^{x_B} \frac{\partial h}{\partial \overline{y}_i} \frac{\partial \overline{y}_i}{\partial \epsilon_j} + \frac{\partial h}{\partial \overline{y}_i'} \frac{\partial \overline{y}_i'}{\partial \epsilon_j} \; \mathrm{d}x = \sum_{i=1}^{n} \int_{x_A}^{x_B} \left[\frac{\partial h}{\partial \overline{y}_i} - \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial h}{\partial \overline{y}_i'} \right] \frac{\partial \overline{y}_i}{\partial \epsilon_j} \; \mathrm{d}x, \qquad j = 1, \ldots, m+1.$$

The derivatives are zero at $\epsilon_j = 0 \; \forall j$, so

$$\sum_{i=1}^{n} \int_{x_A}^{x_B} \left[\frac{\partial h}{\partial y_i} - \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial h}{\partial y_i'} \right] \eta_j^i(x) \; \mathrm{d}x = 0, \qquad j = 1, \ldots, m+1, \tag{11}$$

where $\eta_j^i = \partial \overline{y}_i / \partial \epsilon_j$.

Here, contrary to the derivation in Sec. 2.1, the functions η_j^i are not completely arbitrary. Indeed, the derivative of the i:th constraint with respect to ϵ_j at $\epsilon_k = 0 \; \forall k$ is

$$\left. \frac{\mathrm{d} g_i}{\mathrm{d} \epsilon_j} \right|_{\epsilon_k = 0 \; \forall k} = 0 = \sum_{l=1}^{n} \frac{\partial g_i}{\partial \overline{y}_l} \left. \frac{\partial \overline{y}_l}{\partial \epsilon_j} \right|_{\epsilon_k = 0 \; \forall k} = \sum_{l=1}^{n} \frac{\partial g_i}{\partial y_l} \eta_j^l = 0,$$

so the functions η_j^l are mutually dependent for each $1 \leq l \leq m+1$. However, we should remember the functions $\lambda_i(x), i = 1, \ldots, m$ are arbitrary. Equation (11) can be rewritten

$$\sum_{i=1}^{n} \int_{x_A}^{x_B} \left[\frac{\partial f}{\partial y_i} - \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial f}{\partial y_i'} + \sum_{l=1}^{n} \lambda_l(x) \frac{\partial g}{\partial y_i} \right] \eta_j^i(x) \ \mathrm{d}x = 0, \qquad j = 1, \ldots, m+1.$$

We can now choose the functions λ_i so that the coefficient of $\eta_j^i(x)$ vanishes for all $1 \leq i \leq m$, so

$$\frac{\partial f}{\partial y_i} - \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial f}{\partial y_i'} + \sum_{l=1}^{n} \lambda_l \frac{\partial g_l}{\partial y_i} = 0, \qquad i = 1, \ldots, m.$$

The last $n - m$ coefficients are then independent, and

$$\frac{\partial f}{\partial y_i} - \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial f}{\partial y_i'} + \sum_{l=1}^{n} \lambda_l \frac{\partial g}{\partial y_i} = 0, \qquad i = m+1, \ldots, n,$$

from Eq. (11). Combining the results, we obtain a set of n differential equations

$$\frac{\partial h}{\partial y_i} - \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial h}{\partial y_i'} = 0, \qquad j = i, \ldots, n. \tag{12}$$

Together with the m constraints

$$g_j(x, y_1, \ldots, y_n, y_1', \ldots, y_n') = 0, \qquad j = 1, \ldots, m, \tag{13}$$

and the boundary conditions on the functions y_i, these equations uniquely define the functions $y_1, y_2, \ldots, y_m$ and $\lambda_1, \lambda_2, \ldots, \lambda_n$.

3.3 Example : maximum area

Using the results derived in previous sections, we determine the shape that gives the largest surface area when the perimeter is fixed. That is, we want to find $x(t)$ and $y(t)$ so that

$$A = \frac{1}{2} \int_{t_A}^{t_B} (xy' - x'y) \ \mathrm{d}t,$$

is maximized when

$$L = \int_{t_A}^{t_B} \sqrt{x'^2 + y'^2} \ \mathrm{d}t$$

is fixed. We define

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

where λ is a Lagrange mutliplier. Applying the Euler-Lagrange equations, we get two differential equations

$$\frac{\partial h}{\partial x} - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial h}{\partial x'} = y' - \lambda \frac{\mathrm{d}}{\mathrm{d}t} \frac{x'}{\sqrt{x'^2 + y'^2}} = 0,$$
$$\frac{\partial h}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial h}{\partial y'} = x' + \lambda \frac{\mathrm{d}}{\mathrm{d}t} \frac{y'}{\sqrt{x'^2 + y'^2}} = 0.$$

By integration, we obtain

$$y - y_0 = \lambda \frac{x'}{\sqrt{x'^2 + y'^2}},$$
$$x - x_0 = -\lambda \frac{y'}{\sqrt{x'^2 + y'^2}}$$
$$\Rightarrow (x - x_0)^2 + (y - y_0)^2 = \lambda^2,$$

which is the equation of a circle of radius λ. Since the perimeter is L, $\lambda = L/(2\pi)$. As the area of a figure with perimeter L can be made arbitrarily small, but not arbitrarily large, it is clear that this solution gives a maximum. With a fixed perimeter, the area is therefore maximized in a circle.

4 Application to classical mechanics

In physics, a direct application of variational calculus is in Lagrangian mechanics, which is an alternative to the Newtonian formulation of classical mechanics. It is often advantageous over the latter, as solving constraint forces is not necessary in Lagrangian formalism, and the coordinates can be chosen more conveniently than when calculating with vectors. Moreover, Noether's theorem (see Sec. 5) is easily applicable, and allows to relate symmetries in the system to conserved quantities.

Let us consider a system specified by generalized coordinates $q_1(t), \dots, q_n(t)$, with kinetic energy T and potential energy V. The Lagrangian is defined as

$$L(t, q_i, \dot{q}_i) = T - V, \tag{14}$$

where the dot denotes a time derivative. The motion of the system follows Hamilton's principle : the motion from time t_A to t_B is such that the action

$$S = \int_{t_1}^{t_2} L \, \mathrm{d}t \tag{15}$$

is an extremum.

Based on the results from previous sections, the Lagrangian therefore solves the Euler-Lagrange equations

$$\frac{\partial L}{\partial q_i} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_i} = 0, \qquad i = 1, 2, \dots, n. \tag{16}$$

When the coordinates are constrained by

$$g_j(t, q_1, \dots, q_n) = 0, \qquad j = 1, 2, \dots m,$$

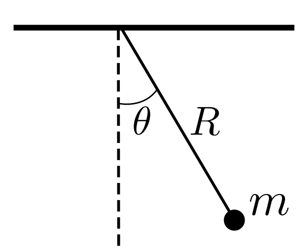
the set of equations to solve is

$$\frac{\partial L}{\partial q_i} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial L}{\partial \dot{q}_i} = \sum_{j=1}^{m} \lambda_j \frac{\partial g_j}{\partial q_i}, \qquad i = 1, \dots, n. \tag{17}$$

In mechanics, the Lagrange multipliers are related to the forces needed to constrain the motion [1, 2] : the generalized reaction forces for the system are

$$Q_i = \sum_{j=1}^{k} \lambda_j \frac{\partial g_j}{\partial q_i}. \tag{18}$$

As an example, we consider a simple pendulum of length R and mass m (see Fig. 3). We use generalized coordinates r and θ, and neglect friction and the mass of the string. The Lagrangian is



$$L = \frac{1}{2}m\dot{r}^2 + \frac{1}{2}mr^2\dot{\theta}^2 + mgr\cos\theta.$$

The motion is constrained so that

$$r - R = 0.$$

The Euler-Lagrange equations yield

$$\ddot{\theta} = -\frac{g}{r}\sin\theta,$$
$$m\ddot{r} = mr\dot{\theta}^2 + mg\cos\theta - \lambda.$$

Using the constraint $r - R = 0$, we obtain

$$\ddot{\theta} = -\frac{g}{R}\sin\theta,$$
$$\lambda = mr\dot{\theta}^2 + mg\cos\theta.$$

The first equation describes the motion of the pendulum, and the second is the constraint force needed to keep r equal to R. Note that the first equation could have been obtained by simply using only one generalized coordinate θ, without solving the force of constraint. The reason to include the generalized coordinate r and the corresponding constraint is to get information on the reaction force.

5 Noether's theorem

Let us consider a functional

$$I = \int_{x_A}^{x_B} f(x, y_q, \ldots y_n, y_1', \ldots y_n')\ \mathrm{d}x,$$

and a transformation

$$\tilde{x} = x + \epsilon\xi(x, y_1, \ldots, y_n, y_1', \ldots, y_n'),$$
$$\tilde{y}_i = y_i + \epsilon\eta_i(x, y_1, \ldots, y_n, y_1', \ldots, y_n').$$

According to Noether's theorem, if the functional I is invariant under this transformation, meaning

$$I(\epsilon) = \int_{x_A}^{x_B} f(\tilde{x}, \tilde{y}_1, \ldots, \tilde{y}_n, \tilde{y}_1', \ldots, \tilde{y}_n') = I\ \ \forall\epsilon,$$

there exists a corresponding conserved quantity

$$\sum_{i=1}^{n}\frac{\partial f}{\partial y_i'}\eta_i + \xi\left(f - \sum_{i=1}^{n} y_i'\frac{\partial f}{\partial y_i'}\right) = C. \tag{19}$$

Put another way, for every differentiable symmetry in the system there exists a corresponding conservation law. A proof of the theorem is given in [1].

As an example, let us consider a two-particle system with an interaction potential that depends only on the separation of the particles. The corresponding Lagrangian is

$$L = \frac{m_1}{2}|\dot{\mathbf{r_1}}|^2 + \frac{m_2}{2}|\dot{\mathbf{r_2}}|^2 - V(\mathbf{r_1} - \mathbf{r_2}).$$

The Lagrangian, and thus the action, is clearly invariant under the transformation

$$\tilde{t} = t + \epsilon\tau, \quad \tilde{\mathbf{r_1}} = \mathbf{r_1} + \epsilon\xi_\mathbf{1}, \quad \tilde{\mathbf{r_2}} = \mathbf{r_2} + \epsilon\xi_\mathbf{2},$$

where

$$\tau = 0, \quad \xi_{\mathbf{1}} = (1,0,0), \quad \xi_{\mathbf{2}} = (1,0,0).$$

According to Noether's theorem,

$$\sum_{i=1}^{2} \frac{\partial L}{\partial \dot{x_i}} \xi_i = \frac{\partial L}{\partial \dot{x_1}} + \frac{\partial L}{\partial \dot{x_2}} = p_{x_1} + p_{x_2} = C,$$

where p_{x_1} and p_{x_2} are the x-component of the linear momentum of particle 1 and 2, respectively. The x component of the total linear momentum of the system is therefore a constant of motion. In a similar way, it can be shown that the other components are also conserved, so the total linear momentum of the system is conserved. In general, translational invariance gives rise to conservation of linear momentum, rotational invariance to conservation of angular momentum, and time invariance to conservation of energy.

References

[1] F.W. Byron, R.W. Fuller, *Mathematical methods of classical and quantum physics*, Dover Publications, New York (1992).

[2] A.L. Fetter, J.D. Walecka, *Theoretical mechanics of particles and continua*, Dover Publications, Mineola, New York (2003).

[3] T. Lancaster, S.J. Blundell, *Quantum field theory for the gifted amateur*, Oxford university press (2014).