Part 7. From the Action to the Laws of Nature

PHYS 2022, Yi Wang, Department of Physics, HKUST

A new way of thinking about physics

We are used to Newton's way to think about physics – for a particle, by knowing the initial position and velocity, we calculate the trajectory of the particle as time evolves. This way of thinking is generalized to gravity, E&M fields, and so on.

These are all good. But if we really want to ask fundamental questions – can these equations of motion be the ultimate codes of nature – there are still a few unpleasant aspects:

1. Equations of motion are not invariant objects in special relativity. Because an equation of motion is time evolution. And in relativity time and space appears with almost equal rights. For different observers, equations of motion are covariant (transforms consistently with Lorentz transformation), but not invariant. Is the ultimate code of nature really so subjective? In the movie "Matrix", the world is a program. Has the programer choose a preferred frame in this movie to write down equations of motion and code the world?
2. Where do conservation laws come from? We are aware of energy, momentum, angular momentum, charge conservation laws. They can be proved given a particular framework of the theory. For example, given Newtonian mechanics we can prove energy and momentum conservations by smart tricks. But is there a universal way to figure out the origin of conservation laws for a general system?
3. Motion with constraints. For example, think about a pendulum (if it's too simple, think about double pendulum). In an ideal theory, constrained motion means that we are reducing possibilities and thus the situation should be simpler. But in the Newtonian way, the more constraints, the more analysis of forces and the more annoying math. Is there a way to make constrained systems simpler to compute?
4. The quantum world is so different from its classical counter part. We haven't introduced to you quantum mechanics yet but later we will address this concern as well.
5. The equations of motion are too "cold blooded" to describe how nature "would like to" behave. Can you use one sentence to conclude the nature of Nature, which applies to every known piece of fundamental physics? For example, "Nature prefers to $\cdots$"? Saying Nature prefers to satisfy its equation of motion does not seem pleasant enough.

A short sentence by Maupertuis (1698-1759) addresses all the above issues:

"Nature is thrifty in all its actions."

We will see how this simple sentence work.

The equation of motion

The equation to determine the particle's trajectory is known as the equation of motion, which is usually a second order differential equation (or a set of such equations) with time. For example, the motion of a particle in a potential $V(q)$ can be described by an equation of motion $\ddot{q} + dV/dq = 0$. To solve a realistic problem, the equation of motion is also packed with initial conditions: at an initial time t_0, the values of $q(t_0)$ and $\dot{q}(t_0)$. To solve an equation of motion with such initial conditions is called a Cauchy problem.

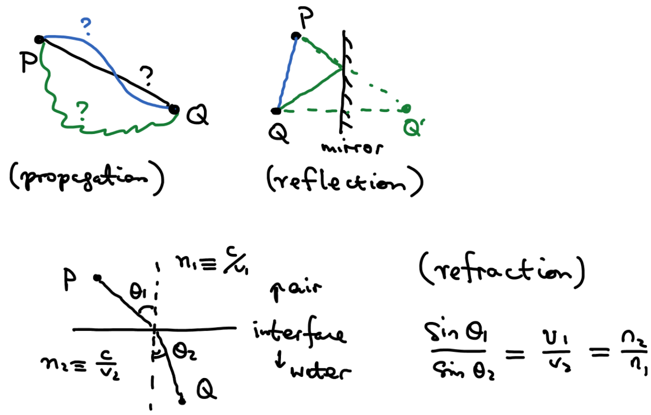
1 Fermat's Principle of Light

Before addressing the action principle, let us first use the Fermat's principle to study the propagation of light. This is not the real action principle yet, but the physical concept and math method are very similar. At the same time it is more intuitive. So let's first do this warm up exercise here.

Fermat's principle

Consider a class of light propagation problems: Given two fixed points P and Q, how does light propagate between those points?

Some examples are given in the figures to the right, including propagation in free space, reflection and refraction.



Fermat proposed a general solution for this type of problems: Light travels between two given points along the path of extremal time.

Pierre de Fermat

Pierre de Fermat (1601(or 1607) - 1665) is a French lawyer. Nobody remembers what case he has defended. But we do remember many things about him, including his conjecture that there is no positive integer solution for $x^n + y^n = z^n$ for $n > 2$. After 300 years, this was finally proven in 1994.

How travel time depends on path? Functionals

You are familiar to the problems of extremal problem in calculus: A function $f(x)$ takes extremal value when $df/dx = 0$. Here our situation is similar, but with a different mathematical object..

We are talking about the space (in math sense) of "path". How to parameterize a path in math? A path can be described as a function. For example, a curve $y = y(x)$.

Of course not all paths can be parametrized by $y = y(x)$, for example $x =$constant. But here let's restrict our attention to paths which can be described by $y = y(x)$, which provides us enough background to proceed to the action principle.

Suppose we know the speed of light (c in vacuum, c/n in media with reflection index n). Given a path, we know the light propagation time T. Here T depends on the whole function y (i.e. not only the value of $y(x)$ at a particular x). We then say that T is a *functional* of y, denoted as $T[y]$ with square brackets.

In short, a functional is a "function of function." A comparison of functions and functionals are sketched below.

Function: $f(x): \mathbb{R} \mapsto \mathbb{R}$

E.g. $\{1, 2, 3, \cdots\} \mapsto \{2, 4, 6, \cdots\}$ E.g. $\Rightarrow f(x) = 2x$

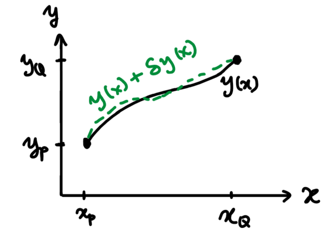
$f(x) = \frac{6}{7}x^2 - \frac{4}{7}x + \frac{12}{7}$

Functional: $g[f]: \{\text{functions}\} \mapsto \mathbb{R}$

E.g. $\{y=x, y=2x, y=3x, \cdots\} \mapsto \{1, 2, 3, \cdots\}$ E.g. $\Rightarrow g[f] = \int_0^{\sqrt{2}} f(x)\, dx,$

Functional programming

In programming language there is a paradigm known as functional programming. The very basic requirement is that you can pass a function as an argument of another function. It is the same functional (function of function) there.



Which path has extremal time? Functional variation

To find extremal time between (x_P, y_P) and (x_Q, y_Q), we first draw a (general) path $y(x)$ satisfying $y(x_P) = y_P$ and $y(x_Q) = y_Q$. Then we find the condition that a extremal path must satisfy with the following procedure:

We vary the path by $y(x) \to y(x) + \delta y(x)$, where δy is an *arbitrary* infinitesimal function (i.e. we ignore $(\delta y)^2$) satisfying $\delta y(x_P) = 0$ and $\delta y(x_Q) = 0$. These "boundary conditions" are needed because we have fixed these two boundary points by the definition of the problem.

We can now write the time of the new path as $T[y + \delta y]$, and thus the functional variation

$$\delta T \equiv T[y + \delta y] - T[y] \ . \tag{1}$$

A path with extremal time must have $\delta T = 0$ for all the possible variations $\delta y(x)$.

Why $\delta T = 0$ means extremal?

The reason is similar to why an extremal point of a function $y(x)$ has $dy/dx = 0$ in calculus. Let's say if the extremal is a local minimal. Let's consider the change of $y(x)$ within $(x, x + dx)$, keeping $y(x)$ outside this region fixed (this is one allowed form of δy). Then the change will be $\delta T = (\cdots)\delta y dx$ if $\delta T \neq 0$ within $(x, x + dx)$. If $(\cdots)$ is negative, the change δy makes T smaller and thus T is not minimal. If $(\cdots)$ is positive, the change $-\delta y$ makes T smaller and thus T is not minimal. Thus, for minimal T, δT vanishes within $(x, x + dx)$. Since $(x, x+dx)$ is a general interval, δT should vanish everywhere. The same argument applies for local extrema.

The above seems dry and empty. Let's consider an example.

Light propagation in the vacuum

We study light freely propagating in the vacuum between (x_P, y_P) and (x_Q, y_Q). Suppose we only know the Fermat principle, and that the speed of light is c. For simplicity, we suppress the z direction and only study x, y spatial dimensions. The propagation time T between this two points can be written as

$$T = \int_{t_P}^{t_Q} dt = \frac{1}{c} \int_{x=x_P, y=y_P}^{x=x_Q, y=y_Q} \sqrt{dx^2 + dy^2} = \frac{1}{c} \int_{x_P}^{x_Q} \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx \ . \tag{2}$$

To find the condition that an extremal path $y(x)$ must satisfy, we vary $y(x) \to y(x) + \delta y(x)$ and insert it into the above equation. As $d\delta y/dx$ is small, we can consider it as a small parameter and do Taylor expansion:

$$\sqrt{1 + \left(\frac{d[y(x) + \delta y(x)]}{dx}\right)^2} = \sqrt{1 + (y')^2} + \frac{y'}{\sqrt{1 + (y')^2}} \frac{d\delta y}{dx} + \mathcal{O}[(\delta y)^2] \ , \tag{3}$$

where $y' \equiv dy(x)/dx$. Thus

$$\begin{aligned} \delta T[y] &= \frac{1}{c} \int_{x_P}^{x_Q} \frac{y'}{\sqrt{1 + (y')^2}} \frac{d\delta y}{dx}\, dx \\ &= \frac{1}{c} \int_{x_P}^{x_Q} \left\{ \frac{d}{dx}\left(\frac{y'}{\sqrt{1 + (y')^2}} \delta y\right) - \delta y \frac{d}{dx}\left(\frac{y'}{\sqrt{1 + (y')^2}}\right) \right\} dx \\ &= -\frac{1}{c} \int_{x_P}^{x_Q} \delta y \frac{d}{dx}\left(\frac{y'}{\sqrt{1 + (y')^2}}\right) dx \ , \end{aligned} \tag{4}$$

where we have dropped the total derivative term because $\delta y = 0$ on the boundary.

$\delta T = 0$ requires that the above equation vanishes for all δy. Thus we require

$$\frac{d}{dx}\left(\frac{y'}{\sqrt{1 + (y')^2}}\right) = 0 \quad \to \quad y' = \text{const} \ . \tag{5}$$

We have thus proven that light travels with straight lines from the Fermat's principle.

With more efforts, one can also derive the light trajectory for the reflection and refraction cases using Fermat's principle and variations. But we shall not do it here.

Here we are doing something trivial using complicated methods. But the techniques (math) developed here will be reused in the next section for the action principle.

2 Principle of Extremal Action

One may live a life in two ways: (1) Given the current state, this is what I'd like to do now. And I will evolve with time and see what future will become; and (2) Given the current state, I have a clear dream of my future. And I will find out the master plan where my dream can come true.

Interestingly, physics can also be understood in these two ways (and they are equivalent): (1) corresponds to the Cauchy problem – solving the equation of motion with initial position and velocity; and (2) corresponds to the action principle, which we are now to introduce.

The action principle

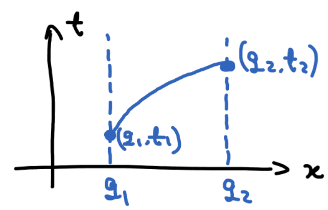
A theory is defined by an action S. The equation of motion of the theory corresponds to the extremal action $\delta S = 0$.

Let us first see how this works in Newtonian mechanics. Later we generalize it to include all known fundamental theories.

Action principle in Newtonian mechanics

Newtonian mechanics can be formulated as the following question:

Given a potential $V(q)$, where $q(t)$ is the position of the particle, what is the trajectory between a starting point (q_1, t_1) and a final point (q_2, t_2)?



First let us define an action

$$S[q] = \int_{t_1}^{t_2} dt \left[\frac{1}{2}m\dot{q}^2 - V(q)\right] . \tag{6}$$

The action has a form of integrating $K - V$ over time, where K is the kinetic energy and V is the potential energy. This is quite general. Do not ask about the physical meaning at the moment. We just define such a functional (functional of paths $q(t)$) and see what it leads to. We will come back to its physical meaning (and also why the extremal action principle at all) in Section 4.

To derive the equation of motion, similarly to Section 1, we vary $q(t) \to q(t) + \delta q(t)$, which satisfies $\delta q(t_1) = \delta q(t_2) = 0$, and see how the action changes:

$$\delta S = \int_{t_1}^{t_2} dt \left[m\dot{q}\delta\dot{q} - \frac{dV}{dq}\delta q\right] = \int_{t_1}^{t_2} dt \left[\frac{d}{dt}(m\dot{q}\delta q) - m\ddot{q}\delta q - \frac{dV}{dq}\delta q\right] . \tag{7}$$

Again the first term is a boundary term that vanishes. The last two terms holds for all $\delta q(t)$ and thus we have re-derived Newtonian equation of motion for a particle:

$$m\ddot{q} + \frac{dV}{dq} = 0 . \tag{8}$$

A remark on integral convention

Physicists sometimes write $\int dt f(t)$. This is identical to $\int f(t)dt$. It is just a matter of notation. But it reflects that physicists tend to think an integral as a limit of summation: $\lim_{\Delta t\to 0}\sum_i \Delta t f(t_i) = \int dt f(t)$. Here Δt and $f(t_i)$ are just multiplied together and thus the order is not important.

Motion with constraints

By now, you should be able to solve the motion with constraints, for example, double pendulum in a smarter way than Newtonian analysis of forces. As an exercise, you may use two angles to parameterize K and V of the system and write $S = \int dt(K - V)$. Variation principle gives you much simpler result than what you would have calaulated using forces.

A General action and the Euler-Lagrange equation

In general, consider a "Lagrangian" $L = L(q_i, \dot{q}_i, t)$, where $i = 1, 2, \ldots, N$. Here q_i denotes the position of the i-th particle. (In fact, in the spirit of general coordinate, the index i can also collectively denote many possible things: different spatial dimensions, different particles, values of fields at a point, and so on. We will not dive into those details.)

The action is defined as

$$S = \int L(q_i, \dot{q}_i, t) dt \ . \tag{9}$$

Here we have not written down the limits of the integral, having in mind that the boundary terms will be dropped. Clearly, this definition includes the Newtonian particle example that we have studied, and is much more general.

The equation of motion of such a general system, known as the Euler-Lagrangian equation, can be derived by the variation

$$\begin{aligned} \delta S &= \sum_i \int \left(\frac{\partial L}{\partial \dot{q}_i} \delta \dot{q}_i + \frac{\partial L}{\partial q_i} \delta q_i \right) dt \\ &= \sum_i \int \left[\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_i} \delta q_i \right) - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_i} \right) \delta q_i + \frac{\partial L}{\partial q_i} \delta q_i \right] dt \ . \end{aligned} \tag{10}$$

Thus the Euler-Lagrangian equation (holds for every i) can be read off as

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_i} \right) - \frac{\partial L}{\partial q_i} = 0 \ . \tag{11}$$

Math of the Lagrangian

By writing $L = L(q_i, \dot{q}_i, t)$, it is in the sense of multi-variable calculus: one can calculate partial derivatives on the three variables "independently" assuming the rest two variables do not change. For example, $\partial_t L \equiv \partial L / \partial t$ assumes that q_i and $\dot{q}_i$ do not change when calculating the partial derivative. This is different from $dL/dt = \sum_i (\partial L/\partial q_i)(dq_i/dt) + \sum_i (\partial L/\partial \dot{q}_i)(d\dot{q}_i/dt) + \partial L/\partial t$.

Dropping the boundary temrs

From now on we will assume that the boundary terms $\int \partial_t (\cdots) dt$ can be dropped. The argument is that this term does not modify the EoM. A careful treatment of the boundary terms is beyond the scope of this lecture; but dropping these boundary terms doesn't hurt for all our current purposes.

All known fundamental physics in one line

All known fundamental physics can be written into an action

$$S \sim \int d^4 x \sqrt{-g} \Big\{ \frac{1}{16\pi G} R - \frac{1}{4} F^2 + i \bar{\psi} \not{D} \psi + |Dh|^2 - V(h) + h \bar{\psi} \psi \Big\} \ . \tag{12}$$

This action covers the laws of gravity (the metric g), E&M and their friends (gauge field in F), electrons and their friends (spinor fields represented by ψ) and origin of mass (the Higgs field h). These fields, put together with certain gauge symmetries, is known as the particle physics Standard Model and describes all known fundamental physics. We shall not explain it here (a course of particle physics or quantum field theory needed). Take it as a piece of art at the moment.

3 Symmetry and Conservation Laws

In the part of special relativity, at relativistic momentum and energy, we have argued why we need conservation laws: (1) physically, allow us to ignore the details happened in the middle; (2) mathematically, reduce 2nd order ODEs into 1st order or 0th order; (3) provide observables for modern physics.

But the question left is, why there exists conservation laws after all? To ask the question in general, we had better to use a universal framework of physical theories to find the root of conservation laws. Thus the general action (9) is a good starting point.

Instead of directly asking why anything is conserved in (9), let's first think about a piece of beauty – symmetry of a theory.

A symmetry of a theory

A symmetry of a theory is a transformation under which the theory does not change.

In terms of (9), the above definition of symmetry can be put more explicitly as: Consider an infinitesimal transformation

$$q_i \to \tilde{q}_i = q_i + \epsilon \delta q_i \ , \tag{13}$$

where ϵ is an infinitesimal constant. If the action does not change

$$\delta S = \int L(\tilde{q}_i, \dot{\tilde{q}}_i, t) dt - \int L(q_i, \dot{q}_i, t) dt = 0 \ , \tag{14}$$

then we consider the transformation (13) a symmetry.

The definitions (13) and (14) are very abstract. We thus consider a few examples:

Examples of transformations

- Time translation, to test if the prediction of a theory is time dependent. The transformation thus relate two situations: doing an experiment now and doing an experiment a bit later (within the same theory) and observe if there is any difference. The time translation can thus be written as

$$q_i(t) \to \tilde{q}_i(t) = q_i(t + \epsilon\delta t) = q_i(t) + \epsilon \dot{q}_i(t) \delta t \ . \tag{15}$$

We thus extract that for time translation

$$\delta q_i = \dot{q}_i(t) \delta t \ . \tag{16}$$

We will later study the consequence of time translation in great detail in this section.

- Space translation, to test if doing an experiment in one place is identical to doing an experiment in a slightly different location. The transformation is thus $q_i(t) \to q_i(t) + \epsilon \delta q_i$, where q_i is a constant shift. The space translation can be related to momentum conservation.
- Lorentz transformation. The boost between (t_B, x_B) and (t_A, x_A) for infinitesimal β (or rotation). When β is small, $\gamma \approx 1$. Then approximately $t_B \simeq t_A + \epsilon v x_A / c^2$, $x_B \simeq x_A + \epsilon v t_A$. We thus have $q_i(t) \to q_i(t + \epsilon v q_i / c^2) + \epsilon v t$ This boost part can be related to (not very interestingly) the initial center of energy. The rotation part of Lorentz transformation can be defined similarly and is related to angular momentum conservation
- Change of phase. If q_i are complex, it makes sense to examine the transformation $q_i \to e^{i\epsilon\alpha} q_i$. This can be related to charge conservation.

We have studied many transformations. Let's move on to explore the requirement for a transformation to be a symmetry. We only take time translation as an example.

When is time translation a symmetry?

Intuitively, if the theory as specified by the Lagrangian L does not explicitly depend on t (i.e. $L(q_i, \dot{q}_i, t) = L(q_i, \dot{q}_i)$, without explicit t dependence), the theory is time translation invariant. Here we will test that in this case, the time translation is indeed a symmetry.

Remarks about symmetry

- The action is a means to derive equations of motion. Thus to examine the change of the action under (13), any equation of motion must not be used (to avoid circular arguments). To be explicit, when talking about $\delta S = 0$, we may mean one of the two things: (1) a symmetry transformation without using equation of motion; or (2) any transformation (may not be a symmetry) as a variation principle, and use the equation of motion. One should be clear about the difference.
- Again we assume to drop boundary terms $\int \partial_t(\cdots) dt$ when considering the change of the action. This is to say, if the Lagrangian $L(\tilde{q}_i, \dot{\tilde{q}}_i, t) \neq L(q_i, \dot{q}_i, t)$, but rather $L(\tilde{q}_i, \dot{\tilde{q}}_i, t) = L(q_i, \dot{q}_i, t) - \epsilon dg/dt$ for arbitrary g, the transformation is still considered a symmetry.
- As ϵ is infinitesimal, we will ignore the $\mathcal{O}(\epsilon^2)$ terms.
- The type of symmetry (13) that we study here is a continuous and global symmetry. There are other types of symmetries, namely discrete symmetries and gauge symmetries, which will not be useful to derive nontrivial conservation laws.

For space translation, is δq_i the same for different i? We will need to examine what i actually means then: If i stands for diffferent directions in space, then δq_i can take different values. If i stands for different particles in the same direction, then δq_i must take the same value for different i.

Note that $L = L(q_i, \dot{q}_i)$ is a sufficient condition of time translation symmetry, but not a necessary condition. We will not explore the necessary condition here.

Under time translation, considering that ϵ is a constant, we have

$$\delta q_i = \dot{q}_i \delta t \ , \qquad \delta \dot{q}_i = \ddot{q}_i \delta t \ . \tag{17}$$

$$\delta L = L(q_i + \epsilon \dot{q}_i \delta t, \dot{q}_i + \epsilon \ddot{q}_i \delta t) - L(q_i, \dot{q}_i) = \left[\frac{\partial L}{\partial q_i} \dot{q}_i + \frac{\partial L}{\partial \dot{q}_i} \ddot{q}_i \right] \epsilon \delta t = \epsilon \frac{d(L \delta t)}{dt} \ . \tag{18}$$

When ϵ is a constant (as we already defined), this is indeed a symmetry. To see that, note that a total derivative is a boundary term (which we neglect) in the action. Thus (14) holds.

We have inserted examples above to make things not too dry. But now let us back to very general discussions following equations (13) and (14). They are not limited to time translation but in general for any symmetry.

Conservation laws from symmetries

Bear with me a mathematical trick: We have defined ϵ as a constant. Now, let's vary it: $\epsilon = \epsilon(t)$. A symmetry transformation keeps the action invariant when $\epsilon =$ const. Now how it should change when $\epsilon = \epsilon(t)$? The change of action should now take the form

$$\delta S = \int (P\epsilon + Q\dot{\epsilon}) dt = \int Q\dot{\epsilon} dt = -\int \dot{Q} \epsilon dt + (\text{neglected boundary terms}) \ . \tag{19}$$

Here the P term vanishes because a symmetry requires this term to vanish when taking $\epsilon =$ const. There are no terms such as $\ddot{\epsilon}$ because $L = L(q, \dot{q}, t)$ and there is no $\ddot{q}$ to generate $\ddot{\epsilon}$. At the last step integration by parts is used.

Now we are ready to interpret the physical meaning of (19). If we allow to use equations of motion, that is to say, $\delta S = 0$ for all possible $\epsilon(t)$ (not because of a symmetry, but because of the action principle). Thus we have $\dot{Q} = 0$. In other words, Q is a conserved quantity when equations of motion are used.

This correspondence between a symmetry and a conserved quantity is known as the Noether's theorem.

When can EoM be used

We did not allow to use the equations of motion (EoM) when testing if a transformation a symmetry. But now, for requiring a conserved quantity to conserve, the equation of motion can indeed be used. For example, energy conservation indeed needs Newton's 2nd law (or relativistic generalizations) to apply. If a particle accelerates freely as it likes, it does not conserve energy.

If you really want to be careful about the boundary terms of the action, here you can choose $\epsilon(t)$ being a function that vanishes on the initial and final boundaries and thus we can indeed get rid of boundary terms.

In the above, we have shown the existence of a conserved quantity without showing actually what it is. This is not our style (unless you are a mathematician). What is the form of a conserved quantity given L and δq?

What is the conserved quantity in general?

We have shown that, a symmetry leaves the action invariant and thus change the Lagrangian by at most a total derivative: $\delta L = -\epsilon dg/dt$ for some quantity g. When $\epsilon = \epsilon(t)$, we cannot drop this term. Also, the change of allowing $\epsilon = \epsilon(t)$ adds a term in the following step:

$$\delta L = \sum_i \left[\frac{\partial L}{\partial q_i} \epsilon \delta q_i + \frac{\partial L}{\partial \dot{q}_i} \partial_t (\epsilon \delta q_i) \right] \supset \sum_i \frac{\partial L}{\partial \dot{q}_i} \delta q_i \dot{\epsilon} \ . \tag{20}$$

They are all the new terms that $\epsilon = \epsilon(t)$ brings and thus

$$\delta S = \int \left(g + \sum_i \frac{\partial L}{\partial \dot{q}_i} \delta q_i \right) \dot{\epsilon} dt \ , \tag{21}$$

where we have performed an integration by part to the g-term. Compare with (19), we have the conserved quantity

$$Q = g + \sum_i \frac{\partial L}{\partial \dot{q}_i} \delta q_i \ . \tag{22}$$

When the equation of motion is used, $\dot{Q} = 0$ (conservation law).

What is (22)? It differs symmetry by symmetry. Let's see one example: time translation in Newtonian mechanics.

Example: energy conservation as a result of time translation symmetry

Under time translation, the change of Lagrangian is (18). Thus, $g = -L\delta t$. The conserved quantity is the "Hamiltonian"

$$H \equiv \frac{Q}{\delta t} = \sum_i \frac{\partial L}{\partial \dot{q}_i} \dot{q}_i - L \ . \tag{23}$$

What's this? Consider more explicitly

$$L = \sum_i \frac{1}{2} \dot{q}_i^2 - V(q_1, \ldots, q_n) \ . \tag{24}$$

The conserved quantity is thus

$$H = \sum_i \frac{1}{2} \dot{q}_i^2 + V \ . \tag{25}$$

This conserved Hamiltonian is indeed the energy of the system.

When is conservation broken?

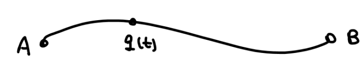
The Noether theorem does not only tell what is the conserved quantity, but also tell explicitly the quantity is conserved under which situation. In the past we state that in an isolated system energy is conserved. Now we know that

- The condition can be relaxed: even the system is not isolated, as long as the time translation symmetry is not broken by the environment, energy is still conserved. (For example, in the case with a fixed gravitational potential, consider the source of gravity, such as the earth to be outside the system.)
- If the system is time-dependent, even if the system is isolated, energy may not be conserved. An example is our expanding universe.

4 The Hidden Quantum Reality

The nature seems "strange"

The action principle offers a new way of thinking about how physics works. Take for example the motion of a particle moving from A to B in a force field:



- Newton's view: The initial position and velocity are given. At each moment, the particle "feels" the force and "adjusts" its velocity according to the force. The adjusted velocity "tells" the particle how to move further.
- The action principle: The starting and end points are fixed, the particle needs to "find" its way between these two points based on an extremal action.

Do you feel the action way "stranger"? It's straightforward for a particle to "adjust" its velocity in a force field (Newton's view). However, a particle cannot calculate (seriously in the theory of computation, since it may not even carry a bit of information or complicated enough for being even one logical gate). So how a particle can actually "follow" the action principle?

Moreover, as we argued the action principle is a more fundamental way to describe all laws of nature. Why nature behaves fundamentally in such a "strange" way?

The nature is natural but quantum

No. The nature is not strange. The nature is natural, but just not classical.

Let's return to the question of the particle motion. Instead of "calculating" the extremal action, what the particle actually does is that, it "tries" *all possible paths* and "take" the extremal one.

The above words are actually nothing but the path integral formulation of quantum mechanics. In quantum mechanics, the probability P for things to happen (here: particle to propagate from A to B) is decomposed as

$$P = |\mathcal{A}|^2 \ , \tag{26}$$

where the complex number $\mathcal{A}$ is known as the probability amplitude. The probability amplitude can be calculated by a weighted average over all paths

$$\mathcal{A} \propto \sum_{\text{all paths}} e^{iS/\hbar} \ . \tag{27}$$

Here by "all paths" we mean all possible lines connecting A and B, not necessarily satisfy the equation of motion.

You may think about a similar question: Why in a circuit, the electric current "knows" which way to go to minimize power? No, the current doesn't know. Rather, the electric field build up all possible paths until reach a stationary situation. More obvious examples include why flood "knows" where to flow, and so on. Definitely they are classical. But a particle can rely on its quantum nature to achieve a similar feature.

The action principle explained

In the quantum world, the physical meaning of the action is clear: $e^{iS/\hbar}$ is the phase factor as a weight of the path in the summation of all paths.

You may be confused here: in classical mechanics, the particle only select one trajectory. In quantum mechanics (path integral formulation), the particle move along all possible paths together. How to reconcile the difference? How the classical trajectory emerge among the quantum trajectories?

Classical mechanics emerges from quantum mechanics when $S \gg \hbar$. In this limit, even a very small change in S (due to choosing a nearby different trajectory) results in a large change in the unit $\hbar$, and thus $e^{iS/\hbar}$ is a fast oscillating function, which cancels contributions of different paths almost for all paths.

There is only one exception: close to the stationary action $\delta S = 0$. Near $\delta S = 0$ different paths do not cancel. Thus the classical trajectory is the only trajectory left over when the classical limit $S \gg \hbar$ is taken.

What if $S \sim \hbar$?

What about $S \sim \hbar = 1.05 \times 10^{-34}\ \mathrm{m^2kg/s}$? Let us jump into the brave new world of quantum mechanics to answer this question in the next part! But in fact, we will use different but equivalent formulations of quantum mechanics (mainly wave mechanics and a bit of matrix mechanics) instead of the path integral. This is because the path integral, though conceptually simple, is harder in many computations. It will be included in a full quantum mechanics course.

5 Epilogue: Summary and What's Next

Further reading about the content

- Similar contents in other textbooks: You can find an extensive discussion of the content here in Part II of Einstein Gravity in a Nutshell by Zee. You may also read the first section and the final appendix of Lecture Notes on Modern Physics by Baumann.
- If you want even more references, I recommend to watch Theoretical Minimum (Video Lectures) (Lectures 3 and 4 of Classical Mechanics) by Susskind; or to read Chapter 2 of Classical Mechanics by Goldstein, Poole Jr and Safko.

What happens next in a university physics program?

- Classical mechanics. The action principle is closely related to the Lagrangian formulation of classical mechanics. There is also a Hamiltonian formulation. You will see how mechanics are written in these ways. You will learn how Lagrangian mechanics is helpful in solving constrained motion problems.
- Quantum mechanics. The path integral mentioned here will be part of quantum mechanics (or sometimes advanced quantum mechanics). The (quantized) Hamiltonian is also a centural part of quantum mechanics.
- Quantum field theory. A model in quantum field theory starts with an action. You will fully see there that the action principle is indeed considered as the first principle.

6 Exercises

E1.1 Refraction from the Fermat's principle

Derive the law of refraction from the Fermat's principle.

E1.2 Extremal but non-minimal paths

Find examples that light travel along extremal, but not minimal paths.

E2.1 A cosmological scalar field

In cosmology, a homogeneous and isotropic scalar field has action

$$S = \int dt\ a^3(t) \left(\frac{1}{2}\dot{\phi}^2 - V(\phi) \right) ,$$

where $a(t)$ is a function of time, and $V(\phi)$ is a function of ϕ. Calculate explicitly the Euler-Lagrange equation of ϕ (i.e. relation between $\ddot{\phi}$, $\dot{\phi}$ and ϕ) in two ways: The Euler-Lagrange equation and the variation principle, respectively.

E3.1 From Euler-Lagrange to Newton

Start from the Euler-Lagrange equation, use the Lagrangian of a particle in Newtonian mechanics, to derive the Newtonian second law for particle motion in a potential.

E3.2 A relativistic free particle

The action for a freely moving (i.e. not moving in a force field or interaction with other particles) relativistic particle is

$$S = \alpha \int d\tau = \alpha \int \sqrt{1 - \frac{\dot{q}^2}{c^2}}\ dt\ , \tag{28}$$

where α is a constant.

- Determine the value of α by taking the Newtonian limit $\dot{q} \ll c$.
- Show that the system has time translation symmetry.
- Derive the relativistic energy as the conserved quantity of time translation.

Index