

We begin with the *Action S*, which is what both Lagrange and Hamilton were after. It is essentially a kind of master functional (a function of functions), from which the equations of motion for any reasonable classical system may be derived. Lagrange defined it as such:

$$S = \int_{t_1}^{t_2} L(t, x(t), \dot{x}(t)) dt$$

Where L is of course the *Lagrangian*.

Conceptually speaking, the physical quantity “action” isn’t far off from the colloquial interpretation: It’s a way of quantifying events – anything from a ball rolling down a hill to a proton decaying into a neutron $p \rightarrow n + e^+ + \nu_e$. In physics, the action is a scalar quantity, meaning it’s just a number, with dimensionality [*energy * time*]. If we undo the integral, by deriving both sides with respect to time, we see

$$\frac{d}{dt} S = L^*$$

So the Lagrangian then can be interpreted as the infinitesimal action – the ‘building-blocks’ for an action, if you will. Technically speaking, both S and L encode the dynamics of the system for all times, but only when properly constrained will they reflect reality – as we’ll see. The dimensionality of the Lagrangian then is simply energy.

The reason why Lagrange, Newton, many others were so interested in the action goes back to Fermat, 17th century, who noticed light always moves in straight lines. Eventually this principle was generalized from “*path of least time*” or “*path of least resistance*” to the “*principle of least action*”.

Basically, the idea is that nature never seems to go out of its way, never wastes energy, is always optimally efficient, somehow ‘knows’ the easiest way to do anything... etc. Personally, I think it’s all a tautology, subsequent to how we define and update the “laws of physics” in the first place, but that’s irrelevant. What’s relevant here is that given this principle, to predict how nature will unfold in any reasonable situation, we simply need to *minimize* the action functional S .

So we’ll do just that, we’ll find the minimum of the action, which should spill out a general equation for kinematics. Thanks to *The Calculus of Variations* (which sounds fancy because it is fancy – it’s the go-to mathematics for optimization), this derivation isn’t so bad. I’ve included a reasonably thorough derivation at the end.

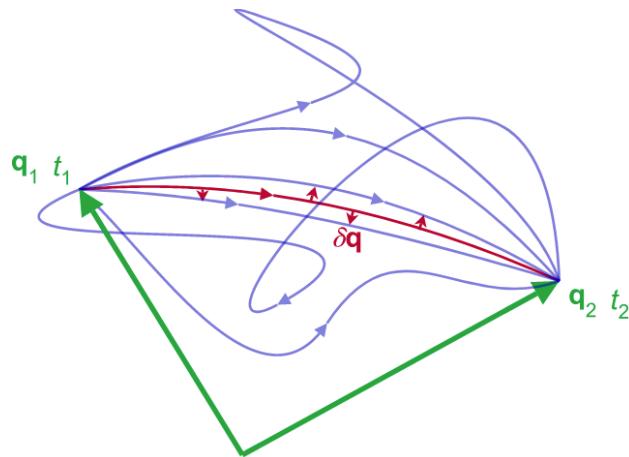
For now, it's sufficient to take my word for the fact that the *Euler-Lagrange Equation* is the result:

$$\frac{\partial L}{\partial x} = \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}} \right)$$

To summate, if we want to find the min or max of the action S , which we can understand as the time integral of the Lagrangian L , we must find a symmetry between the variables of L , which constrains the result. To do this (see derivation), we essentially force it to obey our command, and the Euler-Lagrange Equation is the result.

My favorite interpretation of the Action is that of a *path*. Specifically, a path through *Configuration Space* – which is simply the set of all possible arrangements a system can be in (we'll see more on this in the first example).

Consider the figure below, where multiple paths are drawn between two coordinates in space and time. The paths represent possible actions, which themselves represent possible kinematic trajectories according to the laws of classical physics.



There are many ways for an object to go from left to right. However, nature will only follow one path, the path of least action.

With math,

$$\delta S(L) = 0, \left\{ \frac{\partial L}{\partial x} = \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}} \right) \right.$$

Where δ is a “perturbation” of the action, which is akin to infinitesimal adjustments, here represented as little arrows perpendicular to the central path. This δS can be understood as how when you're at a maximum or minimum of a function, here S , the local slope is zero.

Though to be fair, we should be using generalized coordinates notation (t, q, \dot{q}) . *
This however should be implied whenever we're dealing with Lagrangian mechanics, so please consider my coordinates generalized. ☺

Here is the simplest definition of the Lagrangian, for isolated systems:
The difference in kinetic and potential energy.

$$L = T - V$$

Ideally,

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

Which implies the Lagrangian does not depend on time explicitly. * In this case we say energy is conserved. If the Lagrangian does change with time, say, because the potential does $V(x, t)$, then energy is not conserved.

While not representing the energy directly, like the Hamiltonian does, we see then that the Lagrangian is nevertheless useful, due primarily to its connections with the symmetries of physics: If $\frac{d}{dt} L = 0$, energy is conserved, and if $\frac{d}{dx} L = 0$, momentum is conserved. That is, if the Lagrangian is invariant over time, E is conserved, and if the Lagrangian is invariant over space, p is conserved. E.g., if the right-hand side is zero, i.e., the partial of L with respect to velocity does not depend on time, then the left side says the L does not change over space – therefore p is conserved.

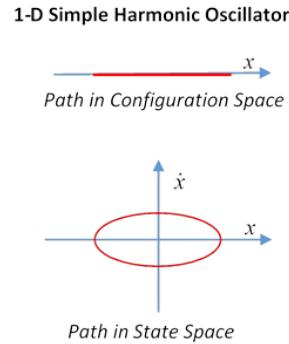
$$\frac{\partial L}{\partial x} = \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}} \right)$$

Although a mouthful, this equation can be read as: *How the Lagrangian changes over space is equivalent to how the Lagrangian changes over the slope of space, over time.*

The utility of the Lagrangian is this property. Remember, the Lagrangian is just *some* function that minimizes the action – in such a way that preserves the laws of classical physics.

Lagrangian and Hamiltonian Methods – Brief Glances

Consider our old friend the Simple Harmonic Oscillator, which we know conserves energy – unless it's driven or damped, of course, but let's just let it oscillate for now.



This can represent a mass on a spring or a pendulum, or perhaps other ideal simple systems, but let's call it a mass / spring. What's important is the ellipse – or circle, because we can always stretch it into a circle by tweaking m or k . What does the circular path represent?

The Lagrangian right? No! Whereas L is a function of these coordinates, $L(t, x(t), \dot{x}(t))$, L is not what is shown. We know this because, for a mass and spring,

$$L = \frac{1}{2}m\dot{x}^2 - \frac{1}{2}kx^2$$

Which, given the fact that $k > 0$, the right term is necessarily negative. Therefore, this is not the equation for an ellipse! Ellipses have the form $1^2 = ax^2 + by^2$ – which you should recognize as none other than a parameterized Pythagorean Theorem: the circle is generalized into an ellipse by the variables a and b .

Our equation for L however, is more like $1^2 = ax^2 - by^2$, which is the equation for a hyperbola. This is because the Lagrangian does *not* represent the energy of our system. Just looking at the equation, it is much more a function of the how the energies differ; what's left given a state of the system.

This is important. To see why let's jump ahead to the corresponding definition for the *Hamiltonian*:

$$H = T + V$$

So we can see the corresponding Hamiltonian for our mass and spring system is,

$$H = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}kx^2$$

Which looks much more like the Total Energy you're familiar with. It is not incorrect then, to say that the difference in the Lagrangian and the Hamiltonian is whether you subtract or add the potential. The distinction is more semantic than you might imagine. Consider this diagram I found from the University of Tokyo, 2016:

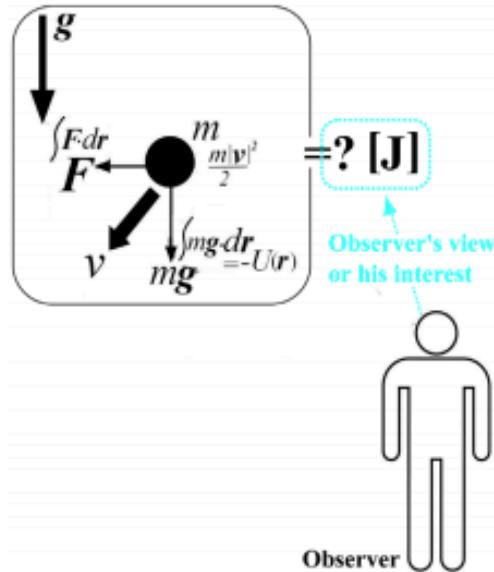


Figure 1. Illustration of a Hamiltonian viewpoint image.

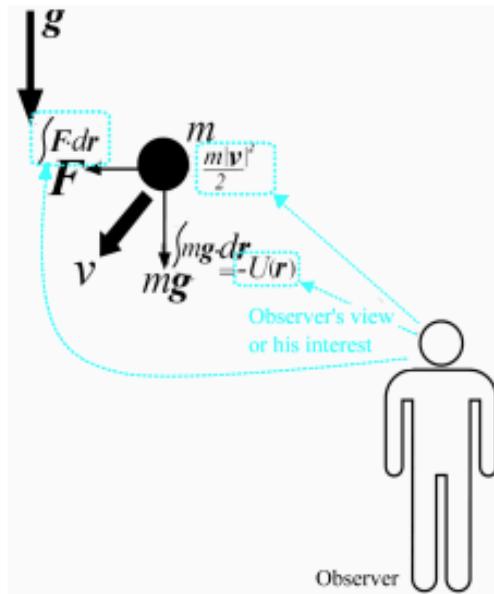


Figure 2. Illustration of a Lagrangian viewpoint image.

Where [J] is for joule, the energy. Contextually, the diagram shows that the difference in the Lagrangian and Hamiltonian is about what you're interested in.

However, that difference is quite crucial. That's because the Hamiltonian formalism is generally easier to deal with, and has been well incorporated into Quantum Mechanics. Curiously enough though, it appears the Lagrangian formalism works better with general relativity.

The benefit of the Hamiltonian over the Lagrangian is well articulated by Taylor. Disregarding time dependence, for now,

$$L(x, \dot{x})$$

$$H(x, p)$$

We see H is a function of position and momentum, instead of position and velocity. Which sounds like a triviality, until we realize H is a function of only 1st order parameters, whereas L involves a 2nd order, \dot{x} .

In general, this means for an n dimensional system, the Lagrangian approach gives n 2nd order equations of motion, whereas the Hamiltonian approach gives $2n$ 1st order equations – twice the equations, yet for exponentially less work.

Here are the 2 equations, Hamilton's Equations, corresponding to the singular Euler-Lagrange Equation.

$$\dot{q}_i = \frac{\partial H}{\partial p_i}$$

$$\dot{p}_i = -\frac{\partial H}{\partial q_i}$$

These equations represent a much tighter relationship, one which is defined in *Phase Space*, as opposed to the *State Space* of the Lagrangian approach. Phase space is essentially bigger and better than State Space. The confusion in reading the State Space earlier is indicative of this inferiority. Moreover, in the end you can represent the Hamiltonian approach by the simple equation,

$$\dot{\mathbf{z}} = \mathbf{h}(\mathbf{z})$$

Where bold z is a point in phase space, and bold z dot is a vector derivative. Therefore you're left with the form

$$\text{first derivative of } \mathbf{z} = \text{function of } \mathbf{z}$$

Which is an incredibly common form for differential equations, and therefore well known territory.

I would love to expand this, but unfortunately my time has run out. Below is a standard derivation of the Euler-Lagrange Equation, which makes for good practice.

First assume L is twice continuously differentiable.

(Please refer to the path diagram)

If $x(t)$ extremizes the functional S , then any perturbation of $x(t)$ which preserves the boundary conditions, $x(a) = A, x(b) = B$, must either increase S (if $x(t)$ is a minimizer) or decrease S (if $x(t)$ is a maximizer).

Let $g_\epsilon(t)$ be the result of such a perturbation $\epsilon\eta(x)$ on $x(t)$, thus

$$g_\epsilon(t) = x(t) + \epsilon\eta(x)$$

Now define

$$S_\epsilon = \int_a^b L_\epsilon dt$$

Where

$$L_\epsilon = L(t, g_\epsilon(t), \dot{g}_\epsilon(t)) dt$$

By evaluating $\frac{d}{d\epsilon} S_\epsilon$, we should be able to isolate our extremums.

$$\frac{d}{d\epsilon} S_\epsilon = \int_a^b \frac{d}{d\epsilon} (L_\epsilon) dt$$

Considering the integrand,

$$\frac{d}{d\epsilon} (L_\epsilon) = \frac{d}{d\epsilon} [L(t, g_\epsilon(t), \dot{g}_\epsilon(t))]$$

=

$$\frac{dt}{d\epsilon} \frac{\partial}{\partial t} + \frac{dg_\epsilon}{d\epsilon} \frac{\partial L_\epsilon}{\partial g_\epsilon} + \frac{d\dot{g}_\epsilon}{d\epsilon} \frac{\partial L_\epsilon}{\partial \dot{g}_\epsilon}$$

=

$$0 + \eta(t) \frac{\partial L_\epsilon}{\partial g_\epsilon} + \dot{\eta}(t) \frac{\partial L_\epsilon}{\partial \dot{g}_\epsilon}$$

Returning to the integral,

$$\frac{d}{d\epsilon} S_\epsilon = \int_a^b \left[\eta(t) \frac{\partial L_\epsilon}{\partial g_\epsilon} + \dot{\eta}(t) \frac{\partial L_\epsilon}{\partial \dot{g}_\epsilon} \right] dt$$

We demand this functional to be invariant under the infinitesimal perturbation ϵ , therefore its total derivative, or change, with respect to ϵ , is zero.

We can do this because we're already demanding $x(t)$ to extremize S .

$$\begin{aligned} & \Rightarrow \frac{d}{d\epsilon} S_\epsilon = 0 \\ & \Rightarrow \int_a^b \left[\eta(t) \frac{\partial L_\epsilon}{\partial g_\epsilon} + \dot{\eta}(t) \frac{\partial L_\epsilon}{\partial \dot{g}_\epsilon} \right] dt = 0 \end{aligned}$$

Another way of looking at this step is to imagine evaluating it in the case $\epsilon = 0$. This returns our original L , unperturbed. Therefore, it's fair to say,

$$\Rightarrow \left[\frac{d}{d\epsilon} S_\epsilon \right]_{\epsilon=0} = \int_a^b \left[\eta(t) \frac{\partial L}{\partial x(t)} + \dot{\eta}(t) \frac{\partial L}{\partial \dot{x}(t)} \right] dt = 0$$

Splitting the integral,

$$\begin{aligned} & \int_a^b \left[\eta(t) \frac{\partial L}{\partial x(t)} + \dot{\eta}(t) \frac{\partial L}{\partial \dot{x}(t)} \right] dt \\ & = \\ & \int_a^b \eta(t) \frac{\partial L}{\partial x(t)} dt + \int_a^b \dot{\eta}(t) \frac{\partial L}{\partial \dot{x}(t)} dt \end{aligned}$$

Consider the second term with integration by parts,

$$\int u dv = uv - \int v du$$

Where,

$$u = \frac{\partial L}{\partial \dot{x}(t)} \Rightarrow du = \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}(t)} \right) dt$$

$$dv = \frac{\dot{\eta}(t) dt}{dt} = \dot{\eta}(t) \Rightarrow v = \eta(t)$$

Substituting,

$$\int \frac{\partial L}{\partial \dot{x}(t)} \dot{\eta}(t) = \left[\frac{\partial L}{\partial \dot{x}(t)} \eta(t) \right]_a^b - \int_a^b \eta(t) \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}(t)} \right) dt$$

Adding back the first term, we can now factor out $\eta(t)$,

$$\int_a^b \eta(t) \left[\frac{\partial L}{\partial x(t)} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}(t)} \right) \right] dt + \left[\frac{\partial L}{\partial \dot{x}(t)} \eta(t) \right]_a^b$$

Here we realize $\eta(t)$ evaluated at the endpoints a and b yields zero, since we demanded the net perturbation $g_\epsilon(t)$ respect the end points of $x(t)$, i.e., $g_\epsilon(a) = x(a) = A$ and $g_\epsilon(b) = x(b) = B$. Therefore, given our definition of $g_\epsilon(t)$, we see $\eta(a) = \eta(b) = 0$.

So we're left with

$$\int_a^b \eta(t) \left[\frac{\partial L}{\partial x(t)} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}(t)} \right) \right] dt = 0$$

Finally we determine, with help of the *Fundamental Lemma of Calculus of Variations*, that the cases where $\int_a^b \eta(t) dt = 0$ are simply not important, since they represent here 'alternative paths'. The path which we demanded initially to make S stationary, is $x(t)$. Therefore, we are only concerned with cases where

$$\int_a^b \left[\frac{\partial L}{\partial x(t)} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}(t)} \right) \right] dt = 0$$

Which occurs pointwise when the value of the brackets vanishes. Hence, the Euler-Lagrange Equation.

$$\frac{\partial L}{\partial x(t)} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}(t)} \right) = 0$$

$$\frac{\partial L}{\partial x} = \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}} \right)$$

Good References: (two per line)

<https://link.springer.com/article/10.1007/s40329-014-0049-x> <http://preetum.nakkiran.org/lagrange.html>
https://en.wikipedia.org/wiki/Principle_of_least_action http://www.feynmanlectures.caltech.edu/II_19.html
https://en.wikipedia.org/wiki/Variational_principle https://file.scirp.org/pdf/WJM_2016032920002016.pdf