These are my notes on many explanations of principles used in Physics. These ideas have emerged as a culmination of what I've learned in lectures and read/watched. I credit the first volume of Leonard Susskind's book series: *The Theoretical Minimum* for the derivation of the conservation of momentum, and I credit Dr. Jorge S Diaz's modern translation of the derivation of the principle of least action by William Rowan Hamilton for the section in which its derivation is presented. This isn't by any means a research paper, just my thoughts on what I felt was an effective way to study these topics.

A Historical Note and My Prior Confusion

Many scientists, engineers, and mathematicians use a tool called the *calculus of variations*, capable of simplifying a multitude of complex problems by determining when a special quantity is at a minimum. This quantity goes by many names depending on your field, but in physics, this quantity is given the name action. Action was first introduced by the French mathematician and philosopher Pierre Louis Maupertuis in an attempt to unify Isaac Newton's mechanics and geometric optics.[1]

At first, Maupertuis' notion of action was thought of by his critics as a non-physical, meaningless quantity, which he claimed to be even more fundamental than the concept of force introduced by Newton, as it was generally capable of describing phenomena in optics in addition to mechanics. Later, the study of action was generalized by the Swiss and French polymaths Leonhard Euler and Joseph-Louis Lagrange. To make matters worse, other mathematicians like Leibniz had been toying with ideas similar to Maupertuis' earlier on, and disliked the fact that the principle was credited to him. Euler framed the groundwork for the tools used in calculus of variations, working with the young Lagrange, who applied the techniques to Physics and the study of action ten years later.

In retrospect, Maupertuis' introduction to action was quite informal, stating that the minimization of action in physical systems (now called the principle of least action) is just another fundamental law. Given the context that he was working in that inspired his pursuit, Maupertuis really couldn't say much more. When I first learned about the principle of least action, I had many questions, namely:

- Where does the definition of action come from?
- Why is action minimized?
- Why does almost every undergraduate resource on mechanics never rigorously prove the principle of least action?

My goal here is to answer every question except the last. Before such questions can be answered, it is first important to discuss what laws are truly fundamental, and how quantities similar to action, like momentum and energy, arise from such fundamental laws, as this is often glossed

[1] Optics is the study of the behavior of light. Geometric optics specifically theorizes that light is a uniform ray (of particles), in contrast to wave optics, which theorizes that light exhibits wavelike properties. In the end, wave optics was proven superior, as it is capable of describing a broader range of physical phenomena. On the other hand, mechanics is the study of motion and interactions of matter, describing almost all everyday objects.

over in undergraduate texts, but not as badly as action. However, these notes also may change the way you think about energy and momentum as *derived* quantities, rather than fundamental objects. Discussing the fundamentals is a difficult pursuit, but constructing a valid theory out of fundamentals has already done in part by Newton, as we will see.

What's a Theory of Physics?

In broad terms, a theory of physics is a set of rules (laws) that physicists use to describe an appropriate set of phenomena. Some laws within a theory are more fundamental than others. We will often refer to laws that are the most fundamental as axioms, as there is often no conceivable way of reasoning as to why these laws are the way they are; they just happen to be good at quantifying what the theory is aiming to describe.

Unsurprisingly, every theory of physics has an analogous principle of least action. In fact, the physics media often describes the principle of least action as 'fundamental' for all theories of physics. This can be quite misleading for two reasons: some theories of physics (called non-Lagrangian theories) do not have a general principle of least action at all; in addition, each principle of least action is actually a consequence of a set of pre-existing axioms for that specific theory. Thus, principles of least action are not themselves a law, but rather a consequence of the mathematical structure of many, but not all, theories. Utilizing this structure is incredibly useful in theoretical physics, making it worthwhile to examine the structural meaning of the principle of least action (among other derived principles), and demonstrating that it truly comes from the axioms of the theory.

Laws of Mechanics

Now that we know what actually qualifies as a fundamental law (axiom) of a theory of physics, it is worth stating what these are in detail for Newtonian mechanics. Some are likely already familiar to you, and are stated as follows:

- (*Newton's 1^{st} Law*) The basic objects of the universe are particles; the location of every particle can be quantified by a coordinate vector at a time t, $\mathbf{x}(t) = (x_1(t), x_2(t), x_3(t))$. There always exists at least one way of describing this position such that $\ddot{\mathbf{x}}(t) = \mathbf{0}$ for every non-interacting particle in the universe. This description is called an inertial frame.
- (*Newton's 2^{nd} Law*) Forces are fundamental quantities that indicate interactions, depending only on the relative positions and properties of the particles involved. For the i^{th} particle in the universe with position $\mathbf{x}_i(t)$ in an inertial frame:

  $$\sum_j \mathbf{F}_{ji}(\mathbf{x}_i) = m_i \ddot{\mathbf{x}}_i. \tag{1}$$

  where $\mathbf{F}_{ji}$ is a force acting on the i^{th} particle caused by the j^{th} particle, and m_i is a proportionality constant between the force and the second time derivative of the position, called the mass.
- (*Strong Form of Newton's 3^{rd} Law*) An additional property of forces is that in an inertial frame:[2]

  $$-\mathbf{F}_{ji} = \mathbf{F}_{ij} \parallel (\mathbf{x}_i - \mathbf{x}_j). \tag{2}$$

[2] You may argue that this is not true, as magnetic forces may act perpendicular to the direction of the object

- (*Isotropy and Homogeneity of Spacetime*) Any individual force law is not allowed to depend on the location/time of a system of objects or the direction at which they are separated, only their relative displacements within the system.

For the second law, recall that $\mathbf{F}_{ji} = \mathbf{0}$ if the j^{th} particle is non-interacting. Additionally, if every force satisfies something called *Lipschitz continuity* as a function of the i^{th} position, then we can demonstrate that the position and velocity at any point in time dictate the position and velocity for every other point and time of a particular particle (i.e., the Newtonian world is deterministic). This is called the Picard–Lindelöf theorem, whose discussion is beyond the scope of this derivation.

With these assumptions, we can, in theory, deduce every single equation of motion within Newtonian mechanics, and thereby predict any system given its position. Some forces, often referred to as phenomenological forces, like friction and tension, may break the axioms at a cursory glance. For example, friction depends on the velocity at which you are moving, not just your position. This occurs because the total number of interactions between the very large number of particles (often of the order of 10^{26}) has a net effect on the system that is undergoing friction that causes it to appear velocity-dependent in approximation. Techniques that provide theoretical treatment to such forces are a part of *Statistical Mechanics*, and are also capable of describing lots of other phenomena, like temperature, fluids, and heat flow.

Momentum

Now we will see how our new laws introduce some subtle underlying structure. First off, take a look at the first part of the third law. It states that

$$-\mathbf{F}_{ji} = \mathbf{F}_{ij} \tag{3}$$

for every particle interacting in the universe. These forces are vectors, and may be added via vector addition per usual, so if we consider the sum over every existing force:

$$\mathbf{F}_{\text{total}} = \sum_{(i,j)} \mathbf{F}_{ij}. \tag{4}$$

Since $\mathbf{F}_{ii} = -\mathbf{F}_{ii}$, $\mathbf{F}_{ii} = 0$. Thus we may get rid of every pair where $i = j$. Additionally, if we consider any arbitrary pair (i, j) where $i \neq j$, there must also exist a force $\mathbf{F}_{ji} = -\mathbf{F}_{ij}$ by Newton's third law. Thus we can split the sum into the above cases and perform the following:

$$\mathbf{F}_{\text{total}} = \sum_{i} \underbrace{\mathbf{F}_{ii}}_{=0} + \sum_{\{i,j\}} \mathbf{F}_{ij} + \sum_{\{i,j\}} \mathbf{F}_{ji} = \sum_{\{i,j\}} \mathbf{F}_{ij} + \sum_{\{i,j\}} (-\mathbf{F}_{ij}) = \mathbf{0}. \tag{5}$$

Here, we have used the curly braces to denote an unordered pair (set) of indices rather than an ordered pair. We have just proven something pretty interesting, that the total force in the universe is zero. Oftentimes, there are particles that don't interact with each other very much at a distance (consider neutrons), and so we may suspect that there are also special cases in which we have a

producing them, but magnetic forces are actually a consequence of special relativity, and are outside of Newton's mechanics. Within special relativity, the strong form of Newton's 3^{rd} law still holds given the added requirement that the inertial frame is specifically a *rest frame* of the object producing the force. Thus, magnetism is somewhat similar to fictitious forces like the centrifugal force in the sense that its existence depends on which coordinates you choose.

subset of interacting particles in the universe that also have zero total force. We call such systems of particles *closed systems*.

So far, it appears that Newtonian mechanics builds a theory in which there exists a special group of particles (including the whole universe) that have a special property that the total force is zero. A notable example of such a group of particles are any ordinary rigid object, like a rock. So it is worth studying what may happen to a rock as it interacts with nothing. To do this, we will employ the methods of calculus.

Recall that Newton's second law says that a consequence of a force is acceleration parallel to that force, whose relative magnitude depends on a quantity called mass (at least in an inertial reference frame). If we consider a particle with no forces acting on it that has a nonzero mass,[3], we can conclude that:

$$m\ddot{\mathbf{x}} = \mathbf{0}. \tag{6}$$

We also know from calculus that

$$\ddot{\mathbf{x}} = \frac{\mathrm{d}\dot{\mathbf{x}}}{\mathrm{d}t}, \tag{7}$$

and consequently,

$$\frac{\mathrm{d}}{\mathrm{d}t}(m\dot{\mathbf{x}}) = \mathbf{0}. \tag{8}$$

Therefore, zero force additionally implies zero derivative in this interesting mathematical quantity involving m and $\dot{\mathbf{x}}$. We can integrate both sides of the above equation, leaving us with an integration constant which we will call $\mathbf{p}$:

$$m\dot{\mathbf{x}} = \mathbf{p}. \tag{9}$$

We call this integration constant the *momentum* of the particle, as it does not change in time. Many people consider momentum to be an intuitive physical quantity, but I feel it is quite the opposite. Momentum is a quantity that arises through mathematical symmetry of Newton's laws, relating to there being a 2^{nd} derivative in his second law. Acceleration, being in his second law, is not a fundamental truth; it is merely a choice of derivative that appears to reconstruct the mechanics of the universe well. Although intuitively obvious for one particle, this demonstration of conservation of momentum becomes less trivial and more interesting when we consider a system of particles.

If a system of particles has a total force $\mathbf{F}_{\text{total}}$, then by Newton's second law, we can set this equal to the sum of the products of masses and accelerations as follows:

$$\mathbf{F}_{\text{total}} = \sum_i m_i \ddot{\mathbf{x}}_i. \tag{10}$$

Using the linearity of the integral and repeating the steps above, we obtain:

$$\sum_i m_i \dot{\mathbf{x}}_i = \mathbf{p}_0 + \mathbf{F}_{\text{total}}. \tag{11}$$

This time, our entire right side is more than just an integration constant; it has a term involving time. However, notice that this term goes away if the vector sum of the forces within the system $\mathbf{F}_{\text{total}} = \mathbf{0}$. This is quite interesting. It tels us that our momentum is invariant with time, if the total force in the system is zero, i.e., our system is closed. This fact allows us to conclude that when there are no external interactions within a closed system like a rock, the sum over every particle in

[3] Newton's laws start to break down when a particle, like light, has zero mass

the rock's mass-velocity product is constant. This is a remarkable fact since if the rock were, say, rotating as it floats in space, none of the individual particles' velocities are non-changing, but the ensemble of these particles creates an unchanging quantity. This is the true meaning behind what momentum is.

Note that if we take the case of $\mathbf{F}_{\text{total}} = \mathbf{0}$ like above and divide both sides by the total mass (a constant), we obtain an equation involving the center of mass of the particle:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\sum_i m_i \mathbf{x}_i}{\sum_i m_i}\right) = \frac{\mathbf{p}}{\sum_i m_i}. \tag{12}$$

It becomes apparent that, similar to a single particle, any closed system may behave like one in the sense that its center of mass will move with a constant velocity equal to the momentum divided by the total mass. Additionally, another nice aspect of the above equation is that it can be easily converted into an equation for a continuous distribution of particles with the mass distributed along any measure of space.

Scalar Functions Representing Force

You may think that, after discussing momentum, we will immediately motivate the concept of energy and angular momentum to complete the picture. However, I think that it is easy to get lost in the discussion of such quantities, as you may get ideas of what these things are from prior knowledge/intuition. My goal here is to first motivate the less intuitive concept of action via the method of Lagrange and Euler, which will help us get an understanding of energy and angular momentum as an additional mathematical structure, rather than a physical quantity. However, before jumping straight into the proof of the principle of least action, it is worth discussing another consequence of forces.

Recall that we stipulated in Newton's 2^{nd} law that the function defining a force for a particle may only depend on the position of that particle and nothing else. This fact, along with some others, allows us to actually remove the necessity of describing a force with a vector every time, as we will see. Before we do this, we must prove a quick lemma:

$$(\nabla \times \mathbf{F})(\mathbf{x}) = \mathbf{0}\ \forall \mathbf{x} \implies \exists \phi(\mathbf{x}),\ \mathbf{F}(\mathbf{x}) = \nabla\phi(\mathbf{x}). \tag{13}$$

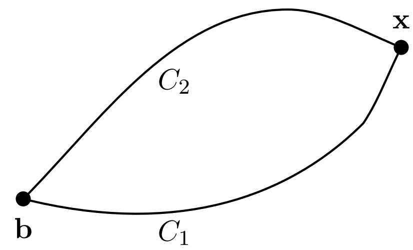
To prove this, we will construct an example of such a scalar function ϕ for every force vector function $\mathbf{F}$. If we consider the function

$$\phi(\mathbf{x}) = \int_{C:\mathbf{b}\to\mathbf{x}} \mathbf{F}\cdot \mathrm{d}\mathbf{x} \tag{14}$$

where C is an arbitrary, yet sufficiently 'nice' path from an arbitrary point $\mathbf{b}$ to the location $\mathbf{x}$. So far, we know very little about ϕ. In fact, we have no idea if changing C changes the quantity of ϕ, which would mean that ϕ is not a function of just $\mathbf{x}$. To deal with this, consider two distinct paths from $\mathbf{b} \to \mathbf{x}$ that define the line integral for ϕ called C_1 and C_2. If we call the result of defining ϕ using C_1 and C_2 respectively ϕ_1 and ϕ_2, we can compute their difference as

$$\phi_2 - \phi_1 = \int_{C_2} \mathbf{F}\cdot \mathrm{d}\mathbf{x} - \int_{C_1} \mathbf{F}\cdot \mathrm{d}\mathbf{x} = \int_{C_2\cup(-C_1)} \mathbf{F}\cdot \mathrm{d}\mathbf{x} \tag{15}$$

The reason for the far right side of the equation is that the difference can be interpreted as a line integral traveling on a closed loop along C_1 and then going backwards on C_2. Thus, we can represent



this difference as traveling along the loop $C_2 \cup (-C_1)$. Since we are physicists and we believe $\mathbf{F}$ and the paths C_1 and C_2 are 'sufficiently nice', we can employ Stokes' theorem to calculate this closed loop integral as

$$\phi_2 - \phi_1 = \int_{\text{int}(C_2 \cup (-C_1))} \underbrace{(\nabla \times \mathbf{F})}_{=\mathbf{0}} \cdot \mathrm{d}\mathbf{S} = 0. \tag{16}$$

Therefore $\phi_1 = \phi_2$, and since C_1, C_2 are arbitrary, it does not matter which path we define C on provided $\mathbf{F}$ has zero curl.

Now we must compute this function's gradient, to demonstrate that we recover the force $\mathbf{F}$. To do this, we will consider a small variation ε in the direction of the i^{th} basis vector $\mathbf{e}_i$ and compute the difference in the function ϕ under that variation. As we have just shown, we don't need to specify a path here, and the difference is

$$\phi(\mathbf{x} + \varepsilon \mathbf{e}_i) - \phi(\mathbf{x}) = \int_{\mathbf{a} \to \mathbf{x} + \varepsilon \mathbf{e}_i} \mathbf{F} \cdot \mathrm{d}\mathbf{x} - \int_{\mathbf{a} \to \mathbf{x}} \mathbf{F} \cdot \mathrm{d}\mathbf{x}. \tag{17}$$

The result of this difference is just an integral from $\mathbf{x} \to \mathbf{x} + \varepsilon \mathbf{e}_i$, which we compute by choosing our path to be a line segment between the two in the direction of $\mathbf{e}_i$.

$$\phi(\mathbf{x} + \varepsilon \mathbf{e}_i) - \phi(\mathbf{x}) = \int_{\mathbf{x} \to \mathbf{x} + \varepsilon \mathbf{e}_i} \mathbf{F} \cdot \mathrm{d}\mathbf{x} = \int_0^{\varepsilon} \mathbf{F}(\mathbf{x} + t\mathbf{e}_i) \cdot (\mathbf{e}_i \mathrm{d}t). \tag{18}$$

We are at a very nice spot here, we just need to divide both sides by ε and take the limit as $\varepsilon \to 0$, which will give us the i^{th} component of the gradient. Doing so and cleaning up the right-hand side a little bit lets us arrive at

$$(\nabla \phi)_i = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^{\varepsilon} (\mathbf{F}(\mathbf{x} + t\mathbf{e}_i) \cdot \mathbf{e}_i) \mathrm{d}t = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^{\varepsilon} F_i(\mathbf{x} + t\mathbf{e}_i) \mathrm{d}t. \tag{19}$$

Without the limit, the right-hand side is just the average value of the i^{th} component of $\mathbf{F}$ along the segment. As the segment shrinks to a point, the average is only computed using one point, simply becoming the value of F_i at $\mathbf{x}$. Thus for $i = 1, 2, 3$

$$(\nabla \phi)_i = F_i, \tag{20}$$

and so $\nabla \phi = \mathbf{F}$, completing our proof.

From a mathematical point of view, ϕ is called a potential of $\mathbf{F}$, which is etymologically related to potential energy, but is not potential energy itself (as we will soon see). Additionally, a curious property of ϕ is that it is not unique. We could easily started from any point $\mathbf{b}$ when defining it with the line integral; choosing a different value for $\mathbf{b}$ would have introduced a constant value (lets call it ϕ_0), that does not change the validity of ϕ as a potential for $\mathbf{F}$. You can easily see why below:

$$\nabla(\phi + \phi_0) = \nabla \phi + \underbrace{\nabla \phi_0}_{=\mathbf{0}} = \mathbf{F} \tag{21}$$

Now, all we need to do to prove that a scalar potential function always exists to describe the force via a gradient is to prove that the curl of $\mathbf{F}$ being zero is a consequence of the strong form of Newton's third law. To do so, consider the force created by some particle located at some position $\mathbf{x}_0$ acting on some particle at a position $\mathbf{x}$. Our choice of coordinates is arbitrary, and it doesn't matter where we place our origin, so it would be useful to place the origin at $\mathbf{x_0}$, making $\mathbf{x_0} = \mathbf{0}$. This makes things nice because then the distance between $\mathbf{x_0}$ and $\mathbf{x}$ is just the magnitude of $\mathbf{x}$, which we will write as x. The strong force of Newton's third law stipulates that the force must be parallel to the line between the two objects. With the object generating the force being placed at zero, the force can be written as

$$\mathbf{F} = f(x)\mathbf{x} \tag{22}$$

for some scalar function f.[4] f is restricted to only depending on the magnitude of $\mathbf{x}$ since the isotropy of space restricts us from using quantities that change if $\mathbf{x}$ were just at a different direction from $\mathbf{x_0}$. We can compute the curl of this by simply expanding $\mathbf{x}$ into its components and by performing the cross product with the ∇ operator per usual:

$$(\nabla \times f(x)\mathbf{x})_1 = \partial_2(f(x)x_3) - \partial_3(f(x)x_2), \tag{23}$$
$$(\nabla \times f(x)\mathbf{x})_2 = \partial_3(f(x)x_1) - \partial_1(f(x)x_3), \tag{24}$$
$$(\nabla \times f(x)\mathbf{x})_3 = \partial_1(f(x)x_2) - \partial_2(f(x)x_1). \tag{25}$$

Note how the derivatives in each component are never taken with respect to the component of $\mathbf{x}$ appearing, thus we can always pull them out. By the Pythagorean theorem,

$$x = \sqrt{x_i \cdot x_i}, \tag{26}$$

and thus by the chain rule,

$$\partial_i f(x) = \frac{\cancel{2}x_i}{\cancel{2}\sqrt{x_i \cdot x_i}} f'(x) = x_i \frac{f'(x)}{x} = . \tag{27}$$

Thus we can express the components of the curl as:

$$(\nabla \times f(x)\mathbf{x})_1 = (x_2\frac{f'(x)}{x})x_3 - (x_3\frac{f'(x)}{x})x_2 = \underbrace{(x_2x_3 - x_3x_2)}_{=0}\frac{f'(x)}{x}, \tag{28}$$

$$(\nabla \times f(x)\mathbf{x})_2 = (x_3\frac{f'(x)}{x})x_1 - (x_1\frac{f'(x)}{x})x_3 = \underbrace{(x_3x_1 - x_1x_3)}_{=0}\frac{f'(x)}{x}, \tag{29}$$

$$(\nabla \times f(x)\mathbf{x})_3 = (x_1\frac{f'(x)}{x})x_2 - (x_2\frac{f'(x)}{x})x_1 = \underbrace{(x_1x_2 - x_2x_1)}_{=0}\frac{f'(x)}{x}, \tag{30}$$

and thus $\nabla \times \mathbf{F} = \mathbf{0}$ for any lawful function f. Therefore, by the previous lemma, for any sufficiently nice force $\mathbf{F}$, there exists a scalar function ϕ such that $\mathbf{F} = \nabla\phi$. Additionally, since the gradient operator is linear, we can construct a function for the total force ϕ_{total} by adding each ϕ from every contributing force. This is a powerful fact that is often part of the modern missing link in textbooks, where the reasoning is often misleading or feels like a trick, rather than a standard property of fundamental forces.

[4] This function f is often determined empirically to characterize different laws, like most inverse square laws. It also may depend on additional properties we define like mass, electric charge, etc.

Work and Kinetic Energy

The integral introduced in Eq. (13) is often called work. For me, work was always a confusing concept, as the only thing that was ever emphasized was that it is equal to a change in kinetic energy, which was said to just 'be defined by the equation $m\dot{x}^2/2$'. Not only did the definitions for work and energy feel artificial, but the way textbooks are often worded make the definitions for both seem rather circular. Here, we will avoid this way of explaining the concepts by again demonstrating the quantities of work and kinetic energy as consequences of the mathematical structure of Newton's laws, rather than something introduced as fundamental.

To begin, we will attempt to evaluate the line integral that defines ϕ. One trick we can use is knowing that, since the total force acting on a particle is represented by ϕ_{total}, we can calculate the integral as:

$$\phi_{\text{total}} = \int_C \mathbf{F}_{\text{total}} \cdot \mathrm{d}\mathbf{x}. \tag{31}$$

Since $\mathbf{F}_{\text{total}}$ is just the sum of all the forces, we can substitute Newton's second law into the integrand:

$$\phi_{\text{total}} = \int_C m\ddot{\mathbf{x}} \cdot \mathrm{d}\mathbf{x} \tag{32}$$

The trick here is to realize via the definition of the differential that $\mathrm{d}\mathbf{x} = \dot{\mathbf{x}}\mathrm{d}t$, where we can then treat $\mathrm{d}t$ as a constant and pull it out of the dot product:

$$\phi_{\text{total}} = \int_{t_1}^{t_2} m(\ddot{\mathbf{x}} \cdot \dot{\mathbf{x}})\mathrm{d}t = \int_{t_1}^{t_2} (m\dot{\mathbf{x}}) \cdot (\ddot{\mathbf{x}}\mathrm{d}t) = \int_C (m\dot{\mathbf{x}}) \cdot \mathrm{d}\dot{\mathbf{x}}. \tag{33}$$

Now, in these changed variables, we can compute the above line integral per usual, arriving at

$$\phi_{\text{total}} = \frac{1}{2}m\dot{\mathbf{x}}^2 + C = \frac{1}{2}m\dot{x}^2 + C \tag{34}$$

where C is just an integration constant. Note that the result of this integral is always positive as $m > 0$ and $\dot{x}^2 > 0$. If we were to rearrange this equation, it's often a convention to keep the term involving the quantity $m\dot{x}^2/2$ positive as it was previously. One rearrangement of this equation is:

$$\frac{1}{2}m\dot{x}^2 + (-\phi_{\text{total}}) = -C. \tag{35}$$

This is quite interesting; it appears that the computation of our integral implies that there are two quantities, one depending solely on the state of the particle (the first term), and another depending on what interactions with that particle are going on (the second term). As you might have assumed, these terms are what we call *kinetic* and *potential* energy, where the quantity $-C$ is called the *total mechanical energy.* I feel that this description and definitions of these quantities, as it avoids blatantly defining each one, and instead shows them as terms in a very beautiful equation. For our purposes, we will call the kinetic energy T, the potential energy U, and the total mechanical energy $\mathcal{H}$ (for reasons we will see later). Therefore:

$$T + U = \mathcal{H}, \tag{36}$$

which we can see by taking the time derivative of both sides, brings us to:

$$\frac{\mathrm{d}}{\mathrm{d}t}(T + U) = 0. \tag{37}$$

This should remind you of momentum, in which we had a fundamentally invariant quantity also arise as an innocent integration constant. A similar process exists for proving the conservation of angular momentum, but we will skip this in place of proving the principle of least action.

I think that a lack of the above derivation in modern undergraduate textbooks is also why people tend to be confused about negative potential energy. Take, for example, gravity; it is easy to show that from its central force law that the work function is:

$$\phi(\mathbf{x}) = \frac{Gm_1m_2}{x}. \tag{38}$$

Since we just *chose* to keep kinetic energy positive, since it's a pure power of 2, we chose to make the potential energy $-\phi$ so we can have a formula for energy that is conserved. We could have easily done it the other way around, but negative kinetic energy feels *even weirder* than negative potential energy. Thus, by negating the work function, we get the familiar expression:

$$U(\mathbf{x}) = -\frac{Gm_1m_2}{x}. \tag{39}$$

We've examined time derivatives of these functions, but what if we took the gradient instead? Consider the potential energy, which we defined as:

$$U = -\phi_{\text{total}}. \tag{40}$$

Taking the gradient of both sides reveals that:

$$\nabla U = -\nabla\phi_{\text{total}}. \tag{41}$$

Recall that we defined ϕ as the special function that recovers the force when its gradient is taken, thus we arrive at the fact that:

$$\nabla U = -\mathbf{F}_{\text{total}}. \tag{42}$$

This is a very neat, intuitive relationship, which we will use later, often allowing us to call the force as a vector that pushes the particle down the potential energy landscape, until it finds a stable state (which we soon shall see). Additionally, we can use this fact with the conservation of energy to find the gradient of T, which is a fun exercise worth doing.

The Action

Maupertuis first wrote down that action was the quantity mvx. This aided in describing the path that a beam of light may take, and Maupertuis' key insight was that it also described the path a free particle[5] takes. Here, x really means the displacement magnitude of the particle, instead of the magnitude of its position at a given time, and v really means the average speed. Notice that if v is constant, like with a particle of light, it is easy to show the principle of least action is equivalent to a principle of least time. This was demonstrated by Maupertuis to emphasize that his principle aligned with Fermat's principle of least time and Newton's laws of motion.

Here we will not only demonstrate where the formula mvx comes from, but also generalize the formula for when the particle is experiencing force, ultimately deriving the principle of least action demonstrated by Lagrange. To begin, we revisit Newton's 2$^{\text{nd}}$ law:

$$\sum_j \mathbf{F}_{ji}(\mathbf{x}_i) = m_i\ddot{\mathbf{x}}_i. \tag{43}$$

[5] This just means zero force acts on the particle

To simplify things, we will denote the total force on the left-hand side acting on the i^{th} particle as $\mathbf{F}_i$, with a corresponding potential energy U_i. By moving the terms in Newton's 2^{nd} law to the same side, we can write the set of equations:

$$m_1\ddot{\mathbf{x}}_1 - \mathbf{F}_1 = \mathbf{0},\ m_2\ddot{\mathbf{x}}_2 - \mathbf{F}_2 = \mathbf{0},\ \dots\ ,\ m_i\ddot{\mathbf{x}}_i - \mathbf{F}_i = \mathbf{0},\ \dots \tag{44}$$

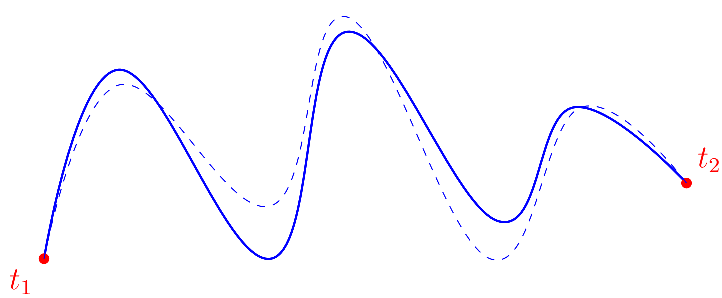
We've played around with the quantities involved above in many ways, but we have yet to add the equations together. Lets explore this and do so:

$$\sum_i (m_i\ddot{\mathbf{x}}_i - \mathbf{F}_i) = \mathbf{0} \tag{45}$$

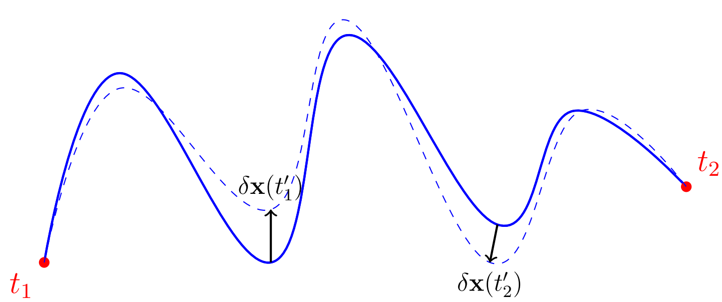
Now, we know that the gradient of the potential energy is the negative force, so we can make the neat substitution as follows:

$$\sum_i (m_i\ddot{\mathbf{x}}_i + \nabla U_i) = \mathbf{0} \tag{46}$$

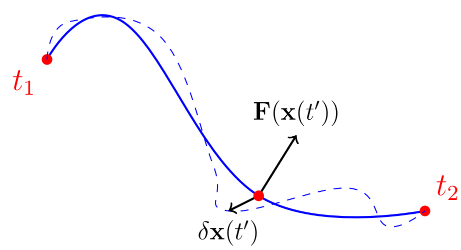
Now we have a quantity that is kind of hard to think about, but does 'capture' the entire system. We could go straight in and compute the left-hand side, which (unsurprisingly) would just give zero if we are summing over the entire universe. Lagrange's insight in getting something meaningful out of this form of Newton's 2^{nd} law was to imagine if the particle takes a slightly different path to what it actually takes while moving according to a certain force (i.e., consider a small *variation* in the path). The stipulation with this variation is that since we know where the particle really starts and ends, we want the new path to also run through those points. This is done so we can really see what would happen to the above quantity if the laws of physics were actually different for the two paths, rather than the two paths just being in different initial states. Such a path might look something like this:



We will label the difference between these two curves as $\delta\mathbf{x}$. We have used the letter δ here instead of d or Δ to signify that this is an arbitrary variation of the curve itself that can change throughout different points of the path. The vectors $\delta\mathbf{x}$ have been pointed out for specific points in time t_1', t_2', $t_1 < t_1' < t_2' < t_2$ for our variation below:



Notice how the total force, dictating the acceleration and therefore curvature, is often aligned in some way with this arbitrary variation so that their dot product is rarely zero for marginally different paths. To motivate this further, let's take a look at another random curve with a random small variation whilst plotting the vector $\delta\mathbf{x}$ and $\mathbf{F}$.



As you can see, the dot product is almost never zero unless the small variation intersects with the true path that the particle takes. Although this statement isn't rigorous and isn't technically always true, it does give us motivation to study the dot product of the variation with quantities related to the force. Additionally, studying the dot product is yet another way to turn our vector problem (Newton's 2^{nd} law) into a scalar problem that is easier to analyze. This might motivate us to try to take the dot product of our small variation with the quantity in Eq. (46). Additionally, as it is clear that there might be some nice emergent properties when $\delta\mathbf{x}$ is small, it might be easiest to study $\delta\mathbf{x}$ as an *infinitesimal variation*. This notion of considering an infinitesimal variation in the path is the idea that coined the name *calculus of variations*.

This time, take the equations in Eq. (44) and perform a dot product with the variation in each path $\delta\mathbf{x}_i$. Adding these together as before gives us the fact:

$$\sum_i (m_i\ddot{\mathbf{x}}_i + \nabla U_i)\cdot\delta\mathbf{x}_i = \mathbf{0} \tag{47}$$

This sum is called the *total virtual work* of the system, where we used the adjective virtual to emphasize that the variations are somewhat non-physical. Before we go in and try to manipulate this expression, let's take a step back and talk about what properties the δ operator might have when it represents an infinitesimal variation. Since this variation is infinitesimal, we might expect it to act similarly to a differential, so it shares the following properties:

- (*Chain Rule*)

$$\mathrm{d}f(x,y) = \partial_x f(x,y)\mathrm{d}x + \partial_y f(x,y)\mathrm{d}y \;\longleftrightarrow\; \delta f(x,y) = \partial_x f(x,y)\delta x + \partial_y f(x,y)\delta y \tag{48}$$

- (*Commutativity With Differentiation*)

$$\frac{\mathrm{d}}{\mathrm{d}x}(\mathrm{d}f(x)) = \mathrm{d}f'(x) \;\longleftrightarrow\; \frac{\mathrm{d}}{\mathrm{d}x}(\delta f(x)) = \delta f'(x) \tag{49}$$

- (*Leibniz's Rule/Commutativity With Integration*)

$$\mathrm{d}\left(\int f(x)\mathrm{d}x\right) = \int \mathrm{d}(f(x)\mathrm{d}x) \;\longleftrightarrow\; \delta\left(\int f(x)\mathrm{d}x\right) = \int \delta(f(x)\mathrm{d}x) \tag{50}$$

Thus, by distributing the dot product in Eq. (47), we can utilize the above properties as follows:

$$\sum_i (m_i \ddot{\mathbf{x}}_i \cdot \delta \mathbf{x}_i + \underbrace{\nabla U_i \cdot \delta \mathbf{x}_i}_{=\delta U_i}) = \mathbf{0}, \tag{51}$$

$$\sum_i (m_i \ddot{\mathbf{x}}_i \cdot \delta \mathbf{x}_i + \delta U_i) = \mathbf{0}. \tag{52}$$

Now, to compute the first term in the summand, we have to make sure we are considering a necessary subtlety. Since δ obeys the chain rule, it also must obey the product rule as well. The first term involves a second derivative and a δ operator on a standalone variable. Thus, we might investigate the time derivative of the product:

$$\frac{\mathrm{d}}{\mathrm{d}t}(\dot{\mathbf{x}} \cdot \delta \mathbf{x}) = \ddot{\mathbf{x}} \cdot \delta \mathbf{x} + \dot{\mathbf{x}} \cdot \underbrace{\frac{\mathrm{d}}{\mathrm{d}t}(\delta \mathbf{x})}_{=\delta \dot{\mathbf{x}}}. \tag{53}$$

As expected, our desired term $(\ddot{\mathbf{x}} \cdot \delta \mathbf{x})$ appears in the sum, where we can isolate it easily. Isolating for our desired term and using the commutativity of the δ operator with differentiation gives us:

$$\ddot{\mathbf{x}} \cdot \delta \mathbf{x} = \frac{\mathrm{d}}{\mathrm{d}t}(\dot{\mathbf{x}} \cdot \delta \mathbf{x}) - \dot{\mathbf{x}} \cdot \delta \dot{\mathbf{x}}. \tag{54}$$

We can go even further and use the chain rule to reduce the second term on the right-hand side even more. To do this, let's consider the following quantity:

$$\delta(\dot{\mathbf{x}}^2) = 2\dot{\mathbf{x}} \cdot \delta \dot{\mathbf{x}} \tag{55}$$

This yields double the second term, so we may write that

$$\ddot{\mathbf{x}} \cdot \delta \mathbf{x} = \frac{\mathrm{d}}{\mathrm{d}t}(\dot{\mathbf{x}} \cdot \delta \mathbf{x}) - \frac{1}{2}\delta(\dot{\mathbf{x}}^2). \tag{56}$$

Now let us substitute this form of the first term into the summand back into Eq. (52) to arrive at:

$$\sum_i (m_i(\frac{\mathrm{d}}{\mathrm{d}t}(\dot{\mathbf{x}}_i \cdot \delta \mathbf{x}_i) - \frac{1}{2}\delta(\dot{\mathbf{x}}_i^2)) + \delta U_i) = \mathbf{0}. \tag{57}$$

From here, we can easily distribute m_i and split the sum to separate the terms involving the time derivative from those without it:

$$\sum_i \frac{\mathrm{d}}{\mathrm{d}t}(m_i \dot{\mathbf{x}}_i \cdot \delta \mathbf{x}_i) - \sum_i (\frac{1}{2} m_i \delta(\dot{x}_i^2)) - \delta U_i) = \mathbf{0}. \tag{58}$$

You may already notice that a familiar quantity showed up after we did this, namely, the variation in kinetic energy. Since we assume each particle's mass doesn't vary, we can use the chain rule in reverse (treating δ as a linear operator) and pull the operator out of the sum. Likewise, we can do the same with the derivative in the first sum. We will also clean things up by moving the second term to the left-hand side. Doing so gives us:

$$\frac{\mathrm{d}}{\mathrm{d}t}\sum_i (m_i \dot{\mathbf{x}}_i \cdot \delta \mathbf{x}_i) = \delta \sum_i (\frac{1}{2} m_i \dot{\mathbf{x}}_i^2 - U_i). \tag{59}$$

Notice here how our squared vector is just the magnitude squared, and so we have effectively de-vectorized the right side of this equation. Now we might be curious as to what would happen if we got rid of the derivative on the left-hand sides. We can do this by integrating both sides from the starting time of the path of each particle (t_1) to the ending time (t_2). We might want to choose these times, because remember that the endpoints of the varied and actual path must coincide, implying that $\delta\mathbf{x} = \mathbf{0}$ at those points.

$$\underbrace{\int_{t_1}^{t_2} \frac{\mathrm{d}}{\mathrm{d}t} \sum_i (m_i \dot{\mathbf{x}}_i \cdot \delta\mathbf{x}_i)\mathrm{d}t}_{=0} = \int_{t_1}^{t_2} \delta \sum_i (\frac{1}{2} m_i \dot{x}_i^2 - U_i)\mathrm{d}t. \tag{60}$$

By the fundamental theorem of calculus, the left-hand side is just the value of the function being acted on by the time derivative at the endpoints t_1, t_2. Since $\delta\mathbf{x}_i = \mathbf{0}$ at these points, the entire left side vanishes. Additionally, we're able to pull out the δ operator from the integral on the right-hand side since it commutes with integration. After doing so we arrive at the neat equation:

$$\delta \int_{t_1}^{t_2} \sum_i (\frac{1}{2} m_i \dot{x}_i^2 - U_i)\mathrm{d}t = 0. \tag{61}$$

This is awesome, now we can finally define the *Action* as the integral being acted on by the δ operator here, which we will denote as S. The integrand is called the *Lagrangian*, which we will denote as $\mathcal{L}$. Additionally, notice that summing every individual particle's kinetic and potential energy just gives the total kinetic energy minus the total potential energy. Therefore we can compactly write the following equation:

$$\delta \int_{t_1}^{t_2} (T - U)\mathrm{d}t = 0 \tag{62}$$

This is the rigorous and generalized principle of least action. Note that δ is kind of telling us here that the true path the particle takes is the one in which there is zero change in the action given an infinitesimal change in the path. This may sound exactly like a local minimum to you (as intended), but it can also be a local maximum, and so a less misleading way of describing the generalized principle of least action is *the principle of stationary action.* The only reason we call it the principle of *least* action, is that the stationary point found using Maupertuis' action is also a minimum.

The idea that the difference caused by a path variation must be zero can be put into notation by using the fact that action is what is called a *functional* in the calculus of variations in the sense that instead of taking in a normal variable, the value of action depends on a function variable. We write the action of the set of paths $\mathbf{x}_i(t)$ as $S[\mathbf{x}_i(t)]$. Therefore, we can use by definition:

$$\delta S[\mathbf{x}_i(t)] = S[\mathbf{x}_i(t) + \delta\mathbf{x}_i(t)] - S[\mathbf{x}_i(t)] = 0. \tag{63}$$

Additionally, what we just defined to be the Lagrangian $\mathcal{L}$ is just a function that takes in details about the path. Namely, it may depend only on the paths, their velocities, and time,[6] often written as $\mathcal{L}(\mathbf{x}_i(t), \dot{\mathbf{x}}_i(t), t)$.

[6] Technically, action should also be written this way, but it the velocity input is often omitted for ease of notation in writing Eq. (63). We will see soon that if the time derivative of the Lagrangian is nonzero, then energy is not conserved. In our case, we derived a conservative Lagrangian since we assumed that $\mathbf{F}$ was a fundamental force, so it doesn't depend on time here.

Now we have the tools to see *why* Maupertuis' action was defined the way it was. For a free particle, $U = 0$, and $T = m\dot{x}^2/2$, so:

$$S = \int_{x_1}^{x_2} \frac{1}{2}m\dot{x}^2 \mathrm{d}t = \int_{t_1}^{t_2} (\frac{1}{2}m\dot{x})(\dot{x}\mathrm{d}t) = \int_{t_1}^{t_2} \frac{1}{2}m\dot{x}\mathrm{d}x = \frac{1}{2}mvx \tag{64}$$

where we let x be the total displacement and v be the average speed over the interval. Notice that minimizing $mvx/2$ is the same as minimizing mvx since the δ operator is linear, and so we recover the same 'action' discovered by Maupertuis.

The Euler-Lagrange Equation

We have just derived an extremely powerful fact, capable of solving virtually any physics problem, despite it being overkill sometimes. To effectively use this fact, we can construct a relationship involving only derivatives from the principle of least action, coined the *Euler-Lagrange* equation. With this equation, we can solve physics problems even using coordinates involved in a very awkward curvilinear system (i.e., spherical, cylindrical, some path parameterization, etc.). All we would have to do is change the integrand by computing x_i^2 for each vector in the system in terms of the new coordinates, so we could compute T. Additionally, U is just a function of $\mathbf{x}_i$, so we just have to make sure that U is configured for the coordinate system as well, which can always be done by a substitution in traditional problems. In fact, finding an expression for T is often the most tedious part when using the principle of least action in alternative coordinates.

The reason I mention this fact about computing T is that in general coordinate systems, T can take on *almost* any form it wants, just like U. Thus, the entire integrand in the principle of least action might also take on any form it likes. Additionally, if we're modeling a system with non-fundamental forces (like friction) there would be an additional dependence on t. So, if we want to solve the problem in general, it's probably best to just refer to this integrand to a just *a function* (which we call the Lagrangian, '$\mathcal{L}$'), and treat it like it could be anything it wants.

Re-writing the principle of least action with the Lagrangian in mind looks like:

$$\delta \int_{t_1}^{t_2} \mathcal{L}(q_i, \dot{q}_i, t)\mathrm{d}t = 0. \tag{65}$$

Note that here I have made the intentional decision to replace our notation x_i with q_i to emphasize that the coordinates being used in the Lagrangian need not be cartesian. Often times, the coordinates q_i are referred to as *generalized coordinates*. Lagrange and Euler's idea was to effectively *undo* our construction of the principle of least action to rediscover some baseline equation like Newton's 2^{nd} Law, while staying in terms of $\mathcal{L}$. They believed this was useful, because it would essentially be an equation which shares the simplicity of Newton's laws, but where everything is now a scalar! So, if we wanted to actually undo what we just did, we could refer back to the steps of the previous derivation. For starters, we last swapped the δ and the integral, lets undo that:

$$\int_{t_1}^{t_2} \delta\mathcal{L}(q_i, \dot{q}_i, t)\mathrm{d}t = 0. \tag{66}$$

Now recall that before we swapped the delta, we did some cumbersome algebraic manipulations, and exploited the properties we gave the δ operator to get a nice form. This was quite cumbersome because we were starting from nothing. Now we actually know that this thing we call the Lagrangian is important, and that moving backwards is actually a lot easier than moving forwards because we know that the δ operator has a chain rule. Using this chain rule gives us:

$$\int_{t_1}^{t_2} \left(\frac{\partial \mathcal{L}}{\partial q_i} \delta q_i + \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \delta \dot{q}_i + \frac{\partial \mathcal{L}}{\partial t} \delta t \right) \mathrm{d}t = 0. \tag{67}$$

We need to do some housekeeping here. In theory, we could keep the time derivative term, but doing so assumes our Lagrangian doesn't necessarily stem from fundamental principles. In fact, I recommend following the rest of this derivation with the term intact on your own if interested. You will notice that your result is completely analogous to an equation involving momentum which we discussed previously. For now, we will assume that the Lagrangian doesn't depend on time, because we are only interested in exploring the structure of the fundamental laws, so we get:

$$\int_{t_1}^{t_2} \left(\frac{\partial \mathcal{L}}{\partial q_i} \delta q_i + \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \delta \dot{q}_i \right) \mathrm{d}t = 0. \tag{68}$$

Now, it would be nice if we could factor something out here. Luckily, since we know that the δ operator commutes with the derivative in the second term, we can actually use the product rule to get rid of it. We can do this by recognizing that the second term is one of the terms resulting from the following derivative (constructed by removing the derivative from $\delta \dot{q}_i$):

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i} \delta q_i \right) = \left(\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \right) \delta q_i + \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \underbrace{\left(\frac{\mathrm{d}}{\mathrm{d}t} \delta q_i \right)}_{= \delta \dot{q}_i}. \tag{69}$$

Isolating for the second term and substituting gives us:

$$\int_{t_1}^{t_2} \left(\frac{\partial \mathcal{L}}{\partial q_i} \delta q_i + \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i} \delta q_i \right) - \left(\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \right) \delta q_i \right) \mathrm{d}t = 0. \tag{70}$$

Although we cant explicitly factor anything else out yet, notice that the second term is a pure derivative. This reminds us of a step we did in the prior version of the derivation, where we used the fact that δx_i was zero at the endpoints t_1, t_2 of the integral, and by using the fundamental theorem of calculus. We can do exactly the same thing here, as it is a mirror image of that step but in terms of the Lagrangian, and the factor δq_i appears. A coordinate system where δq_i isn't zero at the endpoints is invalid, because then it isn't one-to-one with the space you're trying to quantify, thus it must be zero for the same reason as δx_i. The term completely drops out, and we are left with our desired form where we can factor out δq_i like so:

$$\int_{t_1}^{t_2} \left(\frac{\partial \mathcal{L}}{\partial q_i} - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \right) \delta q_i \mathrm{d}t = 0. \tag{71}$$

Now, let us examine this form of the varied action integral. This expression is forced to equal zero, but what could possibly allow it to be zero? Let's consider the factor with the δ operator. This is an arbitrary variation in the path, meaning it can span a multitude of possible infinitesimal functions with time. Just because the variation is infinitesimal does not mean that it is zero. For this integral to be zero, the variation would have to be identically zero, odd about its point, etc. However, this is a finite number of conditions which when changed does not change the value of the first factor of the integrand. Thus there exists a case where the integral is zero and the variation factor does not satisfy one of the conditions to make the integral zero. Since this has nothing to do with the value of the first factor, in all possible cases the first factor must be the culprit, and must satisfy one of the conditions that make the integral zero. Since the endpoints t_1, t_2 is arbitrary, and the equation is true for any selection of endpoints, we know that the first factor (which isn't arbitrary) can't be zero for any of the reasons that depend on local form, such as being odd about a point, as we could easily perform the integral on an interval that doesn't contain this point. The only condition that doesn't depend on local form is the condition that this function is identically zero, which is a function that looks the same everywhere. Thus, the first factor itself must be identically zero. Putting this into equation form looks like:

$$\frac{\partial \mathcal{L}}{\partial q_i} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \dot{q}_i} = 0. \tag{72}$$

This is the simplest form of the Euler-Lagrange Equation(s). Note that in the case where there is more than one coordinate, like in 3D Euclidean space, this is actually multiple equations, for $i = 1, 2, \ldots$. You may argue that this is now no better than a vector equation, as the many equations constitute the same thing as a compactified vector equation. However, not that our coordinates don't have to be cartesian, and we can choose a curvilinear coordinate system that only requires us to use one equation. This makes sense if you have a problem where an object is constrained to a specific line in space. For example, a simple pendulum is always swinging on a circular arc, and only needs one curvilinear coordinate θ, where it would need 2 if we were in cartesian coordinates. Another example is if we have a ball rolling down a hilly landscape (given the constraint that it can't be launched off the hill). If the landscape is in 3D space, we could easily skip the otherwise annoying geometric setup involving 3D vectors by using just 2 Euler-Lagrange Equations corresponding to a 2 dimensional parametric surface that reconstructs the landscape. The only price that you would have to pay when using this method, is that you make sure that your Lagrangian (Kinetic energy and all) complies with the coordinate system.

In fact, let's see that this equation is the same as Newton's 2^{nd} Law when using cartesian coordinates to anchor this idea down. In cartesian coordinates $T = m\dot{x}_i^2/2$ so the Lagrangian for the i^{th} cartesian coordinate is per usual:

$$\mathcal{L} = \frac{1}{2}m\dot{x}_i^2 - U(x_i) \tag{73}$$

Now we should compute the partial derivatives with respect to the coordinates like so:

$$\frac{\partial \mathcal{L}}{\partial x_i} = -U'(x_i), \tag{74}$$

$$\frac{\partial \mathcal{L}}{\partial \dot{x}_i} = \frac{\cancel{1}}{\cancel{2}}m(\cancel{2}\dot{x}_i) = m\dot{x}_i. \tag{75}$$

Before we compute the time derivative of the second equation just written, notice that the second line resembles the momentum in cartesian coordinates. In curvilinear coordinate systems where momentum probably takes a different, we use this derivative of the Lagrangian to *define* the momentum. When we define it this way, we call the partial derivative the *conjugate momentum*. With this in mind, you may instantly know that the time derivative of the second line is just $m\ddot{x}_i$. Thus, the Euler-Lagrange equation is just:

$$-U'(x_i) - m\ddot{x}_i = 0, \tag{76}$$

which can be equivalently expressed as

$$m\ddot{x}_i = -U'(x_i). \tag{77}$$

If we were to vectorize these equations and remove the components, we would clearly get:

$$m\ddot{\mathbf{x}} = -\nabla U(\mathbf{x}). \tag{78}$$

However, the negative gradient of U is just the definition of the net force, and so we have effectively recovered Newton's law!

I hope this has given you an insight on the topic of what the true mathematical (and philosophical) structure of Newton's laws really are. There are so many stack exchange threads that just regard the Euler-Lagrange equations as 'defined' to reconstruct Newton's law, and leave the confused individual (like me) even more confused. I should also point out that this analysis might more easily point out the necessity of Newton's 1$^{\text{st}}$ law (among others), which often gets swept under the rug, and is always worded in a way that makes it seem redundant. It is quite difficult to fathom the intuition that Isaac Newton had when constructing his laws. In fact, I believe some physics-inclined linguists have recently reported that his original text, which was written in Latin, may have lacked translations that did not accurately capture the presentation of his laws for linguistic reasons, but I know much less about this. Additionally, these ideas are very old, and haven't been presented in modern formal writing for a while, and it's unfortunate that the comprehensiveness of some proofs have been lost to time/accessibility. Hopefully this text may act as a bridge for those who wish to climb the ivory tower without pretending like they understand what their undergraduate textbook says when they 'define' momentum, energy, or the Lagrangian. I fear many physics majors loose grip in understanding theory when presented with the principle of least action, as it is the first idea that doesn't have an intuitive explanation. This however, should most definitely be explained.