

Physics 207, Fall 2010

Notes on Mechanics and Waves

Introduction

We have seen that the diffraction of light can be derived by adopting the principle that a light ray, that travels between some given begin and endpoint, follows the path that takes the least amount of time. This rule, called Fermat's principle, is an example of an extremization principle: the path of the light ray minimizes the travel time. As we will see, it is possible to reformulate all of classical mechanics in a similar way, where the classical motion of a given particle or system is determined by the so-called *action principle*: the classical motion extremizes a quantity S , called the action, which is some suitably chosen functional that depends on the path of the particle, or more generally, on the motion of the classical system.

Fermat's principle redux

As a warm-up, let's consider Fermat's principle in the simplest setting: a light ray that travels in vacuum between some given beginpoint A and endpoint B . It follows a straight line. This is clearly the shortest path between the two points, and since the speed of light is fixed, it minimizes the travel time.

Let us write an explicit formula for this shortest path. Let \vec{x}_A and \vec{x}_B denote the coordinates of A and B . Now introduce a parameter s that runs between 0 and 1 and write

$$\vec{X}(s) = (1 - s)\vec{x}_A + s\vec{x}_B. \quad (1)$$

As s runs from 0 to 1, the position $\vec{X}(s)$ traces the straight path between A and B . The light travel time along the path $\vec{X}(s)$ is $1/c$ times the distance d_{AB} between A and B

$$d_{AB} = |\vec{x}_B - \vec{x}_A| = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2 + (z_B - z_A)^2}. \quad (2)$$

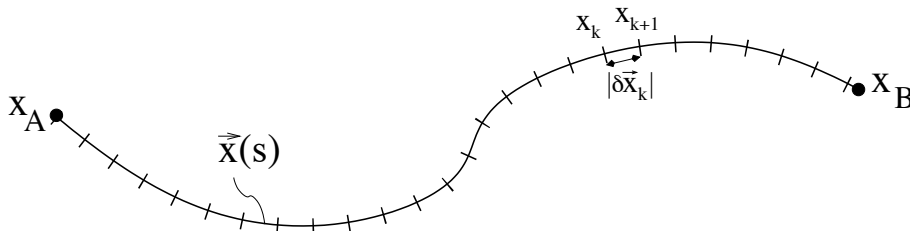
Now suppose we would want to *prove* that the straight line is really the shortest path between A and B . To do this, we would need to take the following three steps:

- (i) Consider all other possible paths $\vec{x}(s)$ between A and B .
- (ii) Compute the *total length* of the path $\vec{x}(s)$. We will denote this *length* by $\ell[\vec{x}(s)]$.
- (iii) Show that $\ell[\vec{x}(s)] \geq d_{AB}$ always.

Let's start with steps (i) and (ii). So consider an arbitrary path between A and B :

$$\vec{x}(s), \quad \text{with} \quad \vec{x}(0) = \vec{x}_A, \quad \vec{x}(1) = \vec{x}_B \quad (3)$$

Now imagine that we want to measure the *total length* of the path $x(s)$. The practical and correct way to do this is to first cut the path into many little steps, then measure the length of each little step, and finally add up the lengths of all the little steps.



So let's cut the interval $0 \leq s \leq 1$ into many extremely small steps of equal length Δs :

$$s_{k+1} - s_k = \Delta s \quad (4)$$

Denote the position along the path after k little steps by \vec{x}_k . In other words: $\vec{x}_k \equiv \vec{x}(s_k)$. The k -th little step is a small 3-vector

$$\Delta \vec{x}_k = \vec{x}_{k+1} - \vec{x}_k \quad (5)$$

with infinitesimal length

$$\Delta \ell_k = |\Delta \vec{x}_k| = \left| \frac{\Delta \vec{x}_k}{\Delta s} \right| \Delta s. \quad (6)$$

The total length of the path is given by the sum over all the infinitesimally small lengths

$$\text{Length of path} = \sum_k \Delta \ell_k = \sum_k \left| \frac{\Delta \vec{x}_k}{\Delta s} \right| \Delta s \quad (7)$$

In the end we wish to take the number of infinitesimal steps to infinity, and the length to zero while keeping the total length of all the steps combined fixed. In this limit, the above sum becomes an integral over s

$$\boxed{\ell[\vec{x}(s)] = \int_0^1 ds \left| \frac{d\vec{x}}{ds} \right|} \quad (8)$$

This formula for the *length* of the path $\vec{x}(s)$ is an example of a *functional*: a quantity that depends on a complete path $\vec{x}(s)$. We will see more examples of such functionals soon, when we discuss the Lagrange and Hamilton formulation of classical mechanics.

Exercise 1: Show that for the straight path (1)

$$\ell[\vec{X}(s)] = d_{AB}$$

with d_{AB} the minimal distance given in eqn (2).

We wish to prove that the *total length* of any path $\vec{x}(s)$ is larger than the minimal distance d_{AB} . To do this, we first derive the somewhat weaker result, that the straight path is the only path for which the length is *extremal*. Then, with a little bit more work, one can show that the *extremum* is in fact a *global minimum*.

An extremum of a function is found by setting the first derivative of the function equal to zero.¹ An extremum of a functional is found in a similar way: we look for a path $\vec{X}(s)$ such that the the first variation of the *Length* functional vanishes. Consider any path $\vec{x}(s)$ from A to B in the neighborhood some candidate shortest path $\vec{X}(s)$. We can write

$$\vec{x}(s) = \vec{X}(s) + \delta\vec{x}(s). \quad (10)$$

with $\delta\vec{x}(s)$ a small variation of the path that leaves the end points A and B fixed: $\delta\vec{x}(0) = \delta\vec{x}(1) = 0$. The first variation of the *length* functional is defined by

$$\delta\ell[\vec{X}(s)] = \ell[\vec{X}(s) + \delta\vec{x}(s)] - \ell[\vec{X}(s)] \quad (11)$$

where on the right-hand side we are instructed to *keep only the linear term in the variation* $\delta\vec{x}$. In other words, we drop all terms of order $(\delta\vec{x})^2$ and higher.

At this point, we will *not* assume that $\vec{X}(s)$ is the straight path. Instead, we wish to derive that $\vec{X}(s)$ must be the straight path from the condition that extremizes the *length*:

$$\delta\ell[\vec{X}(s)] = 0 \quad (12)$$

To simplify the following formulas a bit, let us *temporarily* introduce the abbreviated notation $\dot{x} \equiv \frac{dx}{ds}$, etc. In this notation, the integrand of eqn (8) for the *length* functional

¹The extremum of a function is the point where its first variation vanishes (Taylor expansion):

$$f(x_0 + \delta x) = f(x_0) + f'(x_0) \delta x + \frac{1}{2} f''(x_0) (\delta x)^2 + \dots \quad (9)$$

$f'(x_0) = 0$ is the extremum. If $f''(x_0) > 0$ we have a minimum, while if $f''(x_0) < 0$ we have a maximum. In case of a function of many variables, $f(x_1, \dots, x_n)$, the extrema are characterized by: $\frac{\partial f}{\partial x_i} = 0 \quad \forall i$.

can be written as $|\dot{\vec{x}}| = \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2} = \sqrt{\dot{\vec{x}} \cdot \dot{\vec{x}}}$. We now compute

$$\begin{aligned} |\dot{\vec{X}} + \delta\dot{\vec{x}}| &= \sqrt{\dot{\vec{X}} \cdot \dot{\vec{X}} + 2\dot{\vec{X}} \cdot \delta\dot{\vec{x}} + \delta\dot{\vec{x}} \cdot \delta\dot{\vec{x}}} \\ &= |\dot{\vec{X}}| + \frac{\dot{\vec{X}} \cdot \delta\dot{\vec{x}}}{|\dot{\vec{X}}|} + \dots \end{aligned} \quad (13)$$

where \dots are terms that higher order in the variation $\delta\vec{x}$. Plugging this result into the integral (8), we find the following result for the first variation of the *length*

$$\delta\ell[\vec{X}(s)] = \int_0^1 ds \frac{\dot{\vec{X}} \cdot \delta\dot{\vec{x}}}{|\dot{\vec{X}}|} \quad (14)$$

Finally, by performing a partial integration, we can rewrite the above formula in a somewhat more useful form:

$$\delta\ell[\vec{X}(s)] = - \int_0^1 ds \delta\vec{x} \cdot \frac{d}{ds} \left(\frac{\dot{\vec{X}}}{|\dot{\vec{X}}|} \right) \quad (15)$$

Note that the partial integration step does not produce any boundary terms, since the variation $\delta\vec{x}(s)$ vanishes at the end points of the domain of integration.

Exercise 2: Show that $\vec{X}(s)$ is an extremum of the *length* functional ℓ if and only if $\vec{X}(s)$ is a straight path.

So far, we considered Fermat's principle in vacuum. We could straightforwardly generalize the discussion to the case where the light is traveling through a medium with the space dependent index of refraction $n(\vec{x})$. In this case, Fermat's principle can be put in mathematical form as the condition that light follows the path $X(s)$ that minimizes the *time* functional

$$\tau[\vec{x}(s)] = \frac{1}{c} \int_0^1 ds n(\vec{x}(s)) \left| \frac{d\vec{x}}{ds} \right|, \quad (16)$$

where we used that the speed of light at the location x equals $c(x) = c/n(x)$.

Exercise 3: Show that

$$\delta \tau[\vec{X}(s)] = \frac{1}{c} \int_0^1 ds \delta \vec{x} \cdot \left[\vec{\nabla} n(\vec{X}(s)) |\dot{\vec{X}}| - \frac{d}{ds} \left(n(\vec{X}(s)) \frac{\dot{\vec{X}}}{|\dot{\vec{X}}|} \right) \right]$$

where $\vec{\nabla} n(\vec{x}) = \frac{\partial n(\vec{x})}{\partial \vec{x}}$.

Hamilton's principle

Classical mechanics can be formulated in terms of Hamilton's principle, more commonly known as the "*principle of least action*". It states that the classical trajectory $x(t)$ of a particle, or more generally, the classical motion of any mechanical system, extremizes a certain well-chosen functional S , called *the action*. The action functional S depends on the trajectory $\vec{x}(t)$, or the generalized path that describes motion of the mechanical system. It is carefully chosen such that the equations obtained by extremizing S are identical to the Newton equations of motion of the system.

Let $x(t)$ denote some *arbitrary path* that starts at $x = x_I$ at the initial time $t = t_I$ and ends at $x = x_F$ at final time $t = t_F$, and let $X(t)$ denote the *candidate classical path* with the same initial and final conditions. Formally, we can write the condition that the action takes an extremal value at $x(t) = X(t)$ as

$$\left. \frac{\delta S[x(t)]}{\delta x(t)} \right|_{x(t)=X(t)} = 0 . \quad (17)$$

Notice that, similar as the derivative of a function depends on the point where it is evaluated, the *variation* of a functional depends on the path $X(t)$ at which it is evaluated. We can rewrite (17) in a way which is easier to understand and more practical to use. Define the variation δS of the action as

$$\delta S \equiv \lim_{\epsilon \rightarrow 0} \frac{S[X(t) + \epsilon \delta x(t)] - S[X(t)]}{\epsilon} \quad (18)$$

Hamilton's principle is then formulated as

$$\delta S = 0, \quad \forall \delta x(t) \quad \text{s.t.} \quad \delta x(t_I) = \delta x(t_F) = 0 . \quad (19)$$

This means that we extremize the action among all trajectories that start and end at two specified points x_I and x_F . In the following we will just put $\epsilon = 1$, and drop the $\lim_{\epsilon \rightarrow 0}$, and simply assume that $\delta x(t)$ is already an infinitesimally small variation

The Lagrangian

For mechanical systems, the action functional is expressed as an integral over time:

$$S[x(t)] = \int_{t_I}^{t_F} dt L(x, \dot{x}, t) . \quad (20)$$

The function $L(x, \dot{x}, t)$ inside the integral is called the *Lagrangian* of the system. At any given time t , the Lagrangian $L(x, \dot{x}, t)$ is a function of the position $x(t)$ and the velocity $\dot{x}(t)$ *at that time instant*. In addition it may also have some explicit dependence on t . As will become clear later, the fact that the Lagrangian only depends on x and its first derivative \dot{x} (and not on higher derivatives) is because we observe that the equations of motion of most known systems in Nature only involve up to second derivatives in time.

Euler-Lagrange equations.

The equations of motion come from extremizing the action. Let us work out what are the consequences of imposing (19), with δS as defined in (18). We compute²

$$\begin{aligned} S[X + \delta x] &= \int_{t_I}^{t_F} dt L(X + \delta x, \dot{X} + \delta \dot{x}) dt \\ &= \int_{t_I}^{t_F} dt \left[L(X, \dot{X}) + \delta x(t) \frac{\partial L}{\partial x}(X, \dot{X}) + \delta \dot{x}(t) \frac{\partial L}{\partial \dot{x}}(X, \dot{X}) \right]. \end{aligned} \quad (21)$$

Here we Taylor expanded L . Thus for the variation of S we find

$$\delta S[X] = \int_{t_I}^{t_F} dt \left[\delta x(t) \frac{\partial L}{\partial x} + \delta \dot{x}(t) \frac{\partial L}{\partial \dot{x}} \right] \quad (22)$$

We want to impose the condition that this vanishes for any choice of $\delta x(t)$, that vanishes at the begin and end points. In the above expression, both $\delta x(t)$ and $\delta \dot{x}(t)$ appear; these two functions are not independent. It is better to rewrite the above expression in such a way that only $\delta x(t)$ appears. We can achieve this by integrating by parts:

$$\delta S[X] = \int_{t_I}^{t_F} dt \left[\delta x(t) \frac{\partial L}{\partial x} - \delta x(t) \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} \right] + \frac{\partial L}{\partial \dot{x}} \delta x(t) \Big|_{t_I}^{t_F} . \quad (23)$$

²To avoid cluttering, we temporarily just write X instead of $X(t)$, etc.

Now we extremize among all trajectories such that $x(t_I) = x_I$, $x(t_F) = x_F$, which means $\delta x(t_I) = \delta x(t_F) = 0$. The last boundary term in the above equation therefore vanishes. We are left with:

$$\delta S[X(t)] = \int_{t_I}^{t_F} dt \delta x(t) \left[\frac{\partial L}{\partial x} - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} \right]_{x(t)=X(t)} . \quad (24)$$

Since δS has to vanish for any variation $\delta x(t)$, the factor in brackets in the integrand must vanish. It is indeed clear that if the factor in brackets does not vanish somewhere, then it is always possible to find a function $\delta x(t)$, sharply peaked around that point, such that the integral does not vanish. We end up with the *Euler-Lagrange equations*:

$$\boxed{\left[\frac{\partial L}{\partial x} - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} \right]_{x(t)=X(t)} = 0 .} \quad (25)$$

It is straightforward to generalize this result to the case that the Lagrangian depends on many variables: $L(x_1, \dot{x}_1, \dots, x_n, \dot{x}_n, t)$. These variables could be the three components of the position of a single particle as a vector: $\vec{x} = (x_1, x_2, x_3)$; or they could be the positions of many different particles. Hamilton's principle in this case takes the form

$$\delta S = 0 \quad \forall \delta x_i(t) \quad \text{s.t.} \quad \delta x_i(t_I) = \delta x_i(t_F) = 0 \quad \forall i . \quad (26)$$

This gives rise to the equations:

$$\left[\frac{\partial L}{\partial x_i} - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}_i} \right]_{x(t)=X(t)} = 0, \quad \forall i \quad (27)$$

Momentum and Newton's law

For a given Lagrangian, depending on some set of variables x_i , we will define the *momentum* p_i associated to the variable x_i via

$$\boxed{p_i \equiv \frac{\partial L}{\partial \dot{x}_i}} \quad (28)$$

The Euler-Lagrange equation can then be written as

$$\boxed{\frac{dp_i}{dt} = \frac{\partial L}{\partial x_i}} \quad (\text{E-L eqn}) \quad (29)$$

We can compare this equation with the second law of Newton, $F = ma$, or better (using that $p_i = m\dot{x}_i$):

$$\frac{dp_i}{dt} = F_i \quad (\text{Newton's 2}^{nd}) \quad (30)$$

where F_i denotes the (generalized) force in the direction i . Now suppose that the force F_i is a *conservative* force, that can be derived from a potential energy function $V(x_i)$, via

$$F_i = -\frac{\partial V(x_i)}{\partial x_i} \quad (31)$$

By comparing the eqns (28)-(29) with the more familiar formulas (30)-(31), we are led to make the following guess for a Lagrangian of the single particle moving on one dimension under the influence of a potential $V(x)$:

$$L(x, \dot{x}) = \frac{1}{2} m\dot{x}^2 - V(x) \equiv T - V. \quad (32)$$

In the first term we recognize the *kinetic energy*, $T = \frac{1}{2}mv^2$ with $v = \dot{x}$. The second is the *potential energy*. Notice that the Lagrangian is the difference of the kinetic and potential energy, and not the sum! Thus it is very different from the total energy, $E = T + V$, and should not be confused with it. The Euler-Lagrange equations derived from (32) are:

$$m\ddot{x} = -\frac{\partial V}{\partial x} = F(x). \quad (33)$$

We recognize this as Newton's law. This justifies the use of the Lagrangian formulation: it reproduces the correct laws of mechanics. As we will see, however, it is more powerful and simplifies many problems, and allows a more natural transition from classical mechanics to quantum mechanics.

The generalization to particles in three dimensions and multiple particles is straightforward:

$$L = \sum_{i=1}^n \frac{1}{2} m|\dot{\vec{x}}|^2 - V(\vec{x}_1, \dots, \vec{x}_n). \quad (34)$$

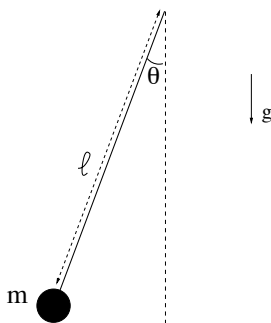
The Euler-Lagrange equations are:

$$m\ddot{\vec{x}}_i = -\frac{\partial V}{\partial \vec{x}_i} = \vec{F}_i. \quad (35)$$

Again, each particle is subject to the conservative force $\vec{F}_i(\vec{x}_1, \dots, \vec{x}_n)$.

Generalized coordinates.

One of the main advantages of the formulation of mechanics in terms of the least action principle, is that it does not refer to any coordinate system. Once we have parameterized the degrees of freedom of our system in terms of our favorite coordinates, and we have expressed the Lagrangian function in terms of them, the fact that a particular trajectory extremizes the action does not depend on the coordinates we use. We are naturally lead to the concept of *generalized coordinates*: the Euler-Lagrange equations have the same form with any coordinate, not just positions in cartesian coordinates.



Let us consider the example of the simple pendulum, consisting of a point mass m attached to a massless rod of length ℓ , moving under the influence of a uniform gravitational acceleration g (see fig). The kinetic and potential energy are:

$$\begin{aligned} T &= \frac{1}{2}mv^2 = \frac{1}{2}m\ell^2\dot{\theta}^2 \\ V &= mgh = -mgl \cos \theta . \end{aligned} \tag{36}$$

The Lagrangian is then simply

$$L = T - V = \frac{1}{2}m\ell^2\dot{\theta}^2 + mgl \cos \theta . \tag{37}$$

The Euler-Lagrange equations are:

$$m\ell^2\ddot{\theta} = -mgl \sin \theta . \tag{38}$$

Let us check whether this agrees with Newton's law. It says that the time derivative of angular momentum is equal to the momentum of the force. The angular momentum is $I\omega = m\ell^2\dot{\theta}$, where $I = m\ell^2$ is the momentum of inertia. We get:

$$\frac{d}{dt}(I\omega) = m\ell^2\ddot{\theta} = lF_{\perp} = -\ell mg \sin \theta . \tag{39}$$

This agrees with the E-L equation.

Energy and momentum conservation

The Lagrangian formulation of mechanics is particularly convenient in case the dynamical system exhibits certain symmetries. We will see that there is a one-to-one correspondence between symmetries of the Lagrangian and conserved quantities. Basic examples of such conserved quantities are energy and momentum.

In general, the Lagrangian may have an explicit dependence on time. For example, the potential energy $V(x)$ and the mass m may depend on time: $L = \frac{1}{2}m(t)\dot{x}^2 - V(x, t)$. In this general case, the total time derivative of the Lagrangian can be expanded, via the chain rule, in terms of partial derivatives as follows

$$\frac{d}{dt}L(x_i, \dot{x}_i, t) = \sum_i \dot{x}_i \frac{\partial L}{\partial x_i} + \sum_i \ddot{x}_i \frac{\partial L}{\partial \dot{x}_i} + \frac{\partial L}{\partial t} \quad (40)$$

The last term is called the *partial time derivative* of L , and differentiates only the *explicit time dependence* of L . The left-hand is the *total time derivative* and differentiates the complete time dependence of L .

In many situations, however, the Lagrangian does not have any explicit time dependence:

$$\frac{\partial L(x_i, \dot{x}_i)}{\partial t} = 0. \quad (41)$$

In this case, we can derive a useful result. We can rewrite the first two terms on the right-hand side of (40), by using eqn (28) and the Euler-Lagrange equation (29), as follows

$$\frac{d}{dt}L(x_i, \dot{x}_i) = \sum_i (\dot{x}_i \dot{p}_i + \ddot{x}_i p_i) = \frac{d}{dt} \left(\sum_i x_i p_i \right)$$

This equation holds when $x(t)$ is a solution of the classical equations of motion. So we find that the quantity

$$E = \sum_i p_i \dot{x}_i - L(x_i, \dot{x}_i) \quad (42)$$

is constant in time

$$\frac{dE}{dt} = 0. \quad (43)$$

This conserved quantity E is the total energy of the system.

Exercise 4: For the lagrangians (32) and (37), show that: $E = T + V$.

In other situations, it may happen that the Lagrangian is independent of one or more of the coordinates x_i :

$$\frac{\partial L}{\partial x_i} = 0 \quad (44)$$

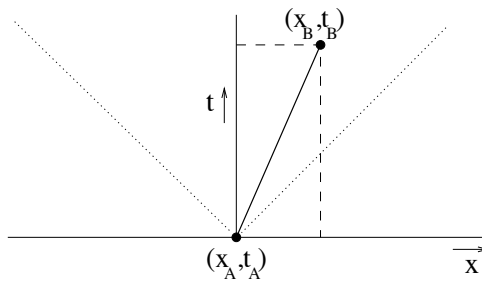
In this case, the E-L equation (29) immediately implies that the corresponding momentum p_i is conserved

$$\frac{dp_i}{dt} = 0 \quad (45)$$

Both energy conservation and momentum conservation can be viewed as consequences of certain *symmetries*. Energy conservation follows if the Lagrangian is invariant under time-translations $t \rightarrow t + \text{const}$. Momentum conservation follows if the Lagrangian is invariant under translations in one (or more) of the coordinates: $x_i \rightarrow x_i + \text{const}$. Since, as discussed above, the x_i are not necessarily cartesian coordinates, but can be generalized coordinates. For example, x_i can also be an angle. The associated conserved momentum does therefore not need to be a linear momentum, but can be a generalized momentum, such as angular momentum.

Relativistic Lagrangian Mechanics

The Lagrangian formulation of Newtonian classical mechanics is naturally generalized to relativistic mechanics. As discussed in class already in the context of the twin paradox, in special relativity in the absence of external force, the classical path of a particle satisfied an extremization principle: it *maximizes* the proper time.



Consider two events $A = (x_A, t_A)$ and $B = (x_B, t_B)$. Let's assume that the two events are time-like separated and that B is in the future of A . This means that it is possible for an observer to follow a space-time path that starts at A and ends at B . In case the observer follows an inertial trajectory, that is, she travels with constant velocity:

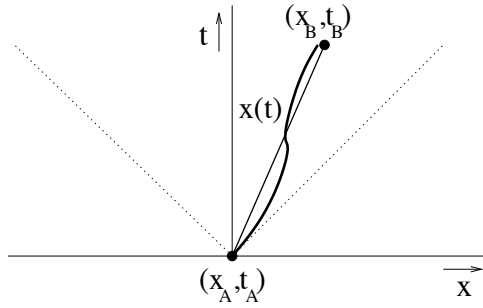
$$X(t) = x_A + v(t - t_A), \quad v = \frac{x_B - x_A}{t_B - t_A}. \quad (46)$$

her clock will indicate that the total time interval between the two events equals (here we put $c = 1$)

$$\tau_{AB} = \frac{t_B - t_A}{\gamma} = \sqrt{(t_B - t_A)^2 - (x_B - x_A)^2} \quad (47)$$

The factor of $1/\gamma = \sqrt{1 - v^2}$ is the relativistic time dilation. The time interval τ_{AB} is called *proper time* between the two events, as measured by the inertial observer. Note that the formula for the proper time distance looks similar to Pythagoras' formula for the distance between two points on a plane, apart from the minus sign.

Now consider another observer that follows some arbitrary path $x(t)$ between A and B :



We wish to write a formula for the total proper time $\tau[x(t)]$ of the path $x(t)$, that is, the total time interval between the two events as measured by the observer following $x(t)$. We proceed in a similar way as before. We choose a parametrization $(x(s), t(s))$ of the space-time path with $0 \leq s \leq 1$ and divide up the s interval into many infinitesimally small steps of size Δs . Let (x_k, t_k) denote the space-time position after k steps. The infinitesimal proper time interval of the k -th step equals

$$\Delta\tau_k = \sqrt{(\Delta t_k)^2 - (\Delta x_k)^2} = \Delta s \sqrt{\left(\frac{\Delta t_k}{\Delta s}\right)^2 - \left(\frac{\Delta x_k}{\Delta s}\right)^2} \quad (48)$$

where $\Delta x_k = x_{k+1} - x_k$ and $\Delta t_k = t_{k+1} - t_k$. The total proper time of the path is found by summing all the $\Delta\tau_k$. In the limit where the step-size Δs is taken to zero and the number of steps to infinity, so that the total length of all the steps combined is kept fixed, $\sum_k \Delta s = 1$, the sum becomes an integral

$$\tau[x(t)] = \int_0^1 ds \sqrt{\left(\frac{dt}{ds}\right)^2 - \left(\frac{dx}{ds}\right)^2} \quad (49)$$

This expression for the total proper time of the path in space-time should be compared with the formula (8) of the total length of a path in 3-d space. The proper time can be thought of as the length of the path $(x(s), t(s))$ in Minkowski space-time.

We can write the result (49) in a somewhat more compact form, by changing integration variables from s to t . This yields

$$\boxed{\tau[x(t)] = \int_{t_A}^{t_B} dt \sqrt{1 - \dot{x}^2}} \quad (50)$$

with $\dot{x} = dx/dt$.

Now our set up is like the twin paradox. Suppose that two twin sisters have the exact same age at $t = t_A$, but one of them travels the inertial path $X(t)$, given in eqn (??), and the other one travels along the arbitrary non-inertial path $x(t)$. They meet again at (x_B, t_B) . By this time, the inertial twin will be aged by an amount τ_{AB} given in (47), while the non-inertial twin will have aged by the amount $\tau[x(t)]$ given in (50). Let us view the situation from the point of view of the inertial twin. She will see that the clock of her twin sister slowed down during the time that the twin sister was traveling, due to the time dilation. So by the time the twins meet, the non-inertial twin will be younger than the inertial twin. In other words:

$$\tau[x(t)] \leq \tau_{AB}. \quad (51)$$

The inertial path $X(t)$ is the path that *maximizes the proper time*:

$$\delta\tau[x(t)] \Big|_{x(t)=X(t)} = 0 \quad (52)$$

This gives us an important hint for how to write an action for a free relativistic particle of mass m : it is a natural guess that $S[x(t)]$ is equal to the proper time $\tau[x(t)]$, times some suitable constant. The action in non-relativistic mechanics has the dimension of *time* \times *energy*. So the natural choice for S is (here we've re-inserted the speed of light c)

$$\boxed{S[\vec{x}(t)] = -mc^2 \int_{t_I}^{t_F} dt \sqrt{1 - \left(\frac{\dot{\vec{x}}}{c}\right)^2}} \quad (53)$$

So the Lagrangian reads

$$L(\vec{x}, \dot{\vec{x}}) = -mc^2 \sqrt{1 - \beta^2}, \quad \beta^2 \equiv \left(\frac{\dot{\vec{x}}}{c}\right)^2. \quad (54)$$

The free Lagrangian does not depend on the positions x . The Euler-Lagrange equations thus reduce to the statement of momentum conservation

$$\frac{d\vec{p}}{dt} = 0, \quad (55)$$

where

$$\vec{p} = \frac{\partial L}{\partial \dot{\vec{x}}} = \frac{m\dot{\vec{x}}}{\sqrt{1-\beta^2}} = \gamma m \dot{\vec{x}}. \quad (56)$$

This is indeed the standard relativistic expression for the momentum of a particle of mass m . In addition, because L does not explicitly depend on t , we derive that energy is conserved

$$\frac{dE}{dt} = 0, \quad E = \vec{x} \cdot \vec{p} - L \quad (57)$$

A simple computation gives that

$$E = \frac{mc^2}{\sqrt{1-\beta^2}} = \gamma mc^2 \quad (58)$$

which is the most famous formula in all of physics. So the formula (53) passes all the necessary tests: it gives rise to the right equations of motion, and to the right expression for the energy and momentum.

A relativistic particle can be placed inside of a potential $V(x)$. (E.g. we can imagine that the particle has a charge q , and turn on a static electric field $\vec{E}(x) = -\vec{\nabla}\phi(x)$, where $\phi(x)$ is the electrostatic potential. In this case, the potential energy equals $V(x) = q\phi(x)$.) The Lagrangian then reads

$$L(\vec{x}, \dot{\vec{x}}) = -mc^2\sqrt{1-\beta^2} - V(\vec{x}) \quad (59)$$

The E-L equation of motion looks like Newton's law

$$\frac{d\vec{p}}{dt} = -\frac{\partial V(x)}{\partial \vec{x}} = \vec{F}(x) \quad (60)$$

but the momentum \vec{p} is defined via the relativistic formula (56). The total energy is

$$E = \vec{x} \cdot \vec{p} - L = \gamma mc^2 + V(\vec{x}) \quad (61)$$