Calculus of Variations

Joel G. Broida
University of Colorado, Boulder

October 1, 2009

Calculus of Variations

1 Functional Derivatives

The fundamental equation of the calculus of variations is the Euler-Lagrange equation

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

There are several ways to derive this result, and we will cover three of the most common approaches. Our first method I think gives the most intuitive treatment, and this will then serve as the model for the other methods that follow.

To begin with, recall that a (real-valued) **function** on $\mathbb{R}^n$ is a mapping $f : U \subset \mathbb{R}^n \to \mathbb{R}$. In other words, f takes a *point* in some subset U of $\mathbb{R}^n$ and gives back a number, i.e., a point in $\mathbb{R}$. In particular, the domain of f is a subset of $\mathbb{R}^n$. We write this mapping as $f(\mathbf{x})$.

In contrast to this, a **functional** F is a "function" whose domain is the space of *curves* in $\mathbb{R}^n$, and hence it depends on the *entire curve*, not just a single point. Very loosely speaking, we will take a **curve** to be a differentiable mapping $y : U \subset \mathbb{R}^n \to \mathbb{R}^m$. So a curve is just a function defined on some interval, and a functional is a "function of a function."

For example, let $y(x)$ be a real valued curve defined on the interval $[x_1, x_2] \subset \mathbb{R}$. Then we can define a functional $F[y]$ by

$$F[y] := \int_{x_1}^{x_2} [y(x)]^2 \, dx \in \mathbb{R}.$$

(The notation $F[y]$ is the standard way to denote a functional.) So a functional is a mapping from the space of curves into the real numbers.

We now want to define the derivative of such a functional. There are several ways to go about this, and we will take the most intuitive approach that is by analogy with the usual notion of derivative.

So, let $f(t)$ be a function of a single real variable, and recall the definition of the derivative $f'(t)$:

$$f'(t) = \frac{df}{dt}(t) = \lim_{h \to 0} \frac{f(t+h) - f(t)}{h}. \tag{1}$$

This is equivalent to saying that f is differentiable at t if there exists some number L (called the **derivative** of f at t) and a function φ with the property that

$$\lim_{h \to 0} \frac{\varphi(h)}{h} = 0$$

such that

$$f(t+h) = f(t) + Lh + \varphi(h). \tag{2}$$

Before proving the equivalence of these formulations, let me make two remarks. First, we say that such a function $\varphi(h)$ is $\mathcal{O}(h^2)$ (**order** h^2). And second, note that the number L is just a linear map from $\mathbb{R}$ to $\mathbb{R}$. (In this case, $L : \mathbb{R} \to \mathbb{R}$ is defined by $L(h) = Lh$ for $h \in \mathbb{R}$.) In fact, it is this formulation of the derivative that is used to generalize differentiation to functions from $\mathbb{R}^n$ to $\mathbb{R}^m$, in which case the linear map L becomes the Jacobian matrix $(\partial y^i/\partial x^j)$.

Let us now show that equations (1) and (2) are equivalent. Note that if we start from (1) and *define* the function φ by

$$\varphi(h) = \begin{cases} f(t+h) - f(t) - f'(t)h & \text{for } h \neq 0 \\ 0 & \text{for } h = 0 \end{cases}$$

then

$$f(t+h) = f(t) + Lh + \varphi(h)$$

where $L = f'(t)$ and (by equation (1)) $\lim \varphi(h)/h = 0$. Conversely, if we start from equation (2), then

$$\frac{f(t+h) - f(t)}{h} = L + \frac{\varphi(h)}{h}$$

and taking the limit as $h \to 0$ we see that $f'(t) = L$.

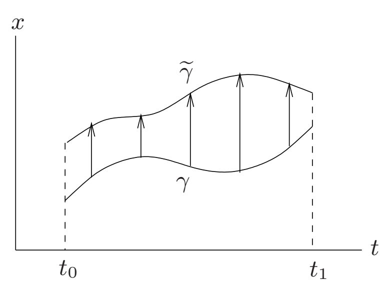
Now let us return to functionals. Let γ be a curve in the plane:

$$\gamma = \{(t, x) : x(t) = x \text{ for } t_0 < t < t_1\}.$$

Let $\widetilde{\gamma}$ be an approximation to γ, i.e.,

$$\widetilde{\gamma} = \{(t, x) : x = x(t) + h(t)\}$$

for some function $h(t)$. We abbreviate this by $\widetilde{\gamma} = \gamma + h$. Let F be a functional and consider the difference $F[\widetilde{\gamma}] - F[\gamma] = F[\gamma + h] - F[\gamma]$.



We say that F is **differentiable** if there exists a linear map L (i.e., for fixed γ we have $L(h_1+h_2) = L(h_1)+L(h_2)$ and $L(ch) = cL(h)$) and a remainder $R(h, \gamma)$ with the property that $R(h, \gamma) = \mathcal{O}(h^2)$ (i.e., for $|h| < \varepsilon$ and $|h'| = |dh/dt| < \varepsilon$ we have $|R| < \text{const} \cdot \varepsilon^2$) such that

$$F[\gamma + h] - F[\gamma] = L(h) + R(h, \gamma) \tag{3}$$

The linear part of equation (3), $L(h)$, is called the **differential** of F.

We now want to prove the following theorem. As is common, we will denote the derivative with respect to t by a dot, although in this case t is not necessarily the time – it is simply the independent variable.

Theorem 1. *Let γ be a curve in the plane, and let $f = f(x(t), \dot{x}(t), t)$ be a differentiable function. Then the functional*

$$F[\gamma] = \int_{t_0}^{t_1} f(x(t), \dot{x}(t), t)\, dt$$

is differentiable and its derivative is given by

$$L(h) = \int_{t_0}^{t_1} \left[\frac{\partial f}{\partial x} - \frac{d}{dx}\left(\frac{\partial f}{\partial \dot{x}}\right)\right] h\, dt + \frac{\partial f}{\partial \dot{x}} h \bigg|_{t_0}^{t_1} \tag{4}$$

Proof. Since f is a differentiable function we have (using equation (2) in the case where f is a function of the two variables x and $\dot{x}$)

$$\begin{aligned} F[\gamma + h] - F[\gamma] &= \int_{t_0}^{t_1} [f(x + h, \dot{x} + \dot{h}, t) - f(x, \dot{x}, t)]\, dt \\ &= \int_{t_0}^{t_1} \left(\frac{\partial f}{\partial x} h + \frac{\partial f}{\partial \dot{x}} \dot{h}\right) dt + \mathcal{O}(h^2) \\ &:= L(h) + R(h, \gamma) \end{aligned}$$

where we have defined $L(h) = \int_{t_0}^{t_1} [(\partial f/\partial x)h + (\partial f/\partial \dot{x})\dot{h}]\, dt$ and $R(h, \gamma) = \mathcal{O}(h^2)$. By equation (3) this shows that F is differentiable.

Integrating the second term of $L(h)$ by parts we have

$$\int_{t_0}^{t_1} \frac{\partial f}{\partial \dot{x}} \frac{dh}{dt}\, dt = \frac{\partial f}{\partial \dot{x}} h \bigg|_{t_0}^{t_1} - \int_{t_0}^{t_1} \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}}\right) h\, dt$$

and therefore

$$L(h) = \int_{t_0}^{t_1} \left[\frac{\partial f}{\partial x} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}}\right)\right] h\, dt + \frac{\partial f}{\partial \dot{x}} h \bigg|_{t_0}^{t_1}. \qquad \blacksquare$$

2 The Euler-Lagrange Equation

Before proving our main result (the Euler-Lagrange equation) we need a lemma. Note that the function $h(t)$ in this lemma must be completely arbitrary (other than being continuous).

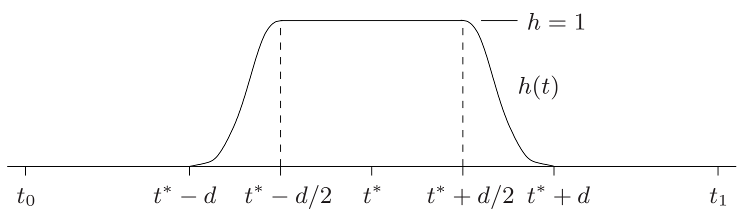
Lemma. *If a continuous function $f(t)$ defined on the interval $[t_0, t_1]$ satisfies $\int_{t_0}^{t_1} f(t)h(t)\,dt = 0$ for* any *continuous function $h(t)$ with $h(t_0) = h(t_1) = 0$, then $f(t) = 0$ for all $t \in [t_0, t_1]$.*

Proof. First note that this makes sense intuitively because if you choose $h(t) \neq 0$ only over a very small interval, then the integral essentially picks out $f(t)$ only over this interval, and therefore $f(t)$ must equal 0 on this very small interval. However, since the interval is arbitrary, it must be that $f = 0$ everywhere.

(A formal way to show this is to let $h(t)$ be the Dirac delta function $\delta(t-t^*)$ so that $f(t^*) = 0$ for all $t^* \in (t_0, t_1)$. But we aren't going to assume any knowledge of the Dirac delta at this point.)

Now for the proof. Assume that f is not identically 0, and there is no loss of generality in assuming that there exists $t^* \in (t_0, t_1)$ such that $f(t^*) > 0$. We first show that if a continuous function is nonzero at a point, then it is nonzero in a neighborhood of that point. Recall that to say f is continuous at t^* means that given $\varepsilon > 0$, there exists $\delta > 0$ such that for all t with $|t^* - t| < \delta$ we have $|f(t^*) - f(t)| < \varepsilon$. So if $0 < f(t^*)$, then choosing $0 < \varepsilon < f(t^*)$ shows that you can't have $f(t) = 0$ if $|t^* - t| < \delta$. Therefore, since f is continuous, there exists a neighborhood Δ of t^* such that $f(t) > c = \text{constant} > 0$ for all $t \in \Delta$.

Write $\Delta = (t^* - d, t^* + d) \subset (t_0, t_1)$ and let $\Delta/2$ denote the interval $\Delta/2 = (t^* - d/2, t^* + d/2)$. Choose the function $h(t)$ so that $h(t) = 0$ outside Δ, $h(t) > 0$ on Δ, and $h(t) = 1$ on $\Delta/2$.



Since $h = 0$ outside Δ we have $\int_{t_0}^{t_1} fh\,dt = \int_\Delta fh\,dt$. And since $f > c > 0$ for all $t \in \Delta$ it follows that

$$\int_{t_0}^{t_1} f(t)h(t)\,dt = \int_\Delta f(t)h(t)\,dt \geq c \cdot 1 \cdot \frac{d}{2} \cdot 2 = cd > 0.$$

But this contradicts the hypothesis that $\int_{t_0}^{t_1} fh\,dt = 0$ and hence it must be that $f(t^*) = 0$ for all $t^* \in (t_0, t_1)$, i.e., $f(t) = 0$ for all $t \in [t_0, t_1]$. ∎

Let me make one remark. The function $h(t)$ defined in the proof of the preceding theorem is sometimes called a "bump function," and such functions are used, for example, in more general integration theory to prove the existence of partitions of unity. As a specific example, let us show that there exists a C^∞

function $h(t)$ that equals 1 on $[-1, 1]$ and 0 on the complement of $(-2, 2)$. (A C^∞ function is one that is infinitely differentiable.) To see this, first let

$$f(t) = \begin{cases} e^{-1/t} & \text{for } t > 0 \\ 0 & \text{for } t \leq 0 \end{cases} .$$

Now let

$$g(t) = \frac{f(t)}{f(t) + f(1-t)}$$

and note that $g(t) = 1$ for $t \geq 1$. Finally, define $h(t) = g(t+2)g(2-t)$. I leave it to you to show that this has the desired properties.

We are now in a position to prove our main result. First some terminology. We say that an **extremal** of a differentiable functional $F[\gamma]$ is a curve γ such that $L(h) = 0$ for all h. (This is like saying x_0 is a stationary point of a function $f(x)$ if $f'(x_0) = 0$.) However, note that if a curve γ is an extremal of F, we don't know whether or not F takes its maximum or minimum (or neither) value on γ. This is analogous to the usual case in calculus where we have to evaluate the second derivative of a function g to decide whether a point x_0 where $g'(x_0) = 0$ is a minimum, maximum or inflection point. Fortunately, in most cases of physical interest we are looking for a minimum, and it will usually be clear from the problem that we have found the desired result.

Theorem 2 (**Euler-Lagrange Equation**). *A curve $\gamma : x = x(t)$ is an extremal of the functional*

$$F[\gamma] = \int_{t_0}^{t_1} f(x(t), \dot{x}(t), t)\, dt$$

on the space of curves passing through the points $x_0 = x(t_0), x_1 = x(t_1)$ if and only if

$$\frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}}\right) - \frac{\partial f}{\partial x} = 0 \tag{5}$$

along the curve $x(t)$.

Proof. When we refer to the space of curves passing through the points x_0 and x_1 we mean that $h(t_0) = h(t_1) = 0$ for all (differentiable) functions $h(t)$. By Theorem 1 we have

$$L(h) = \int_{t_0}^{t_1} \left[\frac{\partial f}{\partial x} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}}\right)\right] h\, dt + \frac{\partial f}{\partial \dot{x}} h \Big|_{t_0}^{t_1} .$$

The second term on the right hand side is the boundary term, and it vanishes by hypothesis. We also have $L(h) = 0$ for all h since γ is an extremal. But then

$$L(h) = \int_{t_0}^{t_1} \left[\frac{\partial f}{\partial x} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}}\right)\right] h\, dt = 0$$

for all h so that by the lemma we have the Euler-Lagrange equation

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

This is clearly equivalent to

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

Conversely, if the Euler-Lagrange equation holds, then clearly $L(h) = 0$ so that γ is an extremal. ∎

It is worth emphasizing that this is a differential equation for $x(t)$ (as opposed to f) since the function f is known. In fact, in a later section we will see that this Euler-Lagrange equation is a second-order differential equation for $x(t)$ (which can be reduced to a first-order equation in the special case that $\partial f/\partial t = 0$, i.e., that f has no explicit dependence on the independent variable t).

So far we have considered only functions of the form $f = f(x(t), \dot{x}(t), t)$. We can easily generalize this to functions of several *dependent* variables $\{x^1, \dots, x^n\}$:

$$f = f(x^1(t), \dot{x}^1(t), x^2(t), \dot{x}^2(t), \dots, x^n(t), \dot{x}^n(t)).$$

Now we have n curves $\gamma^i : x^i = x^i(t)$ and $\widetilde{\gamma}^i = x^i + h^i$ so that

$$\begin{aligned}
F[\gamma^1 + h^1, \dots, \gamma^n + h^n] &- F[\gamma^1, \dots, \gamma^n] \\
&= \int_{t_0}^{t_1} [f(x^1 + h^1, \dots, x^n + h^n) - f(x^1, \dots, x^n)]\, dt \\
&= \int_{t_0}^{t_1} \sum_{i=1}^n \left(\frac{\partial f}{\partial x^i} h^i + \frac{\partial f}{\partial \dot{x}^i} \dot{h}^i\right) dt + \sum_{i=1}^n O\big((h^i)^2\big)
\end{aligned}$$

Again, assume that the boundary terms vanish when we integrate by parts, and we are left with

$$L(h^1, \dots, h^n) = \int_{t_0}^{t_1} \sum_{i=1}^n \left[\frac{\partial f}{\partial x^i} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}^i}\right)\right] h^i\, dt.$$

If we now assume that the variations h^i are all independent (so this ignores constraints), then we conclude that we will have an extremal if

$$\frac{\partial f}{\partial x^i} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}^i}\right) = 0 \qquad \text{for each } i = 1, \dots, n. \tag{6}$$

In other words, the Euler-Lagrange equation applies to each variable (or coordinate) separately.

Example 1. Our first example will be to show that the shortest distance between two points in the plane is a straight line. Thus we want to minimize the arc length

$$\int_{(x_1,y_1)}^{(x_2,y_2)} ds = \int_{(x_1,y_1)}^{(x_2,y_2)} \sqrt{dx^2 + dy^2} = \int_{(x_1,y_1)}^{(x_2,y_2)} \sqrt{1 + y'^2}\, dx$$

where ds is an infinitesimal element of arc length in the plane.

The independent variable is x, the dependent variable is y, and we have $f(y, y', x) = \sqrt{1 + y'^2}$. Then $\partial f/\partial y = 0$ so that

$$\frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) = 0$$

which implies

$$\frac{\partial f}{\partial y'} = \frac{y'}{\sqrt{1 + y'^2}} = \text{const} := c.$$

Squaring this expression yields $y'^2(1 - c^2) = c^2 > 0$ so we must have $c^2 < 1$. Then

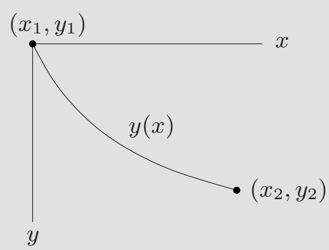
$$y' = \frac{c}{\sqrt{1 - c^2}} := m$$

so integrating this show that $y = mx + b$ where b is the constant of integration.

We will shortly generalize this result to curves in arbitrary metric spaces.

Example 2. Our next example is a famous problem called the **brachistochrone**, and it is the following. Consider a particle moving in a constant gravitational field, starting from rest at the point (x_1, y_1) and falling to a lower point (x_2, y_2). We want to find the path the particle should take to minimize the time of travel (in the absence of friction).

For convenience, we choose the coordinates as shown below, with the origin as the initial location of the particle.



From $\mathbf{F} = m\mathbf{g} = -\nabla V$ we have $V(y) = -mgy$, so by conservation of energy it follows that $0 = (1/2)mv^2 - mgy$ or $v = \sqrt{2gy}$. Denoting an infinitesimal distance in the plane by ds, the time of travel is given by

$$\tau = \int_{(x_1,y_1)}^{(x_2,y_2)} \frac{ds}{v} = \int_{(x_1,y_1)}^{(x_2,y_2)} \frac{\sqrt{dx^2+dy^2}}{v} = \int_{(x_1,y_1)}^{(x_2,y_2)} \frac{\sqrt{x'^2+1}}{\sqrt{2gy}}\,dy$$

$$= \frac{1}{\sqrt{2g}} \int_{(x_1,y_1)}^{(x_2,y_2)} \sqrt{\frac{x'^2+1}{y}}\,dy$$

where we are considering $x = x(y)$ to be a function of y so that $x' = dx/dy$.

The Euler-Lagrange equation (5) is to be applied to the integrand

$$f(x, x', y) = \sqrt{\frac{x'^2+1}{y}}$$

where now y is the independent variable. Because $\partial f/\partial x = 0$, we see that $(d/dy)(\partial f/\partial x') = 0$ or

$$\frac{\partial f}{\partial x'} = \frac{1}{\sqrt{y}} \frac{x'}{\sqrt{x'^2+1}} = \text{const} := \sqrt{\frac{1}{2a}}$$

and therefore

$$\frac{x'^2}{y(x'^2+1)} = \frac{1}{2a}.$$

Solving this for x' and integrating we find

$$x = \int \sqrt{\frac{y}{2a-y}}\,dy = \int \frac{y}{\sqrt{2ay-y^2}}\,dy.$$

Making the substitution $y = a(1-\cos\theta)$ so that $dy = a\sin\theta\,d\theta$, this becomes

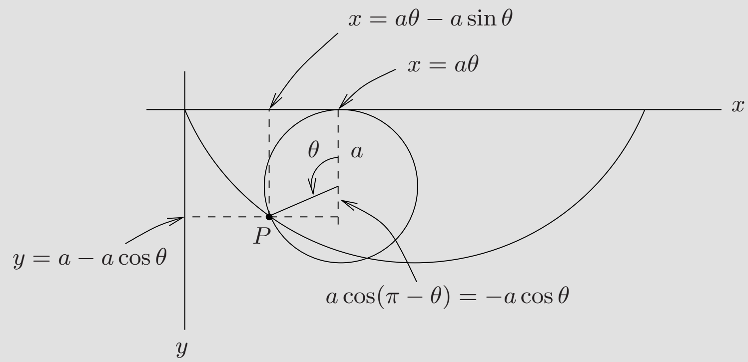
$$x = \int a(1-\cos\theta)\,d\theta = a(\theta - \sin\theta) + \text{const}$$

Our initial condition is $x = y = 0$ which is equivalent to $\theta = 0$, and this shows that the constant of integration is also zero. Therefore the solution to our problem is

$$x = a(\theta - \sin\theta) \quad \text{and} \quad y = a(1-\cos\theta)$$

where the constant a is chosen so that the path passes through the point (x_2, y_2).

These equations are the parametric equations of a curve called a **cycloid**. This is the curve traced out by a point on the rim of a wheel of radius a that is rolling along the underside of the x-axis. See the figure below.



An interesting physical aspect of the cycloid is this: a particle released from rest at any point P will take an amount of time to reach the bottom point of the curve (i.e., $y = 2a$) that is independent of P. I leave the proof of this fact to you.

Example 3. One of the fundamental principles of classical mechanics is called **Hamilton's Principle**: If

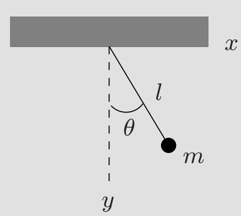
$$L(x, \dot{x}, t) = T(x, \dot{x}) - V(x, t)$$

is the Lagrangian of the system, then the system moves from time t_1 to t_2 in such a way that the integral

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

is an extremum with respect to the functions $x(t)$ where the endpoints $x(t_1)$ and $x(t_2)$ are fixed.

As an easy example, consider the simple plane pendulum:



The kinetic energy is

$$T = \frac{1}{2}mv^2 = \frac{1}{2}ml^2\dot{\theta}^2$$

and the potential energy is given by

$$V = mgl(1 - \cos\theta)$$

where we have defined V to be zero when the mass is at its lowest point. Then the Lagrangian is

$$L = T - V = \frac{1}{2}ml^2\dot{\theta}^2 - mgl(1 - \cos\theta)$$

where the independent variable is t and the dependent variable is θ.

To find the Euler-Lagrange equation we compute

$$\frac{\partial L}{\partial \theta} = -mgl\sin\theta \quad \text{and} \quad \frac{\partial L}{\partial \dot{\theta}} = ml^2\dot{\theta}$$

so we find from equation (5)

$$ml^2\ddot{\theta} + mgl\sin\theta = 0$$

or

$$\ddot{\theta} + \omega^2\sin\theta = 0$$

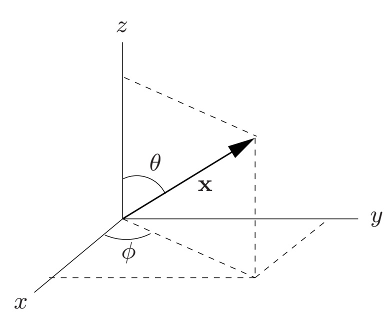
where $\omega^2 := g/l$. In the case that we consider only small oscillations $\theta \ll 1$ we have $\sin\theta \approx \theta$ so our equation becomes $\ddot{\theta} + \omega^2\theta = 0$ with the general solution

$$\theta(t) = A\cos(\omega t + \delta).$$

3 The Geodesic Equation

We now turn to the more general problem mentioned above of finding the equation of the shortest path between two points in what is called a **semi-Riemannian** space. This is a generalization of the usual metric space $\mathbb{R}^3$ with the Pythagorean notion of distance, and will be more carefully defined below. To motivate this definition, we first look at some common special cases that are easy to visualize.

The first thing to formulate is how to find an infinitesimal displacement $d\mathbf{x}$ in a curvilinear coordinate system. Let us consider the usual spherical coordinates as an example.



Writing $\|\mathbf{x}\| = r$, the position vector $\mathbf{x}$ has (x, y, z) coordinates

$$\mathbf{x} = (r \sin\theta \cos\phi, r \sin\theta \sin\phi, r \cos\theta).$$

If we let u^i stand for the ith coordinate of a general curvilinear coordinate system, then a unit vector in the u^i direction is by definition

$$\hat{\mathbf{u}}^i = \frac{\partial \mathbf{x}/\partial u^i}{\|\partial \mathbf{x}/\partial u^i\|}.$$

For our spherical coordinates we have for r:

$$\frac{\partial \mathbf{x}}{\partial r} = (\sin\theta \cos\phi, \sin\theta \sin\phi, \cos\theta)$$

and

$$\left\|\frac{\partial \mathbf{x}}{\partial r}\right\| = \left\langle \frac{\partial \mathbf{x}}{\partial r}, \frac{\partial \mathbf{x}}{\partial r} \right\rangle^{1/2} = 1$$

so that

$$\hat{\mathbf{r}} = (\sin\theta \cos\phi, \sin\theta \sin\phi, \cos\theta) \qquad \text{and} \qquad \frac{\partial \mathbf{x}}{\partial r} = \hat{\mathbf{r}}.$$

For θ:

$$\frac{\partial \mathbf{x}}{\partial \theta} = (r \cos\theta \cos\phi, r \cos\theta \sin\phi, -r \sin\theta)$$

and

$$\left\|\frac{\partial \mathbf{x}}{\partial \theta}\right\| = \left\langle \frac{\partial \mathbf{x}}{\partial \theta}, \frac{\partial \mathbf{x}}{\partial \theta} \right\rangle^{1/2} = r$$

so that

$$\hat{\boldsymbol{\theta}} = (\cos\theta \cos\phi, \cos\theta \sin\phi, -\sin\theta) \qquad \text{and} \qquad \frac{\partial \mathbf{x}}{\partial \theta} = r\,\hat{\boldsymbol{\theta}}.$$

For ϕ:

$$\frac{\partial \mathbf{x}}{\partial \phi} = (-r \sin\theta \sin\phi, r \sin\theta \cos\phi, 0)$$

and

$$\left\|\frac{\partial \mathbf{x}}{\partial \phi}\right\| = \left\langle \frac{\partial \mathbf{x}}{\partial \phi}, \frac{\partial \mathbf{x}}{\partial \phi} \right\rangle^{1/2} = r \sin\theta$$

so that

$$\hat{\boldsymbol{\phi}} = (-\sin\phi, \cos\phi, 0) \qquad \text{and} \qquad \frac{\partial \mathbf{x}}{\partial \phi} = r \sin\theta\,\hat{\boldsymbol{\phi}}.$$

Putting this all together we see that

$$d\mathbf{x} = \frac{\partial \mathbf{x}}{\partial r}\,dr + \frac{\partial \mathbf{x}}{\partial \theta}\,d\theta + \frac{\partial \mathbf{x}}{\partial \phi}\,d\phi$$

or

$$d\mathbf{x} = \hat{\mathbf{r}}\,dr + \hat{\boldsymbol{\theta}}\,r\,d\theta + \hat{\boldsymbol{\phi}}\,r \sin\theta\,d\phi.$$

While this was the "right" way to derive this result, there is a quick way that is usually easiest. In this method, we hold two of the three variables constant and vary the third, and see what $d\mathbf{x}$ is for that variation. So, first hold θ, ϕ constant and vary r to obtain $d\mathbf{x} = dr\,\hat{\mathbf{r}}$ (look at the figure above, and let $\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}}$ be unit vectors in the appropriate directions). Next, hold r, ϕ constant and vary θ to see that $d\mathbf{x} = r\,d\theta\,\hat{\boldsymbol{\theta}}$. Finally, hold r, θ constant and vary ϕ to obtain $d\mathbf{x} = r\sin\theta\,d\phi\,\hat{\boldsymbol{\phi}}$. Putting these together we again find

$$d\mathbf{x} = dr\,\hat{\mathbf{r}} + r\,d\theta\,\hat{\boldsymbol{\theta}} + r\sin\theta\,d\phi\,\hat{\boldsymbol{\phi}}.$$

Note also that if we treat each of these three variations as the edge of a small cube, then the volume element for spherical coordinates is seen to be the product of these displacements and we have the well know result $d^3x = r^2 \sin\theta dr d\theta d\phi$.

Returning to the unit vectors $\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\phi}}$ derived above, it is also not hard to see that these form an orthonormal set. Indeed, they are normalized by construction, and by direct calculation it is easy to see that they are orthogonal (e.g., $\hat{\mathbf{r}} \cdot \hat{\boldsymbol{\theta}} = 0$). Alternatively, we can note that $\hat{\mathbf{r}}$ points in a radial direction, while $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\phi}}$ are both tangent to a sphere as well as orthogonal to each other.

In a similar (but much easier) manner, it is easy to see that in the plane $\mathbb{R}^2$ we have the cartesian coordinate expression

$$d\mathbf{x} = dx\,\hat{\mathbf{x}} + dy\,\hat{\mathbf{y}}$$

as well as the polar coordinate expression

$$d\mathbf{x} = dr\,\hat{\mathbf{r}} + r\,d\theta\,\hat{\boldsymbol{\theta}}$$

By definition, the element of distance ds (or **line element**) is given by

$$ds^2 = \langle d\mathbf{x}, d\mathbf{x} \rangle$$

so for the three examples given above we have (since all basis vectors are orthonormal):

$$ds^2 = dx^2 + dy^2 \qquad \text{for cartesian coordinates in } \mathbb{R}^2$$

$$ds^2 = dr^2 + r^2 d\theta^2 \qquad \text{for polar coordinates in } \mathbb{R}^2$$

$$ds^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta\,d\phi^2 \qquad \text{for spherical polar coordinates.}$$

If we label our coordinates by x^i and the (not generally orthonormal) basis vectors by e_i (e.g., $(x^1, x^2, x^3) = (r, \theta, \phi)$ and $e_1 = \hat{\mathbf{r}}, e_2 = r\hat{\boldsymbol{\theta}}, e_3 = r\sin\theta\hat{\boldsymbol{\phi}}$), then using the summation convention we can write all of these line elements as

$$ds^2 = \langle d\mathbf{x}, d\mathbf{x} \rangle = \langle e_i dx^i, e_j dx^j \rangle = \langle e_i, e_j \rangle dx^i dx^j$$

or simply

$$ds^2 = g_{ij} dx^i dx^j \tag{7}$$

where the *symmetric* matrix (g_{ij}) with components defined by

$$g_{ij} := \langle e_i, e_j \rangle = g_{ji}$$

is called the **metric**.

The metric is both symmetric and diagonal (since we almost always use an orthonormal basis), and in our three examples it takes the forms

$$(g_{ij}) = \begin{bmatrix} 1 & \\ & 1 \end{bmatrix} \qquad \text{for } ds^2 = dx^2 + dy^2$$

$$(g_{ij}) = \begin{bmatrix} 1 & \\ & r^2 \end{bmatrix} \qquad \text{for } ds^2 = dr^2 + r^2 d\theta^2$$

$$(g_{ij}) = \begin{bmatrix} 1 & & \\ & r^2 & \\ & & r^2 \sin^2\theta \end{bmatrix} \qquad \text{for } ds^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta \, d\phi^2$$

Since the metric is diagonal, it is obvious that the inverse metric is given by $(g^{ij}) := (g_{ij})^{-1}$ where $g^{ii} = 1/g_{ii}$. In other words,

$$g^{ij} g_{jk} = \delta^i_k. \tag{8}$$

In particular we have

$$(g^{ij}) = \begin{bmatrix} 1 & \\ & 1 \end{bmatrix} \qquad \text{for } ds^2 = dx^2 + dy^2$$

$$(g^{ij}) = \begin{bmatrix} 1 & \\ & 1/r^2 \end{bmatrix} \qquad \text{for } ds^2 = dr^2 + r^2 d\theta^2$$

$$(g^{ij}) = \begin{bmatrix} 1 & & \\ & 1/r^2 & \\ & & 1/r^2 \sin^2\theta \end{bmatrix} \qquad \text{for } ds^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta \, d\phi^2$$

A **Riemannian space** is a metric space with a *positive definite* metric (g_{ij}) defined on it. (Equivalently, the inner product is **positive definite**, i.e., $\langle u, u \rangle \geq 0$ and $\langle u, u \rangle = 0$ if and only if $u = 0$.) In other words, $g_{ij} = 0$ for $i \neq j$ and $g_{ii} > 0$. In a **semi-Riemannian** space, we remove the requirement that the diagonal entries g_{ii} be greater than zero. (An equivalent way to say this is that the inner product is **nondegenerate**, i.e., $\langle u, v \rangle = 0$ for *all* v if and only if $u = 0$.) For example, the usual Lorentz metric η_{ij} of special relativity takes the form $ds^2 = dt^2 - d\mathbf{x}^2$. This is not positive definite because any nonzero light-like (or null) vector has zero length.

We are now ready to generalize Example 1. We want to find the equation for the path of shortest length between two points in a Riemannian or semi-Riemannian space. In other words, we want to minimize the path length

$$\int ds = \int (g_{ij} dx^i dx^j)^{1/2} = \int \left(g_{ij} \frac{dx^i}{dt} \frac{dx^j}{dt} \right)^{1/2} dt$$

where $x^i = x^i(t)$ and t is an arbitrary curve parameter. Note also that in general $g_{ij} = g_{ij}(x)$.

However, the presence of the square root in the integrand makes things way too messy. It is far easier to realize that whatever path extremizes $\int ds = \int (ds/dt)dt$ will also extremize $\int (ds/dt)^2 dt$. Thus, let us use equation (6) to find the path that extremizes the integral

$$\int ds = \int g_{ij}\dot{x}^i\dot{x}^j \, dt.$$

In general, the path of shortest length between two points is called a **geodesic**, and the resulting equation for $x(t)$ is called the **geodesic equation**.

Letting $f = g_{ij}\dot{x}^i\dot{x}^j$ we have

$$\begin{aligned}
\frac{\partial f}{\partial x^k} &= \frac{\partial g_{ij}}{\partial x^k}\dot{x}^i\dot{x}^j := g_{ij,k}\dot{x}^i\dot{x}^j \\
\frac{\partial f}{\partial \dot{x}^k} &= g_{ij}\frac{\partial \dot{x}^i}{\partial \dot{x}^k}\dot{x}^j + g_{ij}\dot{x}^i\frac{\partial \dot{x}^j}{\partial \dot{x}^k} = g_{ij}(\delta^i_k\dot{x}^j + \dot{x}^i\delta^j_k) \\
&= g_{kj}\dot{x}^j + g_{ik}\dot{x}^i = 2g_{kj}\dot{x}^j \\
\frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}^k}\right) &= 2\left(\frac{dg_{kj}}{dt}\dot{x}^j + g_{kj}\ddot{x}^j\right) = 2(g_{kj,l}\dot{x}^j\dot{x}^l + g_{kj}\ddot{x}^j)
\end{aligned}$$

and hence the Euler-Lagrange equation (6) becomes

$$g_{kj}\ddot{x}^j + g_{kj,l}\dot{x}^j\dot{x}^l - \frac{1}{2}g_{ij,k}\dot{x}^i\dot{x}^j = 0.$$

We rewrite the second term as follows:

$$\begin{aligned}
g_{kj,l}\dot{x}^j\dot{x}^l &= \frac{1}{2}(g_{kj,l}\dot{x}^j\dot{x}^l + g_{kj,l}\dot{x}^j\dot{x}^l) \\
&= \frac{1}{2}(g_{kj,l}\dot{x}^j\dot{x}^l + g_{kl,j}\dot{x}^l\dot{x}^j) && \text{(by relabeling the second term)} \\
&= \frac{1}{2}(g_{kj,l} + g_{kl,j})\dot{x}^l\dot{x}^j && \text{(since } \dot{x}^j\dot{x}^l = \dot{x}^l\dot{x}^j\text{)} \\
&= \frac{1}{2}(g_{kj,i} + g_{ki,j})\dot{x}^i\dot{x}^j && \text{(by relabeling } l \to i\text{)}
\end{aligned}$$

This leaves us with

$$g_{kj}\ddot{x}^j + \frac{1}{2}(g_{kj,i} + g_{ki,j} - g_{ij,k})\dot{x}^i\dot{x}^j = 0$$

and using equation (8) we have

$$\ddot{x}^l + \frac{1}{2}g^{lk}(g_{kj,i} + g_{ki,j} - g_{ij,k})\dot{x}^i\dot{x}^j = 0.$$

Let us define the **Christoffel symbols**

$$\Gamma^l_{ij} := \frac{1}{2}g^{lk}(g_{kj,i} + g_{ki,j} - g_{ij,k}) = \Gamma^l_{ji}.$$

Without the comma notation, this looks like

$$\Gamma^l_{ij} = \frac{1}{2}g^{lk}\left(\frac{\partial g_{kj}}{\partial x^i} + \frac{\partial g_{ki}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k}\right). \tag{9}$$

We will also find it useful to define the Christoffel symbol with all lowered indices:

$$\Gamma_{kij} := g_{kl}\Gamma^l_{ij} = \frac{1}{2}(g_{ki,j} + g_{kj,i} - g_{ij,k}) = \Gamma_{kji}. \tag{10}$$

and hence also

$$\Gamma^l_{ij} = g^{lk}\Gamma_{kij}. \tag{11}$$

In any case, the equation of a geodesic becomes

$$\ddot{x}^l + \Gamma^l_{ij}\dot{x}^i\dot{x}^j = 0$$

or

$$\frac{d^2x^l}{dt^2} + \Gamma^l_{ij}\frac{dx^i}{dt}\frac{dx^j}{dt} = 0. \tag{12}$$

It is left as a homework problem to look at equation (12) under a change of parameter. Solutions of equation (12) yield t as arclength or any other affinely related parameter λ (i.e., $\lambda = at + b$).

Example 4. The simplest nontrivial example is just plane polar coordinates $ds^2 = dr^2 + r^2 d\theta^2$. For clarity, we will sometimes write the indices $i = 1, 2$ as just r, θ. We will also use a hybrid notation that should be obvious from the context.

As we have seen above, the nonzero components of the metric are $g_{rr} = 1$ and $g_{\theta\theta} = r^2$ along with $g^{rr} = 1$ and $g^{\theta\theta} = 1/r^2$. The only nontrivial derivative of the metric components is

$$g_{\theta\theta,r} = \frac{\partial}{\partial r}g_{\theta\theta} = 2r$$

so the only non-vanishing Christoffel symbols are those with exactly two θ's and one r:

$$\Gamma_{r\theta\theta} = \frac{1}{2}(g_{r\theta,\theta} + g_{r\theta,\theta} - g_{\theta\theta,r}) = -r$$
$$\Gamma_{\theta r\theta} = \Gamma_{\theta\theta r} = \frac{1}{2}(g_{\theta\theta,r} + g_{\theta r,\theta} - g_{r\theta,\theta}) = r.$$

Since (g^{ij}) is diagonal, these yield

$$\Gamma^r_{\theta\theta} = g^{ri}\Gamma_{i\theta\theta} = g^{rr}\Gamma_{r\theta\theta} = -r$$

$$\Gamma^\theta_{r\theta} = \Gamma^\theta_{\theta r} = g^{\theta i}\Gamma_{ir\theta} = g^{\theta\theta}\Gamma_{\theta r\theta} = \frac{1}{r}$$

and the geodesic equations become

$$\ddot{r} - r\dot{\theta}^2 = 0 \qquad \text{and} \qquad \ddot{\theta} + \frac{2}{r}\dot{r}\dot{\theta} = 0.$$

This pair of coupled equations is not easy to solve. But by inspection we see that one class of solution is simply $\dot{\theta} = \ddot{r} = 0$. Then $\theta = \text{const}$ and r is of the form of a straight line $r = at + b$. In other words, as we already knew, the geodesics in the plane are just the straight lines. (Actually, in this particular class of solutions we have $\theta = \text{const}$, so these are only those lines that pass through the origin.)

4 Variational Notation and the Second Form of Euler's Equation

So far we have been fairly careful with our notation in explaining just what it means to vary a path. Now we are going to write our equations in another form that is commonly used, especially by physicists. I stress that there is nothing new here, it is only a change of notation that is justified by what we have done up to this point.

Recall that we had the varied curve $\widetilde{\gamma} = \gamma + h$ or, equivalently, $\widetilde{x}(t) = x(t) + h(t)$. Define the variation of x by

$$\delta x(t) := \widetilde{x}(t) - x(t) = h(t).$$

Taking the derivative of this yields

$$\frac{d}{dt}\delta x(t) = \dot{\widetilde{x}}(t) - \dot{x}(t) = \dot{h}(t).$$

But $\delta\dot{x}(t) := \dot{\widetilde{x}}(t) - \dot{x}(t)$ and hence we see that

$$\delta\dot{x}(t) = \frac{d}{dt}\delta x(t) = \dot{h}(t).$$

We also write

$$\delta f := f(x + h, \dot{x} + \dot{h}, t) - f(x, \dot{x}, t)$$

so expanding this to first order (i.e., ignore $O(h^2)$ terms) we have

$$\delta f = \frac{\partial f}{\partial x}h + \frac{\partial f}{\partial \dot{x}}\dot{h} = \frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial \dot{x}}\delta\dot{x}.$$

As we did in Theorem 1, let us consider the functional

$$F[\gamma] = \int_{t_0}^{t_1} f(x(t), \dot{x}(t), t)\, dt.$$

Then

$$\begin{aligned}
\delta F := F[\gamma + h] - F[\gamma] &= \int \delta f \\
&= \int_{t_0}^{t_1} \left(\frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial \dot{x}}\delta \dot{x} \right) dt = \int_{t_0}^{t_1} \left(\frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial \dot{x}}\frac{d}{dt}\delta x \right) dt \\
&= \int_{t_0}^{t_1} \left[\frac{\partial f}{\partial x} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}} \right) \right] \delta x\, dt
\end{aligned}$$

where to get the last line we integrated by parts and used $\delta x(t_0) = \delta x(t_1) = 0$.

Again, at an extremum we have $\delta F = 0$, and since this holds for arbitrary variations $\delta x(t)$ we must have the same result as before (Theorem 2):

$$\frac{\partial f}{\partial x} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}} \right) = 0. \tag{13}$$

As I have said, there is nothing new here, only the symbolic manipulation

$$\delta f = \frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial \dot{x}}\delta \dot{x}.$$

Let's take a closer look at equation (13). We will show that this is actually a second-order differential equation for $x(t)$. To see this, first note that for any function $g(x(t), \dot{x}(t), t)$ we have the total derivative

$$\frac{dg}{dt} = \frac{\partial g}{\partial t} + \frac{\partial g}{\partial x}\frac{dx}{dt} + \frac{\partial g}{\partial \dot{x}}\frac{d\dot{x}}{dt} = \frac{\partial g}{\partial t} + \frac{\partial g}{\partial x}\dot{x} + \frac{\partial g}{\partial \dot{x}}\ddot{x}.$$

Since equation (13) contains the total derivative of $\partial f/\partial \dot{x}$, we can apply the above equation with $g = \partial f/\partial \dot{x}$ to write (13) as

$$\frac{\partial f}{\partial x} - \frac{\partial^2 f}{\partial t \partial \dot{x}} - \frac{\partial^2 f}{\partial x \partial \dot{x}}\dot{x} - \frac{\partial^2 f}{\partial \dot{x}^2}\ddot{x} = 0.$$

This clearly shows that equation (13) is in fact a second-order differential equation for $x(t)$. Fortunately, in many cases of physical interest the function f will have no *explicit* dependence on the independent variable t, and we can then derive a simpler first-order equation. This is a simple consequence of the Euler-Lagrange equation. We have

$$\begin{aligned}
\frac{df}{dt} &= \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x}\dot{x} + \frac{\partial f}{\partial \dot{x}}\frac{d\dot{x}}{dt} \\
&= \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x}\dot{x} + \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}}\dot{x} \right) - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}} \right)\dot{x}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x}\dot{x} + \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}}\dot{x}\right) - \frac{\partial f}{\partial x}\dot{x} \qquad \text{(by equation (13))} \\
&= \frac{\partial f}{\partial t} + \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}}\dot{x}\right)
\end{aligned}
$$

or

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

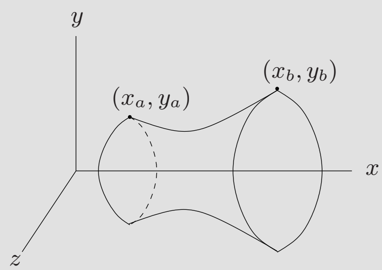
But then if $\partial f/\partial t = 0$ we are left with$(d/dt)[(\partial f/\partial \dot{x})\dot{x} - f] = 0$ or

$$
\frac{\partial f}{\partial \dot{x}}\dot{x} - f = \text{const} \tag{14}
$$

which is a first-order equation, called the **second form** of Euler's equation.

(Some readers might recognize that if f is the Lagrangian L, then this is just the statement that the Hamiltonian $H = \sum p\dot{q} - L$ is conserved. Note however, that this is not the statement that H is the total energy of the system; that fact depends on whether or not the kinetic energy is a homogeneous quadratic function of the $\dot{q}$'s.)

Example 5. Consider the surface of revolution problem. Here we have a curve $y(x)$ with endpoints (x_a, y_a) and (x_b, y_b) that is rotated about the x-axis. We want the curve that will minimize the area.



We assume that $y_a, y_b > 0$ and $y(x) \geq 0$ for $x_a \leq x \leq x_b$, and try to minimize the surface area given by

$$
2\pi \int y\, ds = 2\pi \int_{x_a}^{x_b} y\sqrt{1 + y'^2}\, dx
$$

where $y' = dy/dx$. Here the independent variable is x, and our function is $f = f(y, y', x) = y\sqrt{1 + y'^2}$ (the factor of 2π cancels out of equation (14)).

Since $\partial f/\partial x = 0$, we apply the second form of Euler's equation and write

$$
\frac{\partial f}{\partial y'}y' - f = -\alpha
$$

where we define the constant as $-\alpha$ for convenience (and hindsight). This gives

$$
\frac{yy'^2}{\sqrt{1 + y'^2}} - y\sqrt{1 + y'^2} = \frac{yy'^2 - y(1 + y'^2)}{\sqrt{1 + y'^2}} = \frac{-y}{\sqrt{1 + y'^2}} = -\alpha
$$

so that $(y/\alpha)^2 = 1 + y'^2$ or $y' = (y^2/\alpha^2 - 1)^{1/2}$ which has the solution

$$x - x_0 = \int \frac{dy}{(y^2/\alpha^2 - 1)^{1/2}} = \alpha \cosh^{-1} \frac{y}{\alpha}.$$

Sketch of how to do the integral: If $w = \cosh^{-1} u$ then $u = \cosh w = (e^w + e^{-w})/2$ so that $e^w - 2u + e^{-w} = 0$ or $e^{2w} - 2ue^w + 1 = 0$. Solving for e^w with the quadratic formula yields

$$e^w = \frac{2u \pm \sqrt{4u^2 - 4}}{2} = u + \sqrt{u^2 - 1}$$

where only the positive root is taken because $e^w > 0$ and $u = \cosh w \geq 1$.

Taking the logarithm of both sides we have $w = \ln(u + \sqrt{u^2 - 1}) = \cosh^{-1} u$ so that

$$\frac{d \cosh^{-1} u}{dx} = \left(\frac{1}{u + \sqrt{u^2 - 1}} \right) \left(u' + \frac{uu'}{\sqrt{u^2 - 1}} \right)$$

or

$$\frac{d \cosh^{-1} u}{dx} = \frac{u'}{\sqrt{u^2 - 1}}$$

and this implies that

$$d \cosh^{-1} u = \frac{du}{\sqrt{u^2 - 1}}.$$

In any case, the solution to our problem is

$$y = \alpha \cosh \left(\frac{x - x_0}{\alpha} \right)$$

which is called a **catenary**. Note that $y(x) > 0$ implies $\alpha > 0$. We want to choose the constants α and x_0 such that the curve passes through the points (x_a, y_a) and (x_b, y_b), but this can't always be done. Without going into the details, this is because our theory only determines twice-differentiable solutions. The general solution to this problem is called the **Goldschmidt discontinuous solution**.

5 An Alternative Approach to the Functional Derivative

While we have defined the functional derivative by analogy with the ordinary derivative, there are other definitions that are also frequently used. For example, in the path integral approach to quantum field theory the following approach is quite useful.

Let F be a function of N variables $y_0, y_1, \ldots, y_{N-1}$. If we start at a point

$y^0 = (y_0^0, \dots, y_{N-1}^0)$ and move to another nearby point with coordinates $y_n = y_n^0 + dy_n$, then from calculus we know that

$$dF = \sum_{n=0}^{N-1} \left. \frac{\partial F}{\partial y_n} \right|_{y^0} dy_n.$$

Suppose we have a function $y(x)$ with x defined on some interval $[a, b]$. We break this into $N - 1$ subintervals with a spacing ε between points. Then $(N-1)\varepsilon = b - a$ and the nth point is at $x = x_n = a + n\varepsilon$. The function y takes the values $y_n = y(x_n) = y(a+n\varepsilon)$ which clearly approaches $y(x)$ as $N \to \infty$ and $\varepsilon \to 0$. In this limit, our function $F[\{y_n\}]$ becomes a function of the function $y(x)$, and we write this *functional* as $F[y]$.

Now, what does the above expression for dF look like as $N \to \infty$? Recall the definition of integral:

$$\int_a^b dx\, f(x) = \lim_{\varepsilon \to 0} \sum_{n=0}^{N-1} \varepsilon f(x_n).$$

Write dF as

$$dF = \sum_{n=0}^{N-1} \varepsilon \left(\frac{1}{\varepsilon} \left. \frac{\partial F}{\partial y_n} \right|_{y^0} \right) dy_n.$$

Now take the limit $\varepsilon \to 0$ with $x = a + n\varepsilon$. Introducing the notation $dy_n = \delta y(x)$ we *define* the functional derivative $\delta F/\delta y(x)$ by

$$\delta F[y] = \int_a^b dx \left. \frac{\delta F[y]}{\delta y(x)} \right|_{y^0(x)} \delta y(x) \tag{15}$$

where $y^0(x)$ is the particular function $y(x)$ that is the starting point for the arbitrary infinitesimal variation $\delta y(x)$, and the factor $1/\varepsilon$ was absorbed into the term $\delta F/\delta y(x)$. Since now F is a functional and not a function, we also write δF instead of dF to denote the infinitesimal change in F.

Example 6. Let

$$F[y] = \int_0^1 [y(x)]^2\, dx.$$

Then to first order in δy we have

$$F[y + \delta y] = \int_0^1 [y(x) + \delta y(x)]^2\, dx = \int_0^1 [y(x)^2 + 2y(x)\delta y(x)]\, dx$$

so that

$$\delta F[y] = F[y + \delta y] - F[y] = \int_0^1 2y(x)\delta y(x)\, dx.$$

Comparison with (15) shows that

$$\frac{\delta F[y]}{\delta y(x)} = 2y(x).$$

Generalizing this result to $F[y] = \int_0^1 [y(x)]^n \, dx$ is easy using the binomial theorem

$$(a+b)^n = \sum_{k=1}^{n} \binom{n}{k} a^{n-k}\, b^k$$

where the binomial coefficient is defined by

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Then expanding to first order we have

$$F[y+\delta y] = \int_0^1 (y+\delta y)^n \, dx = \int_0^1 \{y^n + ny^{n-1}\delta y\}\, dx$$

so that

$$F[y+\delta y] - F[y] = \int_0^1 ny^{n-1}\delta y \, dx$$

and hence, as we would expect, comparison with equation (15) shows that

$$\frac{\delta F}{\delta y(x)} = n[y(x)]^{n-1}.$$

In fact, we can now use this result to find the derivative of more general functionals. For example, using the Taylor expansions

$$\sin y = y - \frac{y^3}{3!} + \frac{y^5}{5!} - \cdots \qquad \text{and} \qquad \cos y = 1 - \frac{y^2}{2!} + \frac{y^4}{4!} + \cdots$$

it is easy to see that letting

$$F[y] = \int \sin y(x)\, dx$$

we have

$$\frac{\delta F[y]}{\delta y(x)} = \cos y(x).$$

Example 7. Consider the functional $F[g] = \int f(x)g(x)\, dx$ where $f(x)$ is some fixed function. Then

$$\delta F[g] = F[g + \delta g] - F[g] = \int f(x)[g(x) + \delta g(x)]\, dx - \int f(x)g(x)\, dx$$
$$= \int f(x)\delta g(x)\, dx$$

so that comparison with equation (15) shows that

$$\frac{\delta F[g]}{\delta g(x)} = f(x)\,.$$

We can also use this approach to derive the Euler-Lagrange equation. Let us first consider a simpler problem. Start from

$$F[y] = \int L(y(x), x)\, dx$$

where L is just some function of $y(x)$ and x. Varying $y(x)$ we have (to first order)

$$F[y + \delta y] = \int L(y + \delta y, x)\, dx = \int \left[L(y(x), x) + \frac{\partial L(x, y)}{\partial y}\delta y \right] dx.$$

Therefore

$$\delta F = F[y + \delta y] - F[y] = \int dx\, \frac{\partial L(x, y)}{\partial y}\delta y$$

and hence by (15) again we have

$$\frac{\delta F}{\delta y(x)} = \frac{\partial L(x, y)}{\partial y}. \tag{16}$$

As a simple example, suppose $F[y] = \int x^3 e^{-y(x)}\, dx$. Then from (16) we find $\delta F/\delta y = -x^3 e^{-y(x)}$.

Next, to derive the Euler-Lagrange equation we proceed exactly as we did earlier. Starting from

$$F[y] = \int_a^b L(y(x), y'(x), x)\, dx$$

we have (to first order as usual)

$$\begin{aligned} F[y + \delta y] &= \int_a^b L(y + \delta y, y' + \delta y', x)\, dx \\ &= \int_a^b \left[L(y, y', x) + \frac{\partial L}{\partial y}\delta y + \frac{\partial L}{\partial y'}\delta y' \right] dx \\ &= \int_a^b \left[L(y, y', x) + \frac{\partial L}{\partial y}\delta y - \frac{d}{dx}\left(\frac{\partial L}{\partial y'}\right)\delta y \right] dx + \frac{\partial L}{\partial y'}\delta y(x)\Big|_a^b . \end{aligned}$$

This gives us

$$\delta F = \int_a^b dx \left[\frac{\partial L}{\partial y} - \frac{d}{dx}\left(\frac{\partial L}{\partial y'}\right)\right]\delta y(x) + \frac{\partial L}{\partial y'}\delta y(x)\bigg|_a^b.$$

If we think of (15) as a sum over terms $[\delta L/\delta y(x)]\delta y(x)$ as x goes from a to b, then for $x \neq a, b$ we have the functional derivative as the coefficient of $\delta y(x)$:

$$\frac{\delta L}{\delta y(x)} = \frac{\partial L}{\partial y} - \frac{d}{dx}\left(\frac{\partial L}{\partial y'}\right) \tag{17}$$

whereas if $x = a$ or b, then we must include the boundary term

$$-\frac{\partial L}{\partial y'}\bigg|_a \qquad \text{or} \qquad \frac{\partial L}{\partial y'}\bigg|_b$$

because these are the coefficients of $\delta y(a)$ and $\delta y(b)$.

To again arrive at the Euler-Lagrange equation, we first define the functional (called the **action**)

$$S[x(t)] = \int_{t_0}^{t_1} L(x(t), \dot{x}(t), t)\, dt.$$

Hamilton's principle then states that the true trajectory of the particle is one which minimizes the action subject to the constraint that the endpoints are fixed (i.e., no boundary terms). This just means that

$$\frac{\delta S}{\delta x(t)} = 0$$

and therefore equation (17) yields the usual Euler-Lagrange equation

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

6 The Dirac Delta Function

Let us consider the simplest possible functional

$$F[y(x)] = y(x_0)$$

which just evaluates $y(x)$ at the specific point x_0, and ask about it's functional derivative. Well, since F depends on y at only the single point x_0, if we vary y at some point other than x_0 it will obviously have no effect on F, and therefore we must have

$$\frac{\delta F}{\delta y(x)} = 0 \qquad \text{for } x \neq x_0.$$

In order to deduce what the functional derivative of F is at x_0, let us *assume* there exists some function $\delta(x - x_0)$ such that

$$F[y(x)] = \int \delta(x - x_0) y(x)\, dx. \tag{18}$$

Since F is independent of $y(x)$ for $x \neq x_0$, we must have $\delta(x - x_0) = 0$ for $x \neq x_0$. And since the measure dx goes to zero, the value of $\int \delta(x - x_0)y(x)\,dx$ will be zero for any finite value of $\delta(x - x_0)$ at the single point x_0. Thus $\delta(x - x_0)$ must be infinite at $x = x_0$. Fortunately, the actual value of $\delta(0)$ is irrelevant for practical purposes, because all we really need to know is how to integrate with $\delta(x)$.

Using the definition $F[y(x)] = y(x_0)$, we *define* the "function" $\delta(x)$ by

$$y(x_0) = \int_a^b \delta(x - x_0)y(x)\,dx$$

where $x_0 \in [a, b]$. Note that choosing the particular case $y(x) = 1$, this equation shows that

$$\int_a^b \delta(x - x_0)\,dx = 1.$$

Thus, even though its width is zero, the area under the δ function is 1. Such "functions" can be realized as the limit of a sequence of proper functions such as the gaussians

$$\delta(x) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon\sqrt{\pi}} e^{-x^2/\varepsilon}$$

or as

$$\delta(x) = \lim_{\varepsilon \to 0} \frac{1}{\pi} \frac{\varepsilon}{x^2 + \varepsilon^2}.$$

In any case, from equation (18) we see that

$$\begin{aligned}
\delta F &= F[y + \delta y] - F[y] \\
&= \int \delta(x - x_0)(y(x) + \delta y(x))\,dx - \int \delta(x - x_0)y(x)\,dx \\
&= \int \delta(x - x_0)\delta y(x)\,dx
\end{aligned}$$

and comparing this with equation (15) we conclude that

$$\frac{\delta F}{\delta y(x)} = \delta(x - x_0)$$

or

$$\frac{\delta y(x_0)}{\delta y(x)} = \delta(x - x_0). \tag{19}$$

Given that we have the delta function by one approach or another, we can use it to define the functional derivative in a manner analogous to the ordinary derivative by

$$\frac{\delta F[y]}{\delta y(x_0)} := \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\big\{F[y(x) + \varepsilon\delta(x - x_0)] - F[y(x)]\big\}.$$

To see the equivalence of this definition with equation (15), consider a variation of the "independent variable" $y(x)$ that is localized at x_0 and has strength ε:

$$\delta y(x) = \varepsilon\delta(x - x_0)\,.$$

Using this in (15) we have

$$\delta F[y] = F[y + \varepsilon\delta(x - x_0)] - F[y] = \int dx \frac{\delta F[y]}{\delta y(x)} \varepsilon\delta(x - x_0) = \varepsilon \frac{\delta F[y]}{\delta y(x_0)}$$

so that dividing by ε we obtain

$$\lim_{\varepsilon\to 0} \frac{F[y + \varepsilon\delta(x - x_0)] - F[y]}{\varepsilon} = \frac{\delta F[y]}{\delta y(x_0)}\,.$$

For instance, using the same functional $F[y] = \int_0^1 y(x)^2\,dx$ as we had in Example 6, we again find that (by expanding to first order in ε and assuming that $x_0 \in [0, 1]$)

$$\begin{aligned}
\frac{\delta F[y]}{\delta y(x_0)} &= \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\left\{ \int_0^1 [y(x) + \varepsilon\delta(x - x_0)]^2\,dx - \int_0^1 y(x)^2\,dx \right\} \\
&= \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_0^1 2\varepsilon\delta(x - x_0)y(x)\,dx \\
&= 2y(x_0).
\end{aligned}$$

If $F[y] = \int y(x)^n\,dx$, then this approach along with the binomial theorem gives the result

$$\frac{\delta F[y]}{\delta y(x_0)} = n[y(x_0)]^{n-1}.$$

Note also that if we take $F[y] = y(x)$ then

$$\begin{aligned}
\frac{\delta y(x)}{\delta y(x_0)} &= \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\{y(x) + \varepsilon\delta(x - x_0) - y(x)\} \\
&= \delta(x - x_0)
\end{aligned}$$

in agreement with equation (19). As another way of seeing this, we can write $F[y]$ in the explicitly functional form

$$F[y] = \int y(z)\delta(z - x)\,dz = y(x)$$

and therefore

$$\begin{aligned}
\frac{\delta F[y]}{\delta y(x_0)} &= \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\left\{ \int [y(z) + \varepsilon\delta(z - x_0)]\delta(z - x)\,dz - \int y(z)\delta(z - x)\,dz \right\} \\
&= \delta(x - x_0)\,.
\end{aligned}$$

7 Constraints and Lagrange Multipliers

Here is the general idea behind Lagrange multipliers. Suppose we have a function $f(x, y)$ that we want to extremize subject to a constraint equation of the form $g(x, y) = 0$. Assume that this constrained (relative) extremum occurs at the point (x_0, y_0). Furthermore, let us assume that the constraint equation is parametrized as $x = x(s), y = y(s)$ where s is arc length and $s = 0$ at (x_0, y_0). Then $z = f(x(s), y(s))$ has an extremum at (x_0, y_0) so that $dz/ds = 0$ at that point. But this means that at (x_0, y_0) we have

$$0 = \frac{dz}{ds} = \frac{\partial f}{\partial x}\frac{dx}{ds} + \frac{\partial f}{\partial y}\frac{dy}{ds} = \nabla f \cdot \mathbf{T}$$

where $\mathbf{T} = (x'(s), y'(s))$ is the (unit) tangent to the curve $g(x(s), y(s)) = 0$, and therefore ∇f is orthogonal to $\mathbf{T}$ at (x_0, y_0) (assuming that $\nabla f \neq 0$).

On the other hand, we know that ∇g is orthogonal to surfaces of constant g. (This is just the statement that since $dg = \nabla g \cdot d\mathbf{r}$ in general, it follows that if $d\mathbf{r}$ lies in the direction of constant g, then $dg = 0$ so that ∇g must be orthogonal to $d\mathbf{r}$.) Since $g(x, y) = 0$ is a level curve for g, it must be that ∇g is orthogonal to $\mathbf{T}$ also. But then both ∇f and ∇g are orthogonal to $\mathbf{T}$, and all of them lie in the same plane so that we must have $\nabla f = \lambda \nabla g$ for some scalar λ. In other words, we have $\nabla(f - \lambda g) = 0$.

Now suppose we have a function $f(\mathbf{x})$ defined on $\mathbb{R}^3$, and in addition, suppose that we wish to extremize this function subject to two constraint equations $g(\mathbf{x}) = 0$ and $h(\mathbf{x}) = 0$. These constraint equations are surfaces in $\mathbb{R}^3$, and we also assume that they intersect along some curve with tangent $\mathbf{T}$. (If they don't, then it is impossible to satisfy the conditions of the problem.) Again, we evaluate f along this intersection curve and look for the point where $df = 0$. As before, this means we want the point where $\nabla f \cdot \mathbf{T} = 0$. Since $\mathbf{T}$ is also tangent to both constraint surfaces, it must be that $\nabla g \cdot \mathbf{T} = \nabla h \cdot \mathbf{T} = 0$, and therefore $\nabla f, \nabla g$ and ∇h all lie in the same plane orthogonal to $\mathbf{T}$. This means that two of them can be written as a linear combination of the others so that we can write $\nabla f = \lambda_1 \nabla g + \lambda_2 \nabla h$ or $\nabla(f - \lambda_1 g - \lambda_2 h) = 0$. Points that satisfy equations of this type are called **critical points**.

In the simpler case of a single constraint equation $g(\mathbf{x}) = 0$, we would have to evaluate f over the surface $g = 0$ and look for the point $\mathbf{x}_0$ where ∇f is orthogonal to the entire tangent plane (i.e., $df = 0$ no matter which direction you move). Since ∇g is also orthogonal to the tangent plane, it again follows that ∇f and ∇g must be proportional.

Before working an example, let me remark that the question of local versus absolute extremum depends on the notion of a continuous function on compact spaces. If the domain of f is compact, then f has an absolute max or min on that domain. In the case of $\mathbb{R}^n$, compactness is equivalent to being closed and bounded (this is the famous Heine-Borel theorem). We also ignore the question of how to prove that a given critical point is a max, min or saddle point. Fortunately, this is usually clear in most physical problems.

Example 8. Let $f(x, y, z) = xyz$ and $g(x, y, z) = x^2 + y^2 + z^2 - 1 = 0$. Then $\nabla(f + \lambda g) = 0$ implies

$$\begin{aligned} yz + 2\lambda x &= 0 \\ xz + 2\lambda y &= 0 \\ xy + 2\lambda z &= 0 \end{aligned}$$

together with $x^2 + y^2 + z^2 = 1$. Multiply the first three equations by x, y, z respectively and add them together using the last (constraint) equation to obtain $3xyz + 2\lambda = 0$ or $\lambda = -(3/2)xyz$.

The first equation now yields $yz - 3x^2yz = 0$. If $y, z \neq 0$ then $x = \pm 1/\sqrt{3}$. The second equation is $xz - 3xy^2z = 0$ which implies $y = \pm 1/\sqrt{3}$ (since x and y are nonzero), and the third equation is $xy - 3xyz^2 = 0$ so $z = \pm 1/\sqrt{3}$. Therefore the max and min are the eight points $(\pm 1/\sqrt{3}, \pm 1/\sqrt{3}, \pm 1/\sqrt{3})$.

What if $y = 0$? The second equation then says that $xz = 0$ which implies either $x = 0$ or $z = 0$ also. But if $x = 0$ then the constraint equation says that $z = \pm 1$. If $y = 0$ and $z = 0$ then we similarly have $x = \pm 1$. If we had started with the assumption that $z = 0$ we would have $xy = 0$, so following the same argument we conclude that we also have the additional solutions $(0, 0, \pm 1), (0, \pm 1, 0)$ and $(\pm 1, 0, 0)$.

Do these additional solutions also correspond to either a max or min? No, they don't. To see this, look at the function $f = xyz$ in, for example, a neighborhood of $(0, 0, 1)$. As you move away from this point in the x and y directions, the function $f(x, y, 1) = xy$ will be > 0 if x, y have the same sign, and it will be < 0 if they have opposite signs. Thus these additional points are inflection (or saddle) points.

Now that we have the basic idea behind Lagrange multipliers, we wish to apply it to minimizing functionals (as opposed to simply functions) subject to constraints. To keep the notation as simple as possible, and also to agree with the way many texts treat the subject, we begin by turning to the third formulation of the calculus of variations. (The first two formulations led to equations (5) and (13).)

Recall that we had $x(t) \to \widetilde{x}(t) = x(t) + \eta(t)$ and we varied the curve $\eta(t)$. (We labeled the curve $\eta(t)$ by $h(t)$ before, but I don't want to confuse this with the function h defined below.) This led to the *functional*

$$F[\widetilde{\gamma}] = \int_{t_0}^{t_1} f(\widetilde{x}, \dot{\widetilde{x}}, t)\, dt.$$

Now, instead of varying η, let us choose an *arbitrary* but *fixed* η, and introduce a variation parameter ε so that

$$\widetilde{x}(t, \varepsilon) = x(t) + \varepsilon \eta(t).$$

(Let me remark that the specific form $\widetilde{x} = x + \varepsilon\eta$ is never needed in what follows. All that we really need is a one-parameter family of test functions $\widetilde{x}(t, \varepsilon)$. See the discussion below.) Then the functional becomes an ordinary integral *function* of ε which we denote by $I(\varepsilon)$:

$$I(\varepsilon) = \int_{t_0}^{t_1} f(\widetilde{x}, \dot{\widetilde{x}}, t)\, dt.$$

Be sure to realize just what the difference is between the integrands in these two equations, and why the first is a functional while the second is an ordinary function.

In the present case, $\varepsilon = 0$ corresponds to the desired extremizing function $x(t)$. Recall also that in deriving the Euler-Lagrange equation, our curves $\eta(t)$ all obeyed the boundary conditions $\eta(t_0) = \eta(t_1) = 0$, i.e., all of the varied curves pass through the *fixed* endpoints $x_0 = x(t_0)$ and $x_1 = x(t_1)$. In other words, we have

$$\widetilde{x}(t_0, \varepsilon) = x_0 \quad \text{and} \quad \widetilde{x}(t_1, \varepsilon) = x_1 \quad \text{for all } \varepsilon$$

and $\widetilde{x}(t, 0) = x(t)$ is the desired extremizing function. We shall also assume that $\widetilde{x}(t, \varepsilon)$ is C^2, i.e., that it has continuous first and second derivatives so that

$$\frac{\partial^2}{\partial t\, \partial\varepsilon} = \frac{\partial^2}{\partial\varepsilon\, \partial t}.$$

Since $\widetilde{x}(t, 0) = x(t)$ we see that $I(0) = \int f(x, \dot{x}, t)\, dt$ is the desired extremized functional, and hence we want

$$\left.\frac{dI(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0} = 0.$$

Note that by using the parameter ε we can now take the ordinary derivative of $I(\varepsilon)$, whereas before we were really taking the functional derivative of $F[\gamma]$. Again, be aware that $dI/d\varepsilon = 0$ only gives an extremum, and not necessarily a minimum. Making the distinction between a max and min is tricky, and in our physical problems it will usually be obvious that we have indeed found a minimum.

Before addressing the problem of constraints, let us show how to derive the Euler-Lagrange equation in this new formalism. We have

$$\begin{aligned}
\frac{dI}{d\varepsilon} &= \frac{d}{d\varepsilon} \int_{t_0}^{t_1} f(\widetilde{x}(t, \varepsilon), \dot{\widetilde{x}}(t, \varepsilon), t)\, dt \\
&= \int_{t_0}^{t_1} \left(\frac{\partial f}{\partial \widetilde{x}} \frac{d\widetilde{x}}{d\varepsilon} + \frac{\partial f}{\partial \dot{\widetilde{x}}} \frac{d\dot{\widetilde{x}}}{d\varepsilon} \right) dt \\
&= \int_{t_0}^{t_1} \left[\frac{\partial f}{\partial \widetilde{x}} \frac{d\widetilde{x}}{d\varepsilon} + \frac{\partial f}{\partial \dot{\widetilde{x}}} \frac{d}{dt}\left(\frac{d\widetilde{x}}{d\varepsilon}\right) \right] dt \qquad \text{since} \qquad \frac{d\dot{\widetilde{x}}}{d\varepsilon} = \frac{d}{d\varepsilon}\frac{d\widetilde{x}}{dt} = \frac{d}{dt}\frac{d\widetilde{x}}{d\varepsilon} \\
&= \int_{t_0}^{t_1} \left[\frac{\partial f}{\partial \widetilde{x}} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{\widetilde{x}}}\right) \right] \frac{d\widetilde{x}}{d\varepsilon}\, dt + \left.\frac{\partial f}{\partial \dot{\widetilde{x}}} \frac{d\widetilde{x}}{d\varepsilon}\right|_{t_0}^{t_1}.
\end{aligned}$$

But $\widetilde{x}(t_0, \varepsilon) = x_0$ for *all* ε, so by definition of the derivative we have

$$\left.\frac{d\widetilde{x}}{d\varepsilon}\right|_{t_0} = \lim_{\delta \to 0} \frac{\widetilde{x}(t_0, \varepsilon + \delta) - \widetilde{x}(t_0, \varepsilon)}{\delta} = 0$$

and similarly for $\widetilde{x}(t_1, \varepsilon)$. Therefore the boundary terms vanish and we are left with

$$\frac{dI}{d\varepsilon} = \int_{t_0}^{t_1} \left[\frac{\partial f}{\partial \widetilde{x}} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{\widetilde{x}}}\right)\right] \frac{d\widetilde{x}}{d\varepsilon}\, dt.$$

This then gives us

$$\begin{aligned} 0 = \left.\frac{dI}{d\varepsilon}\right|_{\varepsilon=0} &= \int_{t_0}^{t_1} \left[\frac{\partial f}{\partial \widetilde{x}} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{\widetilde{x}}}\right)\right]_{\varepsilon=0} \left.\frac{d\widetilde{x}}{d\varepsilon}\right|_{\varepsilon=0} dt \\ &= \int_{t_0}^{t_1} \left[\frac{\partial f}{\partial x} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}}\right)\right] \eta\, dt. \end{aligned}$$

Observe that now this last equation is in terms of x and not $\widetilde{x}$. Also, this result doesn't depend on the specific form $\widetilde{x} = x + \varepsilon\eta$. We could just as easily *define* $\eta(t) = d\widetilde{x}/d\varepsilon|_{\varepsilon=0}$ for any one-parameter family of test functions $\widetilde{x}(t, \varepsilon)$ and it would still be an arbitrary function.

Since $\eta(t)$ is arbitrary, our lemma tells us that

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

as before. Note also that taking the limit $\varepsilon \to 0$ and letting $\widetilde{x} \to x$ and $\dot{\widetilde{x}} \to \dot{x}$ depended on $\widetilde{x}$ being C^2. (Recall we showed that the Euler-Lagrange equation is second order.)

Now that we have this last formalism available, let us return to the problem of constraints. We first treat the problem of integral constraints. Thus we want to minimize the functional $I = \int f(x, \dot{x}, t)\, dt$ subject to the constraint $J := \int g(x, \dot{x}, t)\, dt = \text{const}$. If we were to proceed as above and let $x \to \widetilde{x}(t, \varepsilon)$, then the equation $J(\varepsilon) = \text{const}$ would force a specific value for ε and there would be no way to vary $I(\varepsilon)$.

To get around this problem we introduce a *two*-parameter family of functions $\widetilde{x}(t, \varepsilon_1, \varepsilon_2)$ with the following properties:

(1) $\widetilde{x}(t_0, \varepsilon_1, \varepsilon_2) = x_0$ and $\widetilde{x}(t_1, \varepsilon_1, \varepsilon_2) = x_1$ for all $\varepsilon_1, \varepsilon_2$ (i.e., fixed endpoints);

(2) $\widetilde{x}(t, 0, 0) = x(t)$ is the desired extremizing function;

(3) $\widetilde{x}(t, \varepsilon_1, \varepsilon_2)$ has continuous derivatives through second order in all variables.

Define

$$I(\varepsilon_1, \varepsilon_2) = \int_{t_0}^{t_1} f(\widetilde{x}(t, \varepsilon_1, \varepsilon_2), \dot{\widetilde{x}}(t, \varepsilon_1, \varepsilon_2), t)\, dt$$

and

$$J(\varepsilon_1, \varepsilon_2) = \int_{t_0}^{t_1} g(\widetilde{x}(t, \varepsilon_1, \varepsilon_2), \dot{\widetilde{x}}(t, \varepsilon_1, \varepsilon_2), t)\, dt = \text{const}.$$

We want to extremize the function $I(\varepsilon_1, \varepsilon_2)$ subject to the constraint equation $J(\varepsilon_1, \varepsilon_2) = \text{const}$ (or $J(\varepsilon_1, \varepsilon_2) - \text{const} = 0$). This is just the original Lagrange multiplier problem, so we form the quantity

$$K(\varepsilon_1, \varepsilon_2) = I(\varepsilon_1, \varepsilon_2) + \lambda J(\varepsilon_1, \varepsilon_2). \tag{20}$$

The requirement $\nabla(f + \lambda g) = 0$ that we had in our general discussion now becomes

$$\left.\frac{\partial K}{\partial \varepsilon_1}\right|_{\varepsilon_1=\varepsilon_2=0} = \left.\frac{\partial K}{\partial \varepsilon_2}\right|_{\varepsilon_1=\varepsilon_2=0} = 0$$

where

$$K(\varepsilon_1, \varepsilon_2) = \int (f + \lambda g)\, dt := \int h(\widetilde{x}, \dot{\widetilde{x}}, t)\, dt$$

and the gradient is with respect to the variables ε_1 and ε_2.

This is now just the Euler-Lagrange problem that we already solved, so we have for $i = 1, 2$

$$\begin{aligned}
\frac{\partial K}{\partial \varepsilon_i} &= \int_{t_0}^{t_1} \left(\frac{\partial h}{\partial \widetilde{x}} \frac{\partial \widetilde{x}}{\partial \varepsilon_i} + \frac{\partial h}{\partial \dot{\widetilde{x}}} \frac{\partial \dot{\widetilde{x}}}{\partial \varepsilon_i} \right) dt \\
&= \int_{t_0}^{t_1} \left[\frac{\partial h}{\partial \widetilde{x}} - \frac{d}{dt}\left(\frac{\partial h}{\partial \dot{\widetilde{x}}} \right) \right] \frac{\partial \widetilde{x}}{\partial \varepsilon_i}\, dt
\end{aligned}$$

where we used

$$\frac{\partial \dot{\widetilde{x}}}{\partial \varepsilon_i} = \frac{\partial}{\partial \varepsilon_i} \frac{d\widetilde{x}}{dt} = \frac{d}{dt} \frac{\partial \widetilde{x}}{\partial \varepsilon_i}$$

and where, as usual, property (1) above allows us to drop the boundary terms. Writing

$$\eta_i(t) := \left.\frac{\partial \widetilde{x}}{\partial \varepsilon_i}\right|_{\varepsilon_1=\varepsilon_2=0}$$

we have

$$\left.\frac{\partial K}{\partial \varepsilon_i}\right|_{\varepsilon_1=\varepsilon_2=0} = \int_{t_0}^{t_1} \left[\frac{\partial h}{\partial x} - \frac{d}{dt}\left(\frac{\partial h}{\partial \dot{x}} \right) \right] \eta_i(t)\, dt = 0 \qquad i = 1, 2.$$

Since $\eta_i(t)$ is an arbitrary function, by our lemma we again arrive at

$$\frac{\partial h}{\partial x} - \frac{d}{dt}\left(\frac{\partial h}{\partial \dot{x}} \right) = 0 \tag{21}$$

where

$$h(x, \dot{x}, t) = f(x, \dot{x}, t) + \lambda g(x, \dot{x}, t). \tag{22}$$

Equation (21) looks exactly like our previous version of the Euler-Lagrange equation except that now $h = f + \lambda g$ replaces f.

Note that since the Euler-Lagrange equation is second order (as we saw earlier) there are two constants of integration in its solution. These, together with the Lagrange multiplier λ, are just enough to ensure that $x(t)$ passes through the endpoints $x_0 = x(t_0)$ and $x_1 = x(t_1)$, and that J has its correct value.

Generalizing this result is easy. If we have m integral constraint equations

$$J_j = \int g_j(x, \dot{x}, t)\, dt = \text{const} \qquad j = 1, \ldots, m$$

we introduce $m+1$ parameters $\varepsilon_k, k = 1, \ldots, m+1$ and consider the functions $\widetilde{x}(t, \varepsilon_1, \ldots, \varepsilon_{m+1}) := \widetilde{x}(t, \varepsilon)$ where $\widetilde{x}(t, 0) = x(t)$ is the extremizing function (through fixed endpoints). With $I(\varepsilon) = \int f(\widetilde{x}, \dot{\widetilde{x}}, t)\, dt$ we would then form the function

$$h = f + \sum_{j=1}^{m} \lambda_j g_j \qquad \lambda_j = \text{const}$$

so that

$$K(\varepsilon) = \int h(\widetilde{x}, \dot{\widetilde{x}}, t)\, dt = \int f(\widetilde{x}, \dot{\widetilde{x}}, t)\, dt + \sum_{j=1}^{m} \lambda_j \int g_j(\widetilde{x}, \dot{\widetilde{x}}, t)\, dt$$

and again require that

$$\left.\frac{\partial K}{\partial \varepsilon_k}\right|_{\varepsilon=0} = 0 \qquad \text{for all } k = 1, \ldots, m+1.$$

In addition, we could have n dependent variables $x_i(t)$ so that $I = \int f(x_i, \dot{x}_i, t)\, dt$ (where f is a function of $x_1(t), \ldots, x_n(t)$) and $J_j = \int g_j(x_i, \dot{x}_i, t)\, dt$ for $j = 1, \ldots, m$. Now we have

$$\begin{aligned} I(\varepsilon) &= \int f(\widetilde{x}_i, \dot{\widetilde{x}}_i, t)\, dt \qquad \text{where } \widetilde{x}_i = \widetilde{x}_i(t, \varepsilon_1, \ldots, \varepsilon_{m+1}) \\ J_j(\varepsilon) &= \int g_j(\widetilde{x}_i, \dot{\widetilde{x}}_i, t)\, dt \\ K(\varepsilon) &= \int h(\widetilde{x}_i, \dot{\widetilde{x}}_i, t)\, dt \end{aligned}$$

with

$$h(\widetilde{x}_i, \dot{\widetilde{x}}_i, t) = f(\widetilde{x}_i, \dot{\widetilde{x}}_i, t) + \sum_{j=1}^{m} \lambda_j g_j(\widetilde{x}_i, \dot{\widetilde{x}}_i, t).$$

Then

$$\begin{aligned} \frac{\partial K}{\partial \varepsilon_k} &= \int \sum_{i=1}^{n} \left(\frac{\partial h}{\partial \widetilde{x}_i} \frac{\partial \widetilde{x}_i}{\partial \varepsilon_k} + \frac{\partial h}{\partial \dot{\widetilde{x}}_i} \frac{\partial \dot{\widetilde{x}}_i}{\partial \varepsilon_k} \right) dt \qquad k = 1, \ldots, m+1 \\ &= \int \sum_{i=1}^{n} \left[\frac{\partial h}{\partial \widetilde{x}_i} - \frac{d}{dt}\left(\frac{\partial h}{\partial \dot{\widetilde{x}}_i} \right) \right] \frac{\partial \widetilde{x}_i}{\partial \varepsilon_k}\, dt \qquad \text{(since all boundary terms vanish)} \end{aligned}$$

so that

$$\left.\frac{\partial K}{\partial \varepsilon_k}\right|_{\varepsilon=0} = \int \sum_{i=1}^{n} \left[\frac{\partial h}{\partial x_i} - \frac{d}{dt}\left(\frac{\partial h}{\partial \dot{x}_i} \right) \right] \eta_{ik}(t)\, dt = 0 \qquad k = 1, \dots, m+1$$

where

$$\eta_{ik}(t) := \left.\frac{\partial \widetilde{x}_i(t, \varepsilon)}{\partial \varepsilon_k}\right|_{\varepsilon=0} \qquad \text{for } i = 1, \dots, n \text{ and } k = 1, \dots, m+1.$$

Finally, since each $\eta_{ik}(t)$ is arbitrary we have

$$\frac{\partial h}{\partial x_i} - \frac{d}{dt}\left(\frac{\partial h}{\partial \dot{x}_i} \right) = 0 \qquad \text{for each } i = 1, \dots, n \tag{23}$$

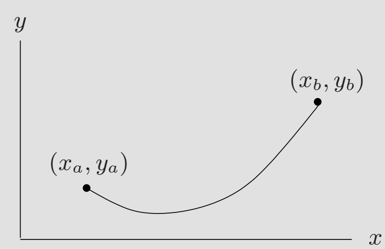
where

$$h(x_i, \dot{x}_i, t) = f(x_i, \dot{x}_i, t) + \sum_{j=1}^{m} \lambda_j g_j(x_i, \dot{x}_i, t). \tag{24}$$

(Equation (23) follows by noting that since each η_{ik} is arbitrary, for any fixed k we can let $\eta_{ik} = 0$ for all i except one, say $i = l$. Then $\partial h/\partial x_l - (d/dt)(\partial h/\partial \dot{x}_l) = 0$.)

These equations can be solved because we have n variables x_i plus m of the λ_j's as unknowns, but we also have n Euler-Lagrange equations plus m constraint equations.

Example 9. Let us solve the problem of a freely hanging perfectly flexible rope with fixed endpoints. This is referred to as an **isoperimetric** problem.



Our approach will be to minimize the potential energy of the rope.

Let ρ be the mass per unit length. The the potential energy relative to the x-axis is

$$\rho g \int y\, ds = \rho g \int_{x_a}^{x_b} y\sqrt{1 + y'^2}\, dx.$$

The constraint is that the length of the rope is fixed:

$$L = \int ds = \int_{x_a}^{x_b} \sqrt{1 + y'^2}\, dx.$$

We form the function

$$
\begin{aligned}
h &= \rho g y\sqrt{1+y'^2} + \lambda\sqrt{1+y'^2} \\
&= \rho g y\sqrt{1+y'^2} - \rho g y_0\sqrt{1+y'^2} \\
&= \rho g\sqrt{1+y'^2}(y-y_0)
\end{aligned}
$$

where we have defined the constant y_0 by $\lambda = -\rho g y_0$ and now y_0 is to be determined.

Change variables from y to $z = y - y_0$. Then $z' = y'$ and we have $h = \rho g z\sqrt{1+z'^2}$. But now this is exactly the same form of function as we had in solving the surface of revolution problem. And as Feynman said, "The same equations have the same solution." So the solution to this problem is the catenary $z = \alpha\cosh[(x-x_0)/\alpha]$ or

$$
y = y_0 + \alpha\cosh\left(\frac{x-x_0}{\alpha}\right)
$$

where x_0, y_0 and α are fixed by the requirements that the ends are at (x_a, y_a) and (x_b, y_b) and that the length of the rope is fixed.

Example 10. Another famous solution solves the **Dido problem**: Find the curve with fixed length that encloses the maximum area. (Actually, this is only one version of the problem.) Here we want the functions $x(t), y(t)$ such that the curve $(x(t), y(t))$ encloses the maximum area. We will use the following formula for the enclosed area:

$$
A = \frac{1}{2}\int_{t_a}^{t_b} (x\dot{y} - \dot{x}y)\, dt.
$$

This formula is a consequence of Green's Theorem, which you should have had in calculus:

$$
\oint_C f(x,y)\, dx + g(x,y)\, dy = \int_R \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}\right) dA.
$$

If we let $f = 0$ and $g = x$ in this formula, then $\oint x\, dy = \int dA = A$. And letting $f = -y$ and $g = 0$ yields $\oint -y\, dx = \int dA = A$. Therefore

$$
A = \frac{1}{2}\int x\, dy - y\, dx = \frac{1}{2}\int (x\dot{y} - \dot{x}y)\, dt.
$$

The length of the curve is given by

$$
L = \int ds = \int_{t_a}^{t_b} (\dot{x}^2 + \dot{y}^2)^{1/2}\, dt
$$

and hence our function h is

$$h(x, y, \dot{x}, \dot{y}, t) = \frac{1}{2}(x\dot{y} - \dot{x}y) + \lambda(\dot{x}^2 + \dot{y}^2)^{1/2}.$$

Note that we have one independent variable, two dependent variables, one constraint equation and hence one Lagrange multiplier. From equation (23) we obtain the equations

$$0 = \frac{\partial h}{\partial x} - \frac{d}{dt}\frac{\partial h}{\partial \dot{x}} = \frac{1}{2}\dot{y} - \frac{d}{dt}\left[-\frac{1}{2}y + \lambda\frac{\dot{x}}{(\dot{x}^2 + \dot{y}^2)^{1/2}}\right]$$

or

$$0 = \dot{y} - \lambda\frac{d}{dt}\left[\frac{\dot{x}}{(\dot{x}^2 + \dot{y}^2)^{1/2}}\right]$$

and

$$0 = \frac{\partial h}{\partial y} - \frac{d}{dt}\frac{\partial h}{\partial \dot{y}} = -\frac{1}{2}\dot{x} - \frac{d}{dt}\left[\frac{1}{2}x + \lambda\frac{\dot{y}}{(\dot{x}^2 + \dot{y}^2)^{1/2}}\right]$$

or

$$0 = \dot{x} + \lambda\frac{d}{dt}\left[\frac{\dot{y}}{(\dot{x}^2 + \dot{y}^2)^{1/2}}\right].$$

Both of these equations are exact differentials and can be easily integrated to yield

$$y - y_0 = \lambda\frac{\dot{x}}{(\dot{x}^2 + \dot{y}^2)^{1/2}} \qquad \text{and} \qquad x - x_0 = -\lambda\frac{\dot{y}}{(\dot{x}^2 + \dot{y}^2)^{1/2}}$$

where x_0, y_0 are determined by the fixed end points.

Clearly, squaring both equations and adding gives

$$(x - x_0)^2 + (y - y_0)^2 = \lambda^2$$

which is the equation of a circle of radius λ centered at (x_0, y_0). Since the length L is fixed, we have $\lambda = L/2\pi$. The points x_0, y_0 are chosen so that the circle passes through the chosen endpoint (which is the same as the starting point for a closed curve). Also, here is an example of the extremum being a maximum and not a minimum (as is obvious from its physical interpretation).

The last topic we shall address is the problem of algebraic (or non-integral) constraints. These are the types of constraints that usually arise in classical mechanics.

The problem we will consider is this: Suppose we want to minimize the functional $I = \int_{t_0}^{t_1} f(x_i, \dot{x}_i, t)\,dt$ (where $i = 1, \ldots, n$) subject to the constraint $g(x_i, t) = 0$. Be sure to note that g does not depend on the $\dot{x}_i$'s. Exactly as we did before, we introduce two parameters $\varepsilon_1, \varepsilon_2$ together with the varied functions $\widetilde{x}_i(t, \varepsilon_1, \varepsilon_2)$ such that

(1) $\widetilde{x}_i(t_0, \varepsilon_1, \varepsilon_2) = x_{i0}$ and $\widetilde{x}_i(t_1, \varepsilon_1, \varepsilon_2) = x_{i1}$ for all $\varepsilon_1, \varepsilon_2$ (i.e., fixed endpoints);

(2) $\widetilde{x}_i(t, 0, 0) = x_i(t)$ are the desired extremizing functions;

(3) $\widetilde{x}_i(t, \varepsilon_1, \varepsilon_2)$ has continuous derivatives through second order in all variables.

Again form $I(\varepsilon_1, \varepsilon_2) = \int_{t_0}^{t_1} f(\widetilde{x}_i, \dot{\widetilde{x}}_i, t)\, dt$ and require that $(\partial I/\partial \varepsilon_j)_{\varepsilon=0} = 0$ subject to $g(\widetilde{x}_i, t) = 0$. We note that the integral constraint considered previously is a *global* constraint in that after the integration we are left with only a relation between the ε variables. However, $g = 0$ is a *local* constraint because it allows us to write one of the $\widetilde{x}_i$ in terms of the other $n-1$ of the $\widetilde{x}_i$'s. If we were to integrate g we would lose most of its content.

But we can turn it into an integral constraint and still keep its generality if we multiply g by an *arbitary* function $\varphi(t)$ and then integrate. This yields the integral constraint

$$J(\varepsilon_1, \varepsilon_2) = \int_{t_0}^{t_1} \varphi(t) g(\widetilde{x}_i, t)\, dt = 0.$$

Now, just as before, we let

$$K(\varepsilon_1, \varepsilon_2) = I(\varepsilon_1, \varepsilon_2) + \lambda J(\varepsilon_1, \varepsilon_2)$$

and require

$$\left.\frac{\partial K}{\partial \varepsilon_1}\right|_{\varepsilon=0} = 0 = \left.\frac{\partial K}{\partial \varepsilon_2}\right|_{\varepsilon=0}$$

where

$$K(\varepsilon_1, \varepsilon_2) = \int_{t_0}^{t_1} h(\widetilde{x}_i, \dot{\widetilde{x}}_i, t)\, dt$$

and $h = f + \lambda \phi g$.

We now define $\lambda(t) := \lambda \phi(t)$ (which is a completely arbitrary function so far) and hence

$$h(x_i, \dot{x}_i, t) = f(x_i, \dot{x}_i, t) + \lambda(t) g(x_i, t).$$

Proceeding in the usual manner we have (for $j = 1, 2$)

$$\begin{aligned}
\frac{\partial K}{\partial \varepsilon_j} &= \sum_{i=1}^{n} \int_{t_0}^{t_1} \left(\frac{\partial h}{\partial \widetilde{x}_i} \frac{\partial \widetilde{x}_i}{\partial \varepsilon_j} + \frac{\partial h}{\partial \dot{\widetilde{x}}_i} \frac{\partial \dot{\widetilde{x}}_i}{\partial \varepsilon_j} \right) dt \\
&= \sum_{i=1}^{n} \int_{t_0}^{t_1} \left[\frac{\partial h}{\partial \widetilde{x}_i} - \frac{d}{dt}\left(\frac{\partial h}{\partial \dot{\widetilde{x}}_i} \right) \right] \frac{\partial \widetilde{x}_i}{\partial \varepsilon_j}\, dt.
\end{aligned}$$

Write

$$\left.\frac{\partial \widetilde{x}_i}{\partial \varepsilon_j}\right|_{\varepsilon_1=\varepsilon_2=0} := \eta_j^i(t)$$

so that evaluating $(\partial K/\partial\varepsilon_j)_{\varepsilon=0}$ yields (for each $j = 1, 2$)

$$\sum_{i=1}^{n} \int_{t_0}^{t_1} \left[\frac{\partial h}{\partial x_i} - \frac{d}{dt}\left(\frac{\partial h}{\partial \dot{x}_i}\right)\right] \eta_j^i(t)\, dt = 0. \tag{25}$$

However, at this point we have to be careful — the $\eta_j^i(t)$ are *not* all independent. The reason for this is easily seen if we differentiate $g(\widetilde{x}_i, t) = 0$ with respect to ε_j:

$$0 = \frac{\partial g}{\partial \varepsilon_j} = \sum_{i=1}^{n} \frac{\partial g}{\partial \widetilde{x}_i} \frac{\partial \widetilde{x}_i}{\partial \varepsilon_j}$$

and hence for each $j = 1, 2$ we have

$$0 = \left.\frac{\partial g}{\partial \varepsilon_j}\right|_{\varepsilon=0} = \sum_{i=1}^{n} \frac{\partial g}{\partial x_i} \eta_j^i(t)$$

so that for each j the functions $\eta_j^i(t)$ are linearly related.

To proceed, we take advantage of the fact that $\lambda(t)$ is still unspecified. Writing out equation (25) we have (remember g is independent of $\dot{x}_i$)

$$\sum_{i=1}^{n} \int_{t_0}^{t_1} \left[\frac{\partial f}{\partial x_i} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}_i}\right) + \lambda(t)\frac{\partial g}{\partial x_i}\right] \eta_j^i(t)\, dt = 0. \tag{26}$$

Choose $i = 1$ (a completely arbitrary choice) and assume that $\eta_j^1(t)$ is given in terms of the rest of the $\eta_j^i(t)$ (where j is fixed). We choose $\lambda(t)$ so that

$$\frac{\partial f}{\partial x_1} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}_1}\right) + \lambda(t)\frac{\partial g}{\partial x_1} = 0.$$

This means we don't need to know η_j^1 because its coefficient in equation (26) is zero anyway, and we are left with

$$\sum_{i=2}^{n} \int_{t_0}^{t_1} \left[\frac{\partial f}{\partial x_i} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}_i}\right) + \lambda(t)\frac{\partial g}{\partial x_i}\right] \eta_j^i(t)\, dt = 0$$

where there is no $i = 1$ term in the sum.

Now the remaining $n - 1$ of the $\eta_j^i(t)$ are completely independent so we may equate their coefficients in (26) to zero:

$$\frac{\partial f}{\partial x_i} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}_i}\right) + \lambda(t)\frac{\partial g}{\partial x_i} = 0 \qquad \text{for } i = 2, \ldots, n.$$

This equation is identical to the result for $i = 1$ even though it was obtained in a completely different manner. In any case, combining these results we have

$$\frac{\partial h}{\partial x_i} - \frac{d}{dt}\left(\frac{\partial h}{\partial \dot{x}_i}\right) = 0 \qquad \text{for all } i = 1, \ldots, n \tag{27}$$

where

$$h(x_i, \dot{x}_i, t) = f(x_i, \dot{x}_i, t) + \lambda(t) g(x_i, t). \tag{28}$$

Note that no derivative of $\lambda(t)$ appears because g is independent of $\dot{x}_i$.

In the case of multiple constraints it should be obvious that we have

$$\frac{\partial f}{\partial x_i} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}_i}\right) + \sum_j \lambda_j(t) \frac{\partial g_j}{\partial x_i} = 0. \tag{29}$$

Also be sure to realize that in the case of integral constraints, the Lagrange multipliers λ are *constants*, whereas in the case of algebraic constraints, the multipliers $\lambda(t)$ are *functions* of the independent variable t.

Example 11. There are a number of ways to show that the shortest path between two points on the surface of a sphere is a great circle. In this example we will prove this using Euler's equation with Lagrange multipliers.

On a sphere of radius a, the path length is given by

$$s = \int ds = \int \sqrt{dx^2 + dy^2 + dz^2} = \int \sqrt{1 + y'^2 + z'^2}\, dx$$

subject to the constraint

$$g(x, y, z) = x^2 + y^2 + z^2 - a^2 = 0.$$

Here x is the independent variable, so we have equation (29) for both y and z:

$$-\frac{d}{dx}\left[\frac{y'}{(1 + y'^2 + z'^2)^{1/2}}\right] + 2\lambda(x) y = 0$$

and

$$-\frac{d}{dx}\left[\frac{z'}{(1 + y'^2 + z'^2)^{1/2}}\right] + 2\lambda(x) z = 0.$$

Solving both of these for 2λ and equating them we obtain

$$\frac{1}{y}\frac{d}{dx}\left[\frac{y'}{(1 + y'^2 + z'^2)^{1/2}}\right] - \frac{1}{z}\frac{d}{dx}\left[\frac{z'}{(1 + y'^2 + z'^2)^{1/2}}\right] = 0$$

which becomes

$$z[y''(1 + y'^2 + z'^2) - y'(y'y'' + z'z'')] - y[z''(1 + y'^2 + z'^2) - z'(y'y'' + z'z'')] = 0$$

or simply

$$zy'' + (yy' + zz')z'y'' - yz'' - (yy' + zz')y'z'' = 0.$$

Taking the derivative of the constraint equation with respect to x yields $x + yy' + zz' = 0$ so that $yy' + zz' = -x$. Substituting this into the previous equation results in

$$(z - xz')y'' = (y - xy')z''. \tag{30}$$

It's not immediately obvious, but this is the intersection of the sphere with a plane passing through the origin. To see this, recall that the equation of a plane through the point $\mathbf{x}_0$ and perpendicular to the vector $\mathbf{n}$ is the set of points $\mathbf{x}$ satisfying $(\mathbf{x}-\mathbf{x}_0)\cdot\mathbf{n} = 0$. In this case we want $\mathbf{x}_0$ to be the origin, and if we write $\mathbf{n} = (n_x, n_y, n_z)$ then the equation of the plane becomes $xn_x + yn_y + zn_z = 0$ which is of the form $Ax + By = z$. Taking the derivative with respect to x gives $A + By' = z'$ and taking the derivative again gives $By'' = z''$. Therefore $B = z''/y''$ so that $A = z' - y'z''/y''$. Using these, the equation of a plane becomes exactly equation (30). In other words, equation (30) represents the intersection of a plane through the origin and the sphere. This defines a great circle on the sphere.

8 Extremization of Multiple Integrals

We now turn our attention to the problem of extremizing an integral with more than one *independent* variable. We start with the simplest case of two such variables.

So, let $y = y(x_1, x_2)$ be a function of two real variables, and consider the integral

$$I = \int_D f(y, \partial_1 y, \partial_2 y, x_1, x_2)\, dx_1 dx_2$$

where we use the shorthand notation $\partial_i = \partial/\partial x_i$. The integral is over some domain D in the x_1x_2-plane, and we let the boundary of this domain be denoted by ∂D. (This is standard notation for the boundary, and does not refer to the partial derivative of anything.) In our previous cases, we specified the value of $x(t)$ at the endpoints t_1, t_2 of an interval. Now we must specify the value of $y(x_1, x_2)$ on ∂D, and hence we write

$$y(x_1, x_2)\big|_{\partial D} = g(\partial D)$$

where $g(\partial D)$ is some function defined on the one-dimensional curve ∂D.

We again introduce a one-parameter family of test functions $\widetilde{y}$ that have the properties

(1) $\widetilde{y}(x_1, x_2, \varepsilon)\big|_{\partial D} = g(\partial D)$ for all ε;

(2) $\widetilde{y}(x_1, x_2, 0) = y(x_1, x_2)$, the desired extremizing function;

(3) $\widetilde{y}(x_1, x_2, \varepsilon)$ has continuous first and second derivatives.

As before, we form

$$I(\varepsilon) = \int_D f(\widetilde{y}, \partial_1\widetilde{y}, \partial_2\widetilde{y}, x_1, x_2)\, dx_1 dx_2$$

and require that

$$\left.\frac{dI}{d\varepsilon}\right|_{\varepsilon=0} = 0.$$

Then we have

$$\begin{aligned}\frac{dI}{d\varepsilon} &= \int_D \left[\frac{\partial f}{\partial \widetilde{y}}\frac{d\widetilde{y}}{d\varepsilon} + \frac{\partial f}{\partial(\partial_1 \widetilde{y})}\frac{d}{d\varepsilon}\left(\frac{\partial \widetilde{y}}{\partial x_1}\right) + \frac{\partial f}{\partial(\partial_2 \widetilde{y})}\frac{d}{d\varepsilon}\left(\frac{\partial \widetilde{y}}{\partial x_2}\right)\right] dx_1 dx_2 \\ &= \int_D \left[\frac{\partial f}{\partial \widetilde{y}}\frac{d\widetilde{y}}{d\varepsilon} + \frac{\partial f}{\partial(\partial_1 \widetilde{y})}\frac{\partial}{\partial x_1}\left(\frac{d\widetilde{y}}{d\varepsilon}\right) + \frac{\partial f}{\partial(\partial_2 \widetilde{y})}\frac{\partial}{\partial x_2}\left(\frac{d\widetilde{y}}{d\varepsilon}\right)\right] dx_1 dx_2\end{aligned}$$

To effect the integration by parts in this case, we again refer to Green's theorem in the form

$$\int_D \left[\frac{\partial P}{\partial x_1} + \frac{\partial Q}{\partial x_2}\right] dx_1 dx_2 = \oint_{\partial D} P\, dx_2 - Q\, dx_1.$$

Let us write

$$P = R(x_1, x_2)A(x_1, x_2) \quad \text{and} \quad Q = R(x_1, x_2)B(x_1, x_2)$$

so that Green's Theorem may be written as

$$\begin{aligned}&\int_D \left[A\frac{\partial R}{\partial x_1} + B\frac{\partial R}{\partial x_2}\right] dx_1 dx_2 \\ &\qquad\qquad = -\int_D \left[\frac{\partial A}{\partial x_1} + \frac{\partial B}{\partial x_2}\right] R\, dx_1 dx_2 + \int_{\partial D} [A\, dx_2 - B\, dx_1] R.\end{aligned}$$

To apply this result to the second and third terms in the integrand of $dI/d\varepsilon$, we let

$$R = \frac{d\widetilde{y}}{d\varepsilon} \qquad A = \frac{\partial f}{\partial(\partial_1 \widetilde{y})} \qquad B = \frac{\partial f}{\partial(\partial_2 \widetilde{y})}$$

to obtain

$$\begin{aligned}\frac{dI}{d\varepsilon} &= \int_D \left[\frac{\partial f}{\partial \widetilde{y}} - \frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial(\partial_1 \widetilde{y})}\right) - \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial(\partial_2 \widetilde{y})}\right)\right] \frac{d\widetilde{y}}{d\varepsilon}\, dx_1 dx_2 \\ &\qquad + \int_{\partial D} \frac{d\widetilde{y}}{d\varepsilon}\left[\frac{\partial f}{\partial(\partial_1 \widetilde{y})}\, dx_2 - \frac{\partial f}{\partial(\partial_2 \widetilde{y})}\, dx_1\right]\end{aligned}$$

But by property (1) above we have $d\widetilde{y}/d\varepsilon = 0$ on ∂D, and hence the second integral above vanishes. Using our extremum condition $dI/d\varepsilon|_{\varepsilon=0} = 0$ we are left with

$$\int_D \left[\frac{\partial f}{\partial y} - \frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial(\partial_1 y)}\right) - \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial(\partial_2 y)}\right)\right] \eta(x_1, x_2)\, dx_1 dx_2 = 0$$

where

$$\eta(x_1, x_2) = \left.\frac{d\widetilde{y}}{d\varepsilon}\right|_{\varepsilon=0}.$$

Since $\eta(x_1, x_2)$ is an arbitrary function on D, we have the Euler-Lagrange equation for two independent variables

$$\frac{\partial f}{\partial y} - \frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial(\partial_1 y)}\right) - \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial(\partial_2 y)}\right) = 0. \tag{31}$$

Example 12. Let us use equation (31) to derive the wave equation for a vibrating string. Our approach will be to use Hamilton's principle as mentioned in Example 3.

We consider a perfectly flexible elastic string stretched under *constant* tension τ along the x-axis with endpoints fixed at $x = 0$ and $x = l$. Assume that the string undergoes small, transverse vibrations in the absence of gravity. Let the amplitude (i.e., string displacement) of the vibrations at time t be $y(x, t)$ where small vibrations means $|\partial y/\partial x| \ll 1$. The velocity at any point is given by $\partial y/\partial t$, and the fixed endpoint condition is $y(0, t) = y(l, t) = 0$ for all t.

Next we must find the Lagrangian for the string. Let $\rho(x)$ be the mass density of the string. Then an element dx of string has mass $\rho(x)\, dx$ with kinetic energy

$$dT = \frac{1}{2}\rho(x)dx\left(\frac{\partial y}{\partial t}\right)^2$$

and hence the total kinetic energy of the string is given by

$$T = \frac{1}{2}\int_o^l \rho(x)\left(\frac{\partial y}{\partial t}\right)^2 dx.$$

Now, potential energy is defined as the work required to put the system into a given configuration. In this case, the work required to stretch the string from its equilibrium length l to another length s against the *constant* force τ is $V = \tau(s - l) = \tau(\int ds - l)$ or

$$V = \tau\left[\int_0^l \sqrt{dx^2 + dy^2} - l\right] = \tau\left[\int_0^l \sqrt{1 + (\partial y/\partial x)^2}\, dx - l\right].$$

But we are assuming $|\partial y/\partial x| \ll 1$ so that

$$\sqrt{1 + (\partial y/\partial x)^2} \cong 1 + \frac{1}{2}\left(\frac{\partial y}{\partial x}\right)^2$$

and hence

$$V = \tau\left[\int_0^l \left\{1 + \frac{1}{2}\left(\frac{\partial y}{\partial x}\right)^2\right\} dx - l\right] = \frac{1}{2}\tau\int_0^l \left(\frac{\partial y}{\partial x}\right)^2 dx.$$

The Lagrangian $L = T - V$ now becomes

$$L = \int_0^l \frac{1}{2}\left[\rho(x)\left(\frac{\partial y}{\partial t}\right)^2 - \tau\left(\frac{\partial y}{\partial x}\right)^2\right] dx.$$

The integrand

$$\mathscr{L} = \frac{1}{2}\left[\rho(x)\left(\frac{\partial y}{\partial t}\right)^2 - \tau\left(\frac{\partial y}{\partial x}\right)^2\right]$$

is called the **Lagrangian density** because its spatial integral is just L. Hamilton's principle says that the actual motion of the system is given by the path that extremizes the integral

$$I = \int L\,dt = \int \mathscr{L}\,dx\,dt.$$

(The integral $\int L\,dt$ is frequently called the **action** and denoted by S.) From equation (31) we have

$$\frac{\partial \mathscr{L}}{\partial y} - \frac{\partial}{\partial x}\left(\frac{\partial \mathscr{L}}{\partial(\partial_x y)}\right) - \frac{\partial}{\partial t}\left(\frac{\partial \mathscr{L}}{\partial(\partial_t y)}\right) = 0$$

and using our Lagrangian density this becomes

$$0 - \tau\frac{\partial^2 y}{\partial x^2} - \rho(x)\frac{\partial^2 y}{\partial t^2} = 0$$

or simply

$$\frac{\partial^2 y}{\partial x^2} = \frac{\rho(x)}{\tau}\frac{\partial^2 y}{\partial t^2}.$$

This is the equation of motion (the **wave equation**) for the string.

Without working through the details, it is easy to write down some generalizations of this formalism. First suppose that we have n independent variables $x_1, \ldots, x_n$. We want to extremize the integral

$$I = \int_D \mathscr{L}(y, \partial_1 y, \ldots, \partial_n y, x_1, \ldots, x_n)\,d^n x$$

where $y = y(x_1, \ldots, x_n)$, D is some region of $\mathbb{R}^n$, $d^n x = dx_1 \cdots dx_n$ is the n-dimensional volume element, and y is given some prescribed value on the $(n-1)$-dimensional boundary ∂D. Then the Euler-Lagrange equation which gives the extremizing path y is

$$\frac{\delta \mathscr{L}}{\delta y} := \frac{\partial \mathscr{L}}{\partial y} - \sum_{i=1}^{n} \frac{\partial}{\partial x_i}\left(\frac{\partial \mathscr{L}}{\partial(\partial_i y)}\right) = 0 \qquad (32)$$

where we have defined the **functional derivative** $\delta\mathscr{L}/\delta y$ by this equation.

Actually, using variational notation and the summation convention, this is easy to derive directly. Indeed, we have

$$\begin{aligned}\delta I &= \int_D \delta\mathscr{L}(y, \partial_1 y, \ldots, \partial_n y, x_1, \ldots, x_n)\,d^n x \\ &= \int_D \left[\frac{\partial \mathscr{L}}{\partial y}\delta y + \frac{\partial \mathscr{L}}{\partial(\partial_i y)}\delta\partial_i y\right] d^n x\end{aligned}$$

$$= \int_D \left[\frac{\partial \mathscr{L}}{\partial y}\delta y + \frac{\partial \mathscr{L}}{\partial(\partial_i y)}\partial_i \delta y \right] d^n x$$

$$= \int_D \left[\frac{\partial \mathscr{L}}{\partial y}\delta y + \partial_i\left(\frac{\partial \mathscr{L}}{\partial(\partial_i y)}\delta y \right) - \partial_i\left(\frac{\partial \mathscr{L}}{\partial(\partial_i y)} \right)\delta y \right] d^n x.$$

But the n-dimensional version of the Divergence Theorem reads

$$\int_D \partial_i f^i \, d^n x = \int_{\partial D} f^i da_i$$

where da_i is the ith component of the n-dimensional area element. (This is just $\int_D \nabla \cdot \mathbf{f} \, d^n x = \int_{\partial D} \mathbf{f} \cdot d\mathbf{a}$.) Therefore the middle term in the integrand becomes a surface integral over ∂D where δy is assumed to vanish, and hence we are left with

$$\delta I = \int_D \left[\frac{\partial \mathscr{L}}{\partial y} - \partial_i\left(\frac{\partial \mathscr{L}}{\partial(\partial_i y)} \right) \right] \delta y \, d^n x$$

Since δy is arbitrary, we see that $\delta I = 0$ implies

$$\frac{\partial \mathscr{L}}{\partial y} - \partial_i\left(\frac{\partial \mathscr{L}}{\partial(\partial_i y)} \right) = 0$$

which is just equation (32).

As a passing remark, in quantum field theory we would have $d^n x = d^4 x = dx^0 dx^1 dx^2 dx^3 = dt \, dx \, dy \, dz$, and the dependent variable y is frequently written as φ in the case of a scalar field. Then the Euler-Lagrange for the quantum field φ is written as

$$\frac{\partial \mathscr{L}}{\partial \varphi} - \partial_\mu\left(\frac{\partial \mathscr{L}}{\partial(\partial_\mu \varphi)} \right) = 0$$

or

$$\frac{\partial \mathscr{L}}{\partial \varphi} - \partial_\mu\left(\frac{\partial \mathscr{L}}{\partial \varphi_{,\mu}} \right) = 0 \tag{33}$$

where it is conventional to use Greek letters to indicate a summation over the four spacetime variables, and $\varphi_{,\mu} = \partial_\mu \varphi$.

Example 13. Consider the Lagrangian density

$$\mathscr{L} = \frac{1}{2}\big[(\partial_\alpha \varphi)(\partial^\alpha \varphi) - m^2 \varphi^2\big]$$
$$= \frac{1}{2}\big[g^{\alpha\beta}(\partial_\alpha \varphi)(\partial_\beta \varphi) - m^2 \varphi^2\big].$$

where m is a constant. Then

$$\frac{\partial \mathscr{L}}{\partial \varphi} = -m^2 \varphi$$

while

$$\frac{\partial \mathscr{L}}{\partial(\partial_\mu \varphi)} = \frac{1}{2} g^{\alpha\beta}(\delta^\mu_\alpha\, \partial_\beta \varphi + \partial_\alpha \varphi\, \delta^\mu_\beta) = \partial^\mu \varphi.$$

Applying equation (33) we now have

$$\partial_\mu \partial^\mu \varphi + m^2 \varphi = 0$$

which is called the **Klein-Gordon equation**, and is a relativistically invariant equation that describes a spinless particle of mass m. The operator $\partial_\mu \partial^\mu$ is called the **d'Alembertian** and is frequently written in various forms as $\partial_\mu \partial^\mu = \partial \cdot \partial = \partial^2 = \Box^2 = \Box$, and the Klein-Gordon equation is often written as

$$(\Box + m^2)\varphi = 0.$$

As an aside, this equation follows from the relativistic expression $E^2 = \mathbf{p}^2 c^2 + m^2 c^4$ and the usual quantum mechanical substitutions $E \to i\hbar(\partial/\partial t)$ and $\mathbf{p} \to -i\hbar\nabla$ so that

$$-\hbar^2 \frac{\partial^2 \varphi}{\partial t^2} = (-c^2 \hbar^2 \nabla^2 + m^2 c^4)\varphi$$

or

$$\left(\frac{\partial^2}{\partial(ct)^2} - \nabla^2 + \frac{m^2 c^2}{\hbar^2}\right)\varphi = 0.$$

Letting $\mu = mc/\hbar$ (a frequently used notation) this is

$$\left(\frac{\partial^2}{\partial(x^0)^2} - \boldsymbol{\nabla}^2 + \mu^2\right)\varphi = (\Box + \mu^2)\varphi = 0$$

which is the same as $(\Box + m^2)\varphi = 0$ if we are using the units $c = \hbar = 1$.

Carrying this even further, suppose that in addition we have m dependent variables $y_1, \ldots, y_m$ where $y_i = y_i(x_1, \ldots, x_n)$. Then extremizing the integral

$$I = \int_D \mathscr{L}(y_j, \partial_i y_j, x_i) d^n x$$

leads to the set of m coupled second order PDEs

$$\frac{\partial \mathscr{L}}{\partial y_j} - \sum_{i=1}^{n} \frac{\partial}{\partial x_i}\left(\frac{\partial \mathscr{L}}{\partial(\partial_i y_j)}\right) = 0, \qquad j = 1, \ldots, m. \tag{34}$$

Finally, suppose we have m dependent variables, n independent variables and p integral equations of constraint of the form

$$J_k = \int_D g_k(y_j, \partial_i y_j, x_i)\, d^n x = \text{const}$$

where $i = 1, \ldots, n; j = 1, \ldots, m$ and $k = 1, \ldots, p$. We again want to extremize the integral

$$I = \int_D \mathscr{L}(y_j, \partial_i y_j, x_i)\, d^n x$$

subject to these constraints. In analogy to what we have done previously, let us form the function

$$h = \mathscr{L} + \sum_{k=1}^{p} \lambda_k g_k$$

where the λ_k are constant Lagrange multipliers. Now h will satisfy the system of Euler-Lagrange equations

$$\frac{\partial h}{\partial y_j} - \sum_{i=1}^{n} \frac{\partial}{\partial x_i}\left(\frac{\partial h}{\partial(\partial_i y_j)}\right) = 0, \qquad j = 1, \ldots, m. \tag{35}$$

Example 14. Consider the Lagrange density

$$\mathscr{L} = \frac{\hbar^2}{2m}\left[\frac{\partial\psi^*}{\partial x}\frac{\partial\psi}{\partial x} + \frac{\partial\psi^*}{\partial y}\frac{\partial\psi}{\partial y} + \frac{\partial\psi^*}{\partial z}\frac{\partial\psi}{\partial z}\right] + V(x,y,z)\psi^*\psi$$

where $\psi = \psi(x,y,z)$ is a complex-valued function and $V(x,y,z)$ is real. We want to extremize the functional

$$I = \int_D \mathscr{L}\, d^3x$$

subject to the constraint

$$J = \int_D \psi^*\psi\, d^3x = 1.$$

Note this is a variational problem with three independent variables (x,y,z), two dependent variables (the real and imaginary parts $\psi_R = y_1$ and $\psi_I = y_2$ of ψ) and one constraint.

Observing that, e.g.,

$$\begin{aligned}\frac{\partial\psi^*}{\partial x}\frac{\partial\psi}{\partial x} &= \left(\frac{\partial\psi_R}{\partial x} - i\frac{\partial\psi_I}{\partial x}\right)\left(\frac{\partial\psi_R}{\partial x} + i\frac{\partial\psi_I}{\partial x}\right) \\ &= \left(\frac{\partial\psi_R}{\partial x}\right)^2 + \left(\frac{\partial\psi_I}{\partial x}\right)^2\end{aligned}$$

and

$$\psi^*\psi = \psi_R^2 + \psi_I^2$$

we form

$$
\begin{aligned}
h &= \mathscr{L} + \lambda\psi^*\psi \\
&= \frac{\hbar^2}{2m}\left[\left(\frac{\partial\psi_R}{\partial x}\right)^2 + \left(\frac{\partial\psi_R}{\partial y}\right)^2 + \left(\frac{\partial\psi_R}{\partial z}\right)^2 + \left(\frac{\partial\psi_I}{\partial x}\right)^2 + \left(\frac{\partial\psi_I}{\partial y}\right)^2 + \left(\frac{\partial\psi_I}{\partial z}\right)^2\right] \\
&\quad + (V(x,y,z) + \lambda)(\psi_R^2 + \psi_I^2) \\
&= \frac{\hbar^2}{2m}\sum_{i=1}^{3}\left[\left(\frac{\partial\psi_R}{\partial x_i}\right)^2 + \left(\frac{\partial\psi_I}{\partial x_i}\right)^2\right] + (V+\lambda)(\psi_R^2 + \psi_I^2).
\end{aligned}
$$

Using this in equation (35) with $y_1 = \psi_R$ and $y_2 = \psi_I$ we obtain the equations

$$
V\psi_R + \lambda\psi_R - \frac{\hbar^2}{2m}\sum_{i=1}^{3}\frac{\partial}{\partial x_i}\left(\frac{\partial\psi_R}{\partial x_i}\right) = V\psi_R + \lambda\psi_R - \frac{\hbar^2}{2m}\nabla^2\psi_R = 0
$$

and

$$
V\psi_I + \lambda\psi_I - \frac{\hbar^2}{2m}\sum_{i=1}^{3}\frac{\partial}{\partial x_i}\left(\frac{\partial\psi_I}{\partial x_i}\right) = V\psi_I + \lambda\psi_I - \frac{\hbar^2}{2m}\nabla^2\psi_I = 0.
$$

Writing $\lambda = -E$ we can combine these into the single equation

$$
-\frac{\hbar^2}{2m}\nabla^2\psi + V\psi = E\psi
$$

which you should recognize as the time-independent Schrödinger equation.

9 Symmetries and Noether's Theorem

In this section we will derive the discrete mechanical version of the famous theorem due to Emmy Noether. In other words, we will consider only systems with a single independent variable x and a finite number of dependent variables $y_1, \dots, y_n$. While this will allow us to derive some of the basic conservation laws of classical mechanics, the real power and use of Noether's theorem lies in its formulation in field theory in 4-dimensional spacetime. We will take up that formulation in the following section.

Before diving into the general theory of variations, let us first take a look at the particular case of a function $f = f(x_i, \dot{x}_i, t)$ that satisfies the Euler-Lagrange equations

$$
\frac{\partial f}{\partial x_i} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}_i}\right) = 0, \qquad i = 1, \dots, n
$$

and where f has no explicit time dependence, i.e, $\partial f/\partial t = 0$. Then an arbitrary variation in f yields

$$\begin{aligned}\delta f &= \sum_{i=1}^{n}\left[\frac{\partial f}{\partial x_i}\delta x_i + \frac{\partial f}{\partial \dot{x}_i}\delta \dot{x}_i\right] = \sum_{i=1}^{n}\left[\frac{\partial f}{\partial x_i}\delta x_i + \frac{\partial f}{\partial \dot{x}_i}\frac{d}{dt}\delta x_i\right] \\ &= \sum_{i=1}^{n}\left[\frac{\partial f}{\partial x_i}\delta x_i + \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}_i}\delta x_i\right) - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}_i}\right)\delta x_i\right].\end{aligned}$$

The first and last terms cancel by the Euler-Lagrange equation and we are left with

$$\delta f = \frac{d}{dt}\sum_{i=1}^{n}\frac{\partial f}{\partial \dot{x}_i}\delta x_i.$$

Now, if f is invariant under the transformation $x_i \to x_i + \delta x_i$, then by definition we have $\delta f = 0$ so that

$$\sum_{i=1}^{n}\frac{\partial f}{\partial \dot{x}_i}\delta x_i = \text{const.}$$

This is a simple version of Noether's theorem: a symmetry (invariance) of a function satisfying the Euler-Lagrange equation leads to a conserved quantity.

With this simple example as a model, we now turn our attention to the more general problem of arbitrarily varied paths.

In all of our work so far we have only considered variations where the varied paths were given fixed, specified values on the boundary regions (e.g., fixed endpoints for a path $x(t)$), and the independent variables themselves were never varied. Now we want to relax both of these requirements.

For example, we could ask for the path that minimizes the travel time taken by a particle moving under gravity from a fixed point to a given vertical line. In this case, one end is fixed and the other is free to lie anywhere on the vertical line. Also, we will see that if the independent variable is the time t, then invariance of the Lagrangian under time translation leads to a conserved Hamiltonian (and hence to conservation of energy if $H = E$).

Up to this point, we have looked at the integral (with a slight change of notation)

$$I = \int_{x_a}^{x_b} f(y(x), y'(x), x)\, dx \tag{36}$$

under variations $y(x) \to \widetilde{y}(x, \varepsilon)$ with fixed endpoints, and required that the integral be an extremum with respect to ε. This resulted in the Euler-Lagrange equation (a second-order differential equation for $y(x)$). Now we will look at this integral under more general variations where the boundary terms do not necessarily vanish, and where the independent variable x is also varied. Furthermore, instead of requiring that the integral be an extremum, we will investigate the consequences of requiring that it be *invariant* under the transformations. In other words, if $I \to \bar{I}(\varepsilon)$, then we will require that $\bar{I}(\varepsilon) = I$. This will give us a first-order equation that is a generalization of the second form of Euler's equation (see equation (14)).

Recall that our derivations dealt in general with a family of functions $\widetilde{y}(x, \varepsilon)$. A particular example of this was a choice of the form $\widetilde{y}(x, \varepsilon) = y(x) + \varepsilon\eta(x)$ where $\eta(x_a) = \eta(x_b) = 0$. Now we consider more general variations where η can depend on the dependent variables as well, and where the independent variable is also varied. Thus we will consider the variations

$$y(x) \to \widetilde{y}(x, y, y') = y(x) + \varepsilon\eta(x, y, y') \tag{37}$$

and

$$x \to \bar{x} = x + \varepsilon\xi(x, y, y') \tag{38}$$

and use them to form the integral

$$\bar{I}(\varepsilon) = \int_{\bar{x}_a}^{\bar{x}_b} f(\widetilde{y}(\bar{x}), \widetilde{y}'(\bar{x}), \bar{x})\, d\bar{x}. \tag{39}$$

Be sure to note that $\bar{I}(\varepsilon = 0) = I$. Also note that the limits of integration here are defined by evaluating equation (38) at the endpoints x_a and x_b.

It is important to understand just what the function $\widetilde{y}(\bar{x})$ really means. It does *not* mean to replace x by $\bar{x}$ in equation (37). It means that we use equation (38) to solve for x as a function of $\bar{x}$, and then use *that* x in equation (37). Also, since we will be expanding $\bar{I}(\varepsilon)$ only through first order in ε, we can freely switch between x and $\bar{x}$ in the functions η and ξ. This is because η and ξ already have a factor of ε in front of them, so their arguments need only be to zeroth order in ε, and in zeroth order x and $\bar{x}$ are the same. Furthermore, a prime denotes differentiation with respect to the appropriate independent variable. Thus the prime on $\widetilde{y}'(\bar{x})$ denotes differentiation with respect to $\bar{x}$, while the prime on $y'(x)$ denotes differentiation with respect to x.

Now let us follow through with this program. We first use equation (38) to write x in terms of $\bar{x}$, keeping in mind what we just said about the argument of ξ and η. Then $x = \bar{x} - \varepsilon\xi(\bar{x}, y, y')$ so that equation (37) becomes

$$\widetilde{y}(\bar{x}) = y(\bar{x} - \varepsilon\xi(\bar{x}, y, y')) + \varepsilon\eta(\bar{x}, y, y')$$

where y and y' are functions of $\bar{x}$. Expanding through first order in ε we have

$$\widetilde{y}(\bar{x}) = y(\bar{x}) - \varepsilon\xi(\bar{x}, y, y')\frac{dy}{d\bar{x}} + \varepsilon\eta(\bar{x}, y, y')$$

or

$$\widetilde{y}(\bar{x}) = y(\bar{x}) + \varepsilon\rho(\bar{x}, y, y') \tag{40}$$

where for simplicity we have defined

$$\rho(\bar{x}, y, y') = \eta(\bar{x}, y, y') - \xi(\bar{x}, y, y')\frac{dy}{d\bar{x}}.$$

Using equation (40) in (39) we then have

$$\bar{I}(\varepsilon) = \int_{\bar{x}_a}^{\bar{x}_b} f(y(\bar{x}) + \varepsilon\rho(\bar{x}, y, y'), y'(\bar{x}) + \varepsilon\rho'(\bar{x}, y, y'), \bar{x})\, d\bar{x}.$$

From equation (38) we write

$$\bar{x}_a = x_a + \varepsilon\xi(x,y,y')\big|_{x=x_a} := x_a + \delta_a$$
$$\bar{x}_b = x_b + \varepsilon\xi(x,y,y')\big|_{x=x_b} := x_b + \delta_b$$

and then break up the integral as follows:

$$\int_{\bar{x}_a}^{\bar{x}_b} = \int_{x_a+\delta_a}^{x_b+\delta_b} = \int_{x_a+\delta_a}^{x_b} + \int_{x_b}^{x_b+\delta_b} = \int_{x_a}^{x_b} - \int_{x_a}^{x_a+\delta_a} + \int_{x_b}^{x_b+\delta_b}.$$

Since the second and third integrals on the right hand side of this equation have an integration range proportional to ε (by definition of δ_a and δ_b), we can drop all terms in their integrands that are of order ε or higher. Then we can write

$$\bar{I}(\varepsilon) = \int_{x_a}^{x_b} f(y(\bar{x}) + \varepsilon\rho(\bar{x},y,y'), y'(\bar{x}) + \varepsilon\rho'(\bar{x},y,y'), \bar{x})\, d\bar{x}$$
$$+ \int_{x_b}^{x_b+\delta_b} f(y(\bar{x}), y'(\bar{x}), \bar{x})\, d\bar{x} - \int_{x_a}^{x_a+\delta_a} f(y(\bar{x}), y'(\bar{x}), \bar{x})\, d\bar{x}.$$

The first of these integrals we expand through order ε. For the second and third, note that everything is continuous and the integration range is infinitesimal. Then each of these two integrals is just the integration range times the integrand evaluated at x_b and x_a respectively. This yields

$$\bar{I}(\varepsilon) = \int_{x_a}^{x_b} f(y(\bar{x}), y'(\bar{x}), \bar{x})\, d\bar{x}$$
$$+ \varepsilon \int_{x_a}^{x_b} \left[\rho(\bar{x},y,y')\frac{\partial}{\partial y} f(y,y',\bar{x}) + \rho'(\bar{x},y,y')\frac{\partial}{\partial y'} f(y,y',\bar{x})\right] d\bar{x}$$
$$+ \delta_b f(y,y',\bar{x})\big|_{\bar{x}=x_b} - \delta_a f(y,y',\bar{x})\big|_{\bar{x}=x_a}.$$

Now consider the second term in the second integral. Integrating by parts and using the definition of ρ we have

$$\int_{x_a}^{x_b} \rho' \frac{\partial f}{\partial y'}\, d\bar{x} = \int_{x_a}^{x_b} \left[\frac{d}{d\bar{x}}\left(\rho\frac{\partial f}{\partial y'}\right) - \rho\frac{d}{d\bar{x}}\left(\frac{\partial f}{\partial y'}\right)\right] d\bar{x}$$
$$= \left(\eta - \xi\frac{dy}{d\bar{x}}\right)\frac{\partial f}{\partial y'}\bigg|_{\bar{x}=x_a}^{\bar{x}=x_b} - \int_{x_a}^{x_b}\left(\eta - \xi\frac{dy}{d\bar{x}}\right)\frac{d}{d\bar{x}}\left(\frac{\partial f}{\partial y'}\right) d\bar{x}$$

Using the definitions of δ_a, δ_b and ρ, we can now write

$$\bar{I}(\varepsilon) = \int_{x_a}^{x_b} f(y(\bar{x}), y'(\bar{x}), \bar{x})\, d\bar{x}$$
$$+ \varepsilon \int_{x_a}^{x_b} \left(\eta - \xi\frac{dy}{d\bar{x}}\right)\left[\frac{\partial f}{\partial y} - \frac{d}{d\bar{x}}\left(\frac{\partial f}{\partial y'}\right)\right] d\bar{x}$$
$$+ \varepsilon\left[\left(\eta - \xi\frac{dy}{d\bar{x}}\right)\frac{\partial f}{\partial y'} + \xi f\right]_{x_a}^{x_b}$$

But now realize that $\bar{x}$ has become just a dummy variable of integration because the limits of integration only depend on x_a and x_b. This means we can replace $\bar{x}$ by simply x throughout, and in particular the first integral is just $I = \bar{I}(\varepsilon = 0)$. Therefore we may write

$$\bar{I}(\varepsilon) - I = \varepsilon \int_{x_a}^{x_b} \left[\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'} \right) \right] (\eta - \xi y')\, dx \\ + \varepsilon \left[\eta \frac{\partial f}{\partial y'} + \xi \left(f - y' \frac{\partial f}{\partial y'} \right) \right]_{x_a}^{x_b}. \quad (41)$$

Let's look at a special case of this result. If we let $\xi = 0$ (which means that we don't vary the independent variable) and require that $\eta(x, y, y')\big|_{x_a} = \eta(x, y, y')\big|_{x_b} = 0$ (so the endpoints are fixed) then we obtain

$$\bar{I}(\varepsilon) - \bar{I}(0) = \varepsilon \int_{x_a}^{x_b} \left[\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'} \right) \right] \eta\, dx.$$

If we now require that

$$\lim_{\varepsilon \to 0} \frac{\bar{I}(\varepsilon) - \bar{I}(0)}{\varepsilon} = \frac{d\bar{I}}{d\varepsilon}\bigg|_{\varepsilon = 0} = 0$$

then the fact that η is arbitrary implies that we arrive at the usual Euler-Lagrange equation

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'} \right) = 0.$$

It should be clear that this special case just duplicates the conditions of the original Euler-Lagrange problem, and hence we should expect this result.

It should also be easy to see that the generalization of equation (41) to the case of several dependent variables $y_1, \dots, y_n$ yields

$$\bar{I}(\varepsilon) - I = \varepsilon \int_{x_a}^{x_b} \sum_{i=1}^{n} \left[\frac{\partial f}{\partial y_i} - \frac{d}{dx}\left(\frac{\partial f}{\partial y_i'} \right) \right] (\eta_i - \xi y_i')\, dx \\ + \varepsilon \left[\sum_{i=1}^{n} \eta_i \frac{\partial f}{\partial y_i'} + \xi \left(f - \sum_{i=1}^{n} y_i' \frac{\partial f}{\partial y_i'} \right) \right]_{x_a}^{x_b} \quad (42)$$

where

$$\begin{aligned} I &= \int_{x_a}^{x_b} f(y_1, \dots, y_n, y_1', \dots, y_n', x)\, dx \\ \widetilde{y}_i(x) &= y_i(x) + \varepsilon \eta_i(x, y_1, \dots, y_n, y_1', \dots, y_n') \\ \bar{x} &= x + \varepsilon \xi(x, y_1, \dots, y_n, y_1', \dots, y_n'). \end{aligned}$$

Before stating and proving Noether's theorem, let us look at a special case that illustrates the ideas involved. Recall that the conserved quantity defined in equation (14) was derived under the assumption that $f(y, y', x)$ (or $f(x, \dot{x}, t)$ in our earlier notation) satisfied the Euler-Lagrange equation had no *explicit* dependence on the independent variable. We will now show that we can obtain this same result as a consequence of the fact that *if* $f(y, y', x) = f(y, y')$, then the integral $I = \int f\,dx$ is invariant under the particular transformation

$$\widetilde{y} = y(x) + \varepsilon\eta \qquad \text{and} \qquad \bar{x} = x + \varepsilon\xi \tag{43}$$

with

$$\eta = 0 \quad \text{and} \quad \xi = 1. \tag{44}$$

By invariant we mean that $\bar{I}(\varepsilon) = \bar{I}(\varepsilon = 0) = I$.

To see this, we start from

$$I = \int_{x_a}^{x_b} f(y(x), y'(x))\,dx$$

and form the integral

$$\bar{I}(\varepsilon) = \int_{\bar{x}_a}^{\bar{x}_b} f(\widetilde{y}(\bar{x}), \widetilde{y}'(\bar{x}))\,d\bar{x}.$$

Just as we did above, we now write (using $\eta = 0$ and $\xi = 1$)

$$\widetilde{y}(\bar{x}) = y(\bar{x} - \varepsilon)$$

with

$$\bar{x}_a = x_a + \varepsilon \qquad \text{and} \qquad \bar{x}_b = x_b + \varepsilon.$$

Then we have

$$\begin{aligned} \bar{I}(\varepsilon) &= \int_{x_a+\varepsilon}^{x_b+\varepsilon} f\left[y(\bar{x} - \varepsilon), \frac{dy(\bar{x} - \varepsilon)}{d\bar{x}}\right] d\bar{x} \\ &= \int_{x_a+\varepsilon}^{x_b+\varepsilon} f\left[y(\bar{x} - \varepsilon), \frac{dy(\bar{x} - \varepsilon)}{d(\bar{x} - \varepsilon)}\right] d\bar{x} \end{aligned}$$

where we used the fact that

$$\frac{dy(\bar{x} - \varepsilon)}{d\bar{x}} = \frac{dy(\bar{x} - \varepsilon)}{d(\bar{x} - \varepsilon)} \frac{d(\bar{x} - \varepsilon)}{d\bar{x}} = \frac{dy(\bar{x} - \varepsilon)}{d(\bar{x} - \varepsilon)}.$$

Changing variables to $x = \bar{x} - \varepsilon$ yields

$$\bar{I}(\varepsilon) = \int_{x_a}^{x_b} f(y(x), y'(x))\,dx = \bar{I}(0) = I$$

which shows that I is indeed invariant under the given transformation (43) and (44).

To see that this again leads to equation (14), first suppose that we have the general case where the functions $y_i(x)$ that we are varying are solutions to the Euler-Lagrange equations, i.e., that $f(y_i, y_i', x)$ satisfies the Euler-Lagrange equations. Then the integral in equation (42) vanishes and we are left with

$$\bar{I}(\varepsilon) - \bar{I}(0) = \varepsilon\left[\sum_{i=1}^{n} \eta_i \frac{\partial f}{\partial y_i'} + \xi\left(f - \sum_{i=1}^{n} y_i' \frac{\partial f}{\partial y_i'}\right)\right]_{x_a}^{x_b}. \tag{45}$$

But we just showed that under the transformation (43) the left hand side of this equation vanishes (in the present case there is no sum on the right since we are dealing with only one dependent variable $y(x)$), and hence using equation (44) we are left with

$$0 = \left(f - y'\frac{\partial f}{\partial y'}\right)\bigg|_{x_a}^{x_b} = \left(f - y'\frac{\partial f}{\partial y'}\right)\bigg|_{x_b} - \left(f - y'\frac{\partial f}{\partial y'}\right)\bigg|_{x_a}$$

and therefore, since x_a and x_b were arbitrary we have

$$f - y'\frac{\partial f}{\partial y'} = \text{const} \tag{46}$$

which is just equation (14) as we wanted to show.

It is worth pointing out that if f is the Lagrangian $L = T - V$ of a system of particles, then the independent variable is the time t so (as we also mentioned after equation (14)) the conserved quantity in this case is the Hamiltonian $H = \dot{x}(\partial L/\partial \dot{x}) - L := p\dot{x} - L$. And if (as is usually the case) H represents the total energy $H = T + V$ of the system, then we have shown that *invariance of the Lagrangian under time translation leads to the conservation of energy.*

In summary, we have shown that the integral $I = \int f\,dx$ is invariant under the particular transformation (43) and (44), and as a consequence we have the conserved quantity defined by equation (46). We now state the generalization of this result, called **Noether's theorem**.

Theorem 3. *Suppose the integral*

$$I = \int_{x_a}^{x_b} f(y_1, \dots, y_n, y_1', \dots, y_n', x)\,dx$$

is invariant under the transformation

$$\widetilde{y}_i(x) = y_i(x) + \varepsilon\eta_i(x, y_1, \dots, y_n, y_1', \dots, y_n') \tag{47}$$

$$\bar{x} = x + \varepsilon\xi(x, y_1, \dots, y_n, y_1', \dots, y_n') \tag{48}$$

by which we mean

$$\bar{I}(\varepsilon) = \int_{x_a}^{x_b} f(\widetilde{y}_1(\bar{x}), \dots, \widetilde{y}_n(\bar{x}), \widetilde{y}_1'(\bar{x}), \dots, \widetilde{y}_n'(\bar{x}), \bar{x})\,d\bar{x} = I.$$

Then there exists a first integral of the related Euler-Lagrange equations which is given by

$$\sum_{i=1}^{n} \eta_i \frac{\partial f}{\partial y_i'} + \xi\left(f - \sum_{i=1}^{n} y_i' \frac{\partial f}{\partial y_i'}\right) = \text{const.} \tag{49}$$

Proof. By hypothesis, the functions $y_i(x)$ are solutions to the Euler-Lagrange equations, and hence equation (45) applies:

$$\bar{I}(\varepsilon) - I = \varepsilon\left[\sum_{i=1}^{n} \eta_i \frac{\partial f}{\partial y_i'} + \xi\left(f - \sum_{i=1}^{n} y_i' \frac{\partial f}{\partial y_i'}\right)\right]_{x_a}^{x_b}.$$

But by the assumed invariance of I, the left hand side of this equation vanishes so that

$$\left[\sum_{i=1}^{n} \eta_i \frac{\partial f}{\partial y_i'} + \xi\left(f - \sum_{i=1}^{n} y_i' \frac{\partial f}{\partial y_i'}\right)\right]_{x_a} = \left[\sum_{i=1}^{n} \eta_i \frac{\partial f}{\partial y_i'} + \xi\left(f - \sum_{i=1}^{n} y_i' \frac{\partial f}{\partial y_i'}\right)\right]_{x_b}.$$

Since x_a and x_b are arbitrary it follows that

$$\sum_{i=1}^{n} \eta_i \frac{\partial f}{\partial y_i'} + \xi\left(f - \sum_{i=1}^{n} y_i' \frac{\partial f}{\partial y_i'}\right) = \text{const.} \qquad \blacksquare$$

The term "first integral" is justified because, as we explained in our derivation of equation (14), the expression (49) is a first-order equation, in contrast to the second-order Euler-Lagrange equation.

Also observe that in the case where $\xi = 0$ (which corresponds to leaving the independent variable alone), we obtain exactly the same result as we did at the beginning of this section. In other words, leaving the *independent* variable alone means the invariance of I is the same as the invariance of f.

Example 15. Consider a two-particle system where the potential energy depends only on the vector joining the particles. Then the Lagrangian of this system is given by

$$L = \frac{1}{2}m_1(\dot{x}_1^2 + \dot{y}_1^2 + \dot{z}_1^2) + \frac{1}{2}m_2(\dot{x}_2^2 + \dot{y}_2^2 + \dot{z}_2^2) - V(\mathbf{r}_1 - \mathbf{r}_2).$$

This Lagrangian is clearly invariant under the transformation (where $i = 1, 2$)

$$\bar{t} = t + \varepsilon\tau \qquad \widetilde{x}_i = x_i + \varepsilon\xi_i \qquad \widetilde{y}_i = y_i + \varepsilon\eta_i \qquad \widetilde{z}_i = z_i + \varepsilon\zeta_i$$

with $\tau = \eta_i = \zeta_i = 0$ and $\xi_i = 1$. In other words, L is invariant under spatial translation in the x-direction.

Applying equation (49) we easily obtain (with the slight change in notation)

$$\sum_{i=1}^{2} \xi_i \frac{\partial L}{\partial \dot{x}_i} = \frac{\partial L}{\partial \dot{x}_1} + \frac{\partial L}{\partial \dot{x}_2} = m_1\dot{x}_1 + m_2\dot{x}_2 = \text{const}$$

which is just the statement of conservation of linear momentum in the x-direction. This result obviously applies to any one of the three directions, so we see that translational invariance of the Lagrangian leads to conservation of linear momentum.

Example 16. Now consider a particle moving in a central potential $V(r)$ where $r = (x^2 + y^2 + z^2)^{1/2}$. Then the Lagrangian is given by

$$L = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - V(r)$$

and it is easy to see that this is rotationally invariant. Indeed, under a rotation we have

$$x_i \to \widetilde{x}_i = \sum_{j=1}^{3} a_{ij} x_j$$

where the orthogonal rotation matrix $A = (a_{ij})$ satisfies $A^T A = I$ so that $\widetilde{r} = \sum_{i=1}^{3} (\widetilde{x}_i)^2 = \sum_{i=1}^{3} (x_i)^2 = r$. This shows that the potential energy is invariant under rotations. And since (a_{ij}) is a *constant* matrix, it is also clear that

$$\dot{\widetilde{x}}_i = \sum_{j=1}^{3} a_{ij} \dot{x}_j$$

so that $\sum_{i=1}^{3} (\dot{\widetilde{x}}_i)^2 = \sum_{i=1}^{3} (\dot{x}_i)^2$ and kinetic energy is also invariant under rotations.

For definiteness, consider a rotation by the angle ε about the z-axis. This corresponds to the transformation

$$\bar{t} = t \qquad \widetilde{x} = x\cos\varepsilon + y\sin\varepsilon \qquad \widetilde{y} = -x\sin\varepsilon + y\cos\varepsilon \qquad \widetilde{z} = z.$$

Since we are considering infinitesimal transformations $\varepsilon \ll 1$, these become

$$\bar{t} = t \qquad \widetilde{x} = x + \varepsilon y \qquad \widetilde{y} = y - \varepsilon x \qquad \widetilde{z} = z$$

so using the same notation as in the previous example, we have $\tau = \zeta = 0, \xi = y$ and $\eta = -x$. Plugging these values into equation (49) we obtain

$$\xi \frac{\partial L}{\partial \dot{x}} + \eta \frac{\partial L}{\partial \dot{y}} = my\dot{x} - mx\dot{y} = \text{const}$$

which is just the statement that the z-component of angular momentum $L_z = (\mathbf{r} \times \mathbf{p})_z = xp_y - yp_x$ is a constant.

You can repeat this calculation for rotations about the x- and y-axes to see that L_x and L_y are also conserved. In other words, the invariance of the Lagrangian under rotations leads to the conservation of angular momentum.

Summarizing the physics of what we have shown, the invariance (symmetry) of the Lagrangian under translations in time, spatial translations and rotations has lead to the conservation of energy, linear momentum and angular momentum respectively. These are specific examples of how Hamilton's principle

$$\delta S = \delta \int L\, dt = 0$$

allows us to find conservation laws as a consequence of the invariance of the action under various symmetry transformations.

10 Noether's Theorem in Field Theory

We now turn our attention to formulating Noether's theorem in 4-dimensional spacetime. There are many ways to approach this result, and almost every text follows a different method with different hypotheses, but they all seem to arrive at essentially the same result. Because of this, we will prove the theorem from more than one point of view so that when you run across it you should be able to understand whichever approach that author is following.

Let us first consider the variation of a scalar field $\phi = \phi(x)$ where points in spacetime are labeled by the point x. In other words, x stands for all four variables $x^\mu = \{x^0, x^1, x^2, x^3\} = \{t, x, y, z\}$. We will show that if a system is described by the Lagrangian

$$L = \int d^3x\, \mathscr{L}(\phi(x), \partial_\mu \phi(x))$$

with the equation of motion

$$\frac{\partial \mathscr{L}}{\partial \varphi} - \partial_\mu \left(\frac{\partial \mathscr{L}}{\partial(\partial_\mu \varphi)} \right) = 0$$

then any *continuous* symmetry transformation that leaves the action

$$S = \int L\, dt = \int \mathscr{L}\, d^4x$$

invariant implies the existence of a conserved current

$$\partial_\mu j^\mu(x) = 0$$

with a "charge" defined by

$$Q(t) = \int d^3x\, j^0(x)$$

which is a constant of the motion

$$\frac{dQ}{dt} = 0.$$

To begin with, we consider two types of variation. In the first case we vary the field at a particular spacetime point:

$$\phi(x) \to \phi'(x) = \phi(x) + \delta\phi(x) \quad \text{or} \quad \delta\phi(x) = \phi'(x) - \phi(x). \tag{50}$$

The second case (which we call the "total variation") varies both the field and the point at which it is evauated:

$$\phi(x) \to \phi'(x') = \phi(x) + \Delta\phi(x) \quad \text{or} \quad \Delta\phi(x) = \phi'(x') - \phi(x). \tag{51}$$

Of course we also have

$$x \to x' = x + \delta x \quad \text{or} \quad x'^\mu = x^\mu + \delta x^\mu. \tag{52}$$

It is easy (but important) to note that

$$\delta\partial_\mu\phi := \partial_\mu\phi'(x) - \partial_\mu\phi(x) = \partial_\mu(\phi'(x) - \phi(x)) = \partial_\mu\delta\phi.$$

In particular, equation (50) yields (note the arguments are all x and not x')

$$\partial_\mu\delta\phi(x) = \delta\partial_\mu\phi(x) = \partial_\mu\phi'(x) - \partial_\mu\phi(x). \tag{53}$$

It is trickier to find the analogous result for equation (51). First note that (52) implies

$$\frac{\partial x'^\mu}{\partial x^\alpha} = \delta^\mu_\alpha + \partial_\alpha\delta x^\mu \tag{54}$$

and hence the inverse matrix is given by

$$\frac{\partial x^\alpha}{\partial x'^\mu} = \delta^\alpha_\mu - \partial_\mu\delta x^\alpha \tag{55}$$

because to first order

$$\frac{\partial x'^\mu}{\partial x^\alpha}\frac{\partial x^\alpha}{\partial x'^\nu} = (\delta^\mu_\alpha + \partial_\alpha\delta x^\mu)(\delta^\alpha_\nu - \partial_\nu\delta x^\alpha) = \delta^\mu_\nu.$$

Using the notation $\phi_{,\mu}(x) := \partial_\mu\phi(x)$ we have

$$\begin{aligned}
\Delta\phi_{,\mu} := \Delta\partial_\mu\phi(x) &:= \frac{\partial\phi'(x')}{\partial x'^\mu} - \frac{\partial\phi(x)}{\partial x^\mu} \\
&= \frac{\partial x^\alpha}{\partial x'^\mu}\frac{\partial\phi'(x+\delta x)}{\partial x^\alpha} - \frac{\partial\phi(x)}{\partial x^\mu} \quad \text{(by the chain rule and (52))} \\
&= \frac{\partial x^\alpha}{\partial x'^\mu}\partial_\alpha[\phi(x+\delta x) + \delta\phi(x+\delta x)] - \partial_\mu\phi(x) \quad \text{(by (50))} \\
&= \frac{\partial x^\alpha}{\partial x'^\mu}\partial_\alpha[\phi(x) + \partial_\nu\phi(x)\delta x^\nu + \delta\phi(x)] - \partial_\mu\phi(x) \quad \text{(expand to 1}^{\text{st}}\text{ order)} \\
&= (\delta^\alpha_\mu - \partial_\mu\delta x^\alpha)[\partial_\alpha\phi(x) + \partial_\alpha\phi_{,\nu}(x)\delta x^\nu + \phi_{,\nu}(x)\partial_\alpha\delta x^\nu + \partial_\alpha\delta\phi(x)] - \partial_\mu\phi(x) \\
&\qquad\qquad \text{(by (55) and acting thru with } \partial_\alpha\text{)} \\
&= \partial_\mu\phi + \partial_\mu\phi_{,\nu}\delta x^\nu + \phi_{,\nu}\partial_\mu\delta x^\nu + \partial_\mu\delta\phi - \partial_\alpha\phi\,\partial_\mu\delta x^\alpha - \partial_\mu\phi. \quad \text{(to 1}^{\text{st}}\text{ order)}
\end{aligned}$$

Cancelling terms we are left with our desired result

$$\Delta\phi_{,\mu}(x) = \partial_\mu\delta\phi(x) + \partial_\mu\phi_{,\nu}\delta x^\nu. \tag{56}$$

It will also be useful to write (51) as follows:

$$\begin{aligned}\Delta\phi(x) &= \phi'(x') - \phi(x) = \phi'(x+\delta x) - \phi(x) = \phi'(x) + \partial_\mu\phi'(x)\delta x^\mu - \phi(x) \\ &= \delta\phi(x) + \partial_\mu\phi'(x)\delta x^\mu = \delta\phi(x) + \partial_\mu[\phi(x) + \delta\phi(x)]\delta x^\mu\end{aligned}$$

or (to first order)

$$\Delta\phi(x) = \delta\phi(x) + \partial_\mu\phi(x)\delta x^\mu. \tag{57}$$

The total variation applied to the Lagrangian density $\mathscr{L} = \mathscr{L}(\phi(x), \partial_\mu\phi(x))$ yields $\mathscr{L}' = \mathscr{L}(\phi'(x'), \partial'_\mu\phi'(x'))$ where $\partial'_\mu\phi'(x') = \partial\phi'(x')/\partial x'^\mu$. Our basic hypothesis is that $\mathscr{L}$ remains invariant under this total variation, i.e. $\mathscr{L}' = \mathscr{L}$, and that the action also remains invariant, i.e. $\int d^4x'\,\mathscr{L}' = \int d^4x\,\mathscr{L}$. But the standard change of variables formula says $d^4x' = |\partial x'/\partial x|d^4x$, so if $\mathscr{L}' = \mathscr{L}$, then we must also have the Jacobian $|\partial x'/\partial x| = 1$. In other words,

$$\left|\frac{\partial x'}{\partial x}\right| = \det\left(\frac{\partial x'^\mu}{\partial x^\alpha}\right) = \det(\delta^\mu_\alpha + \partial_\alpha\delta x^\mu) = 1.$$

(Of course, if we are dealing with a coordinate transformation that is simply a Lorentz transformation (which is the most common situation), then $x'^\mu = \Lambda^\mu{}_\nu x^\nu$ where $\Lambda^T g\Lambda = g$ and we automatically have $|\partial x'/\partial x| = |\det\Lambda| = 1$. This also follows from the fact that the invariant volume element is $\sqrt{|\det g|}d^4x$ where $g = \operatorname{diag}(1,-1,-1,-1)$.)

In any case, I claim this implies that

$$\operatorname{tr}(\partial_\alpha\delta x^\mu) = \partial_\mu\delta x^\mu = 0. \tag{58}$$

To see this, simply note from the definition of determinant, we have for any matrix $A = (a_{ij})$

$$\begin{aligned}\det(I+A) &= \varepsilon^{i_1\cdots i_n}(\delta_{1i_1} + a_{1i_1})(\delta_{2i_2} + a_{2i_2})\cdots(\delta_{ni_n} + a_{ni_n}) \\ &= \varepsilon^{i_1\cdots i_n}(\delta_{1i_1}\delta_{2i_2}\cdots\delta_{ni_n} + a_{1i_1}\delta_{2i_2}\cdots\delta_{ni_n} + \delta_{1i_1}a_{2i_2}\cdots\delta_{ni_n} \\ &\qquad + \cdots + \delta_{1i_1}\cdots\delta_{n-1i_{n-1}}a_{ni_n} + \text{terms of higher order in } a_{ij} \\ &\qquad + a_{1i_1}a_{2i_2}\cdots a_{ni_n}) \\ &= \varepsilon^{i_1\cdots i_n}\delta_{1i_1}\delta_{2i_2}\cdots\delta_{ni_n} + \varepsilon^{i_1\cdots i_n}a_{1i_1}\delta_{2i_2}\cdots\delta_{ni_n} \\ &\qquad + \varepsilon^{i_1\cdots i_n}\delta_{1i_1}a_{2i_2}\cdots\delta_{ni_n} + \cdots + \varepsilon^{i_1\cdots i_n}\delta_{n-1i_{n-1}}a_{ni_n} \\ &\qquad + \text{terms of higher order in } a_{ij} + \varepsilon^{i_1\cdots i_n}a_{1i_1}a_{2i_2}\cdots a_{ni_n} \\ &= \det I + \varepsilon^{i_1 2\cdots n}a_{1i_1} + \varepsilon^{1i_2\cdots n}a_{2i_2} + \cdots + \varepsilon^{1\cdots n-1i_n}a_{ni_n} \\ &\qquad + \text{higher order terms} + \det A \\ &= \det I + a_{11} + a_{22} + \cdots + a_{nn} + \text{higher order terms} + \det A \\ &= 1 + \operatorname{tr}A + \cdots + \det A.\end{aligned}$$

Thus if $|a_{ij}| \ll 1$, to first order we have $\det(I + A) \approx 1 + \operatorname{tr} A$, and hence $\det(I + A) = 1$ implies $\operatorname{tr} A = 0$ as claimed.

Before proving Noether's theorem, let us briefly recall the usual derivation of the Euler-Lagrange equation (33). In this case, we consider only variations of the *fields* at a particular point x as in equation (50), not the spacetime points x themselves. Then

$$\begin{aligned}
\delta\mathscr{L} &= \mathscr{L}(\phi'(x), \partial_\mu\phi'(x)) - \mathscr{L}(\phi(x), \partial_\mu\phi(x)) \\
&= \frac{\partial\mathscr{L}}{\partial\phi}\delta\phi + \frac{\partial\mathscr{L}}{\partial\phi_{,\mu}}\delta\phi_{,\mu} \\
&= \frac{\partial\mathscr{L}}{\partial\phi}\delta\phi + \frac{\partial\mathscr{L}}{\partial\phi_{,\mu}}\partial_\mu\delta\phi \\
&= \left[\frac{\partial\mathscr{L}}{\partial\phi} - \partial_\mu\left(\frac{\partial\mathscr{L}}{\partial\phi_{,\mu}}\right)\right]\delta\phi + \partial_\mu\left(\frac{\partial\mathscr{L}}{\partial\phi_{,\mu}}\delta\phi\right). \qquad (59)
\end{aligned}$$

By hypothesis we have $\delta S = \delta\int d^4x\,\mathscr{L} = \int d^4x\,\delta\mathscr{L} = 0$ where $\delta\phi = 0$ on the bounding spacetime surface. Then the last term in (59) doesn't contribute to the integral and we are left with

$$\int d^4x\left[\frac{\partial\mathscr{L}}{\partial\phi} - \partial_\mu\left(\frac{\partial\mathscr{L}}{\partial\phi_{,\mu}}\right)\right]\delta\phi = 0.$$

Since $\delta\phi$ was arbitrary, we conclude that

$$\frac{\partial\mathscr{L}}{\partial\phi} - \partial_\mu\left(\frac{\partial\mathscr{L}}{\partial\phi_{,\mu}}\right) = 0. \qquad (60)$$

Furthermore, it is easy to see that if we have multiple fields labeled by ϕ^r, then each field satisfies this equation also, i.e.,

$$\frac{\partial\mathscr{L}}{\partial\phi^r} - \partial_\mu\left(\frac{\partial\mathscr{L}}{\partial\phi^r_{,\mu}}\right) = 0. \qquad (61)$$

Example 17. Consider the electromagnetic four-potential $A^\mu = (\varphi, \mathbf{A})$ along with the electromagnetic field tensor $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu = -F_{\nu\mu}$ and the four-current $J^\mu = (\rho, \mathbf{J})$. Define the Lagrangian density

$$\begin{aligned}
\mathscr{L} &= -\frac{1}{4}F^2 - J\cdot A \\
&= -\frac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu) - J_\mu A^\mu.
\end{aligned}$$

Here the fields A^μ correspond to the ϕ^r.

Applying equation (61) we first easily find

$$\frac{\partial\mathscr{L}}{\partial A^\mu} = -J_\mu.$$

Now for the term

$$\partial_\nu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\nu A^\mu)}\right)=\partial^\nu\left(\frac{\partial\mathscr{L}}{\partial(\partial^\nu A^\mu)}\right).$$

We just have to be a little bit careful and realize that the indices in $\mathscr{L}$ are dummy indices. Noting we can write $\partial_\alpha A_\beta = g_{\alpha\rho}g_{\beta\lambda}\partial^\rho A^\lambda$ and so forth, we have

$$-\frac{1}{4}F^2 = -\frac{1}{4}g_{\alpha\rho}g_{\beta\lambda}(\partial^\rho A^\lambda - \partial^\lambda A^\rho)(\partial^\alpha A^\beta - \partial^\beta A^\alpha)$$

and therefore

$$\begin{aligned}
-\frac{1}{4}\frac{\partial F^2}{\partial(\partial^\nu A^\mu)} &= -\frac{1}{4}g_{\alpha\rho}g_{\beta\lambda}\big[(\delta^\rho_\nu\delta^\lambda_\mu - \delta^\lambda_\nu\delta^\rho_\mu)(\partial^\alpha A^\beta - \partial^\beta A^\alpha) \\
&\qquad\qquad + (\partial^\rho A^\lambda - \partial^\lambda A^\rho)(\delta^\alpha_\nu\delta^\beta_\mu - \delta^\beta_\nu\delta^\alpha_\mu)\big] \\
&= -\frac{1}{4}\big[(\delta^\rho_\nu\delta^\lambda_\mu - \delta^\lambda_\nu\delta^\rho_\mu)(\partial_\rho A_\lambda - \partial_\lambda A_\rho) \\
&\qquad\qquad + (\partial_\alpha A_\beta - \partial_\beta A_\alpha)(\delta^\alpha_\nu\delta^\beta_\mu - \delta^\beta_\nu\delta^\alpha_\mu)\big] \\
&= -(\partial_\nu A_\mu - \partial_\mu A_\nu) = -F_{\nu\mu}.
\end{aligned}$$

Using these results, the Euler-Lagrange equations become simply $\partial^\nu F_{\nu\mu} = J_\mu$ or, equivalently

$$\partial_\mu F^{\mu\nu} = J^\nu \tag{62}$$

which you may recognize as two of Maxwell's equations.

To see this, first recall that we are using the metric $g = \operatorname{diag}(1,-1,-1,-1)$ so that $\partial/\partial t = \partial/\partial x^0 = \partial_0 = \partial^0$ and $\boldsymbol{\nabla}^i := \partial/\partial x^i = \partial_i = -\partial^i$. Using $\mathbf{E} = -\boldsymbol{\nabla}\varphi - \partial\mathbf{A}/\partial t$ we have

$$E^i = \partial^i A^0 - \partial^0 A^i = F^{i0}$$

and also $\mathbf{B} = \boldsymbol{\nabla}\times\mathbf{A}$ so that

$$B^1 = \boldsymbol{\nabla}^2 A^3 - \boldsymbol{\nabla}^3 A^2 = -\partial^2 A^3 + \partial^3 A^2 = -F^{23}$$

plus cyclic permutations $1\to 2\to 3\to 1$. Then the electromagnetic field tensor is given by

$$F^{\mu\nu} = \begin{bmatrix} 0 & -E^1 & -E^2 & -E^3 \\ E^1 & 0 & -B^3 & B^2 \\ E^2 & B^3 & 0 & -B^1 \\ E^3 & -B^2 & B^1 & 0 \end{bmatrix}.$$

(Be sure to note that this is the form of $F^{\mu\nu}$ for the metric $\operatorname{diag}(1,-1,-1,-1)$. If you use the metric $\operatorname{diag}(-1,1,1,1)$ then all entries of $F^{\mu\nu}$ change sign. In addition, you frequently see the matrix $F^\mu{}_\nu$ which also has different signs.)

For the $\nu = 0$ component of equation (62) we have $J^0 = \partial_\mu F^{\mu 0} = \partial_i F^{i0} = \partial_i E^i$ which is Coulomb's law

$$\boldsymbol{\nabla}\cdot\mathbf{E} = \rho.$$

Now consider the $\nu = 1$ component of equation (62). This is $J^1 = \partial_\mu F^{\mu 1} = \partial_0 F^{01} + \partial_2 F^{21} + \partial_3 F^{31} = -\partial_0 E^1 + \partial_2 B^3 - \partial_3 B^2 = -\partial_t E^1 + (\boldsymbol{\nabla} \times \mathbf{B})^1$ and therefore we have

$$\boldsymbol{\nabla} \times \mathbf{B} - \frac{\partial \mathbf{E}}{\partial t} = \mathbf{J}.$$

The last two Maxwell equations follow directly by taking the divergence of $\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A}$ and the curl of $\mathbf{E} = -\boldsymbol{\nabla}\phi - \partial \mathbf{A}/\partial t$, i.e.,

$$\boldsymbol{\nabla} \cdot \mathbf{B} = 0$$

and

$$\boldsymbol{\nabla} \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0.$$

I leave it as an exercise for you to show that these last two equations can be written as (note that the superscripts are cyclic permutations)

$$\partial^\mu F^{\nu\sigma} + \partial^\nu F^{\sigma\mu} + \partial^\sigma F^{\mu\nu} = 0$$

or simply

$$\partial^{[\mu} F^{\nu\sigma]} = 0.$$

Now for Noether's theorem. In this case we consider the total variation and assume that it is a symmetry of $\mathscr{L}$, i.e. $\mathscr{L}' = \mathscr{L}$. Recall that by definition we have $\Delta\phi_{,\mu} = \partial\phi'(x')/\partial x'^\mu - \partial\phi(x)/\partial x^\mu$ and therefore

$$\partial'_\mu \phi'(x') = \frac{\partial \phi'(x')}{\partial x'^\mu} = \partial_\mu \phi(x) + \Delta\phi_{,\mu}.$$

Then we have

$$\begin{aligned}
0 = \delta\mathscr{L} &= \mathscr{L}(\phi'(x'), \partial'_\mu \phi'(x')) - \mathscr{L}(\phi(x), \partial_\mu \phi(x)) \\
&= \frac{\partial \mathscr{L}}{\partial \phi} \Delta\phi + \frac{\partial \mathscr{L}}{\partial \phi_{,\mu}} \Delta\phi_{,\mu} \\
&= \frac{\partial \mathscr{L}}{\partial \phi} (\delta\phi + \partial_\alpha \phi \, \delta x^\alpha) + \frac{\partial \mathscr{L}}{\partial \phi_{,\mu}} (\partial_\mu \delta\phi + \partial_\mu \phi_{,\alpha} \delta x^\alpha) \quad \text{(by (56) and (57))} \\
&= \frac{\partial \mathscr{L}}{\partial \phi} \delta\phi + \frac{\partial \mathscr{L}}{\partial \phi} \partial_\alpha \phi \, \delta x^\alpha + \frac{\partial \mathscr{L}}{\partial \phi_{,\mu}} \partial_\mu \delta\phi + \frac{\partial \mathscr{L}}{\partial \phi_{,\mu}} \partial_\alpha \phi_{,\mu} \delta x^\alpha \\
&\qquad\qquad\qquad\qquad \text{(since } \partial_\mu \phi_{,\alpha} = \partial_\alpha \phi_{,\mu}\text{)} \\
&= \partial_\mu \left(\frac{\partial \mathscr{L}}{\partial \phi_{,\mu}} \right) \delta\phi + \frac{\partial \mathscr{L}}{\partial \phi_{,\mu}} \partial_\mu \delta\phi + \left(\frac{\partial \mathscr{L}}{\partial \phi} \frac{\partial \phi}{\partial x^\alpha} + \frac{\partial \mathscr{L}}{\partial \phi_{,\mu}} \frac{\partial \phi_{,\mu}}{\partial x^\alpha} \right) \delta x^\alpha \\
&\qquad\qquad\qquad\qquad \text{(by (60) and } \partial_\alpha = \partial/\partial x^\alpha\text{)} \\
&= \partial_\mu \left(\frac{\partial \mathscr{L}}{\partial \phi_{,\mu}} \delta\phi \right) + \frac{\partial \mathscr{L}}{\partial x^\alpha} \delta x^\alpha \quad \text{(by the product and chain rules)}
\end{aligned}$$

$$
\begin{aligned}
&= \partial_\mu \left[\frac{\partial \mathscr{L}}{\partial \phi_{,\mu}} (\Delta\phi - \phi_{,\alpha} \delta x^\alpha) \right] + g^\mu{}_\alpha \partial_\mu \mathscr{L}\, \delta x^\alpha \quad \text{(by (57))} \\
&= \partial_\mu \left[\frac{\partial \mathscr{L}}{\partial \phi_{,\mu}} \Delta\phi - \frac{\partial \mathscr{L}}{\partial \phi_{,\mu}} \phi_{,\alpha}\, \delta x^\alpha + g^\mu{}_\alpha \mathscr{L} \delta x^\alpha \right] \quad \text{(by (58))}\,.
\end{aligned}
$$

(Note that $g^\mu{}_\alpha = \delta^\mu_\alpha$.)

Let us define the **canonical energy-momentum tensor** $T^\mu{}_\alpha$ by

$$
T^\mu{}_\alpha = \frac{\partial \mathscr{L}}{\partial \phi_{,\mu}} \phi_{,\alpha} - g^\mu{}_\alpha \mathscr{L} \tag{63}
$$

and the **current** j^μ by

$$
j^\mu = \frac{\partial \mathscr{L}}{\partial \phi_{,\mu}} \Delta\phi - T^\mu{}_\alpha \delta x^\alpha. \tag{64}
$$

Then we have shown that

$$
\partial_\mu j^\mu = 0. \tag{65}
$$

In other words, the current j^μ obeys a continuity equation, and is also referred to as a "conserved current."

We now define the **charge** Q by

$$
Q := \int_{\text{all space}} d^3x\, j^0.
$$

This charge is conserved (i.e., constant in time) because $\partial_0 j^0 + \partial_i j^i = \partial_0 j^0 + \boldsymbol{\nabla} \cdot \vec{j} = 0$ so that

$$
\begin{aligned}
\frac{dQ}{dt} &= \frac{d}{dt} \int_{\text{all space}} d^3x\, j^0 = \int d^3x\, \partial_0 j^0 = -\int d^3x\, \boldsymbol{\nabla} \cdot \vec{j} \\
&= -\int_{\text{surface at } \infty} da\, \hat{n} \cdot \vec{j} \\
&= 0
\end{aligned}
$$

where the last integral vanishes because the fields are assumed to vanish at infinity. This then is the statement of Noether's theorem. In other words, if the Lagrangian is invariant under a symmetry transformation, then there exists a conserved charge Q.

Note that from (57) we may also write j^μ in the form

$$
\begin{aligned}
j^\mu &= \frac{\partial \mathscr{L}}{\partial \phi_{,\mu}} (\delta\phi + \phi_{,\alpha} \delta x^\alpha) - \frac{\partial \mathscr{L}}{\partial \phi_{,\mu}} \phi_{,\alpha} \delta x^\alpha + \mathscr{L} \delta x^\mu \\
&= \frac{\partial \mathscr{L}}{\partial \phi_{,\mu}} \delta\phi + \mathscr{L} \delta x^\mu.
\end{aligned}
$$

In addition, if the fields have several components ϕ_r, then these would also be summed over. For example, (using the summation convention on the index r also)

$$T^{\mu}{}_{\alpha} = \frac{\partial \mathscr{L}}{\partial \phi_{r,\mu}} \phi_{r,\alpha} - g^{\mu}{}_{\alpha} \mathscr{L}. \tag{66}$$

and

$$j^{\mu} = \frac{\partial \mathscr{L}}{\partial \phi_{r,\mu}} \Delta \phi_r - T^{\mu}{}_{\alpha} \delta x^{\alpha}. \tag{67}$$

Example 18. Let us first take a look at the simple case of translational invariance, and derive the energy-momentum tensor directly. Under the translation

$$x^{\mu} \to x'^{\mu} = x^{\mu} + \varepsilon^{\mu}$$

(where ε is a constant) we can say in general that the Lagrangian will change by an amount

$$\delta \mathscr{L} = \mathscr{L}' - \mathscr{L} = (\partial_{\mu} \mathscr{L}) \varepsilon^{\mu} = \partial_{\mu} (g^{\mu}_{\nu} \mathscr{L}) \varepsilon^{\nu}.$$

On the other hand however, if $\mathscr{L}$ is translationally invariant, then it can have no *explicit* dependence on x^{μ} so that $\mathscr{L} = \mathscr{L}(\varphi_r, \partial \varphi_r / \partial x^{\mu})$ and we can write

$$\delta \mathscr{L} = \frac{\partial \mathscr{L}}{\partial \varphi_r} \delta \varphi_r + \frac{\partial \mathscr{L}}{\partial \varphi_{r,\mu}} \delta \varphi_{r,\mu}$$

where

$$\delta \varphi_r = \varphi_r(x + \varepsilon) - \varphi_r(x) = \varepsilon^{\nu} \partial_{\nu} \varphi_r(x) = \varepsilon^{\nu} \varphi_{r,\nu}$$

and

$$\delta \varphi_{r,\mu} = \delta \partial_{\mu} \varphi_r = \partial_{\mu} \delta \varphi_r = \partial_{\mu} (\partial_{\nu} \varphi_r) \varepsilon^{\nu} = \partial_{\mu} \varphi_{r,\nu} \varepsilon^{\nu}.$$

(Remember that we sum over the index r, but its placement as a superscript or subscript doesn't make any difference.) Using equation (61) to replace the first term in $\delta \mathscr{L}$ we then find

$$\begin{aligned} \delta \mathscr{L} &= \partial_{\mu} \left(\frac{\partial \mathscr{L}}{\partial \varphi_{r,\mu}} \right) \varphi_{r,\nu} \varepsilon^{\nu} + \frac{\partial \mathscr{L}}{\partial \varphi_{r,\mu}} \partial_{\mu} \varphi_{r,\nu} \varepsilon^{\nu} \\ &= \partial_{\mu} \left(\frac{\partial \mathscr{L}}{\partial \varphi_{r,\mu}} \varphi_{r,\nu} \right) \varepsilon^{\nu} \end{aligned}$$

Equating the general form for the variation of $\mathscr{L}$ with the specific form that we just derived for translational invariance, and using the fact that ε^{ν} is arbitrary, we obtain

$$\partial_{\mu} \left(\frac{\partial \mathscr{L}}{\partial \varphi_{r,\mu}} \varphi_{r,\nu} - g^{\mu}_{\nu} \mathscr{L} \right) = 0.$$

Defining the canonical energy-momentum tensor

$$T^{\mu}{}_{\nu} = \frac{\partial \mathscr{L}}{\partial \varphi_{r,\mu}} \varphi_{r,\nu} - g^{\mu}_{\nu} \mathscr{L}$$

we have the conservation equation

$$\partial_\mu T^\mu{}_\nu = 0$$

which can also be written as

$$\partial_\mu T^{\mu\nu} = 0.$$

Note also that in general $T^{\mu\nu}$ is not symmetric, i.e., $T^{\mu\nu} \neq T^{\nu\mu}$. We will have some additional comments on this below.

To interpret $T^{\mu\nu}$, recall from particle mechanics that the momentum is defined by $p = \partial L/\partial \dot{q}$. For fields, we similarly define the **canonical momentum** by

$$\pi_r(x) := \frac{\partial \mathscr{L}}{\partial \dot{\varphi}_r}.$$

The particle mechanics Hamiltonian is defined by $H = \sum p_i \dot{q}_i - L$, and now we have the **Hamiltonian density**

$$\mathscr{H} = \pi\dot{\varphi} - \mathscr{L}$$

so that

$$H = \int d^3x\, \mathscr{H}.$$

Let us define the 4-vector

$$P^\nu = \int d^3x\, T^{0\nu} = \int d^3x\, [\pi_r(x)\partial^\nu \varphi_r(x) - g^{0\nu}\mathscr{L}].$$

This is a conserved quantity because from $\partial_\mu T^{\mu\nu} = \partial_0 T^{0\nu} + \partial_i T^{i\nu} = 0$ we have

$$\frac{dP^\nu}{dt} = \partial_0 P^\nu = \int d^3x\, \partial_0 T^{0\nu} = -\int d^3x\, \partial_i T^{i\nu}$$

and changing to a surface integral at infinity this must vanish (since the fields are assumed to vanish at infinity). Now observe that

$$T^{00} = \pi\dot{\varphi} - \mathscr{L} = \mathscr{H}$$

is the Hamiltonian density, and the Hamiltonian is then the spatial integral of this:

$$H = \int d^3x\, \mathscr{H} = \int d^3x\, T^{00} = P^0.$$

In other words, P^0 is just the (conserved) total energy. Since the zeroth component of a 4-vector is the energy, it follows that P^ν is in fact the energy-momentum 4-vector for the field. This is the justification for calling $T^{\mu\nu}$ the energy-momentum tensor.

Example 19. Recall from Example 13 that the Lagrangian for the Klein-Gordon equation is

$$\mathscr{L} = \frac{1}{2}\big[(\partial_\alpha\varphi)(\partial^\alpha\varphi) - m^2\varphi^2\big].$$

The scalar field $\varphi(x)$ represents a spinless particle of mass m. By 'scalar field' we mean that it is invariant under a **Poincaré transformation**

$$x \to x'^\mu = \Lambda^\mu{}_\nu x^\nu + a^\mu.$$

In other words, $\varphi(x) \to \varphi'(x') = \varphi(x)$ so that $\Delta\varphi(x) = 0$.

In particular, let us first consider an infinitesimal translation $x \to x'^\mu = x^\mu + \varepsilon^\mu$ so $\delta x^\mu = \varepsilon^\mu$. We have seen that (see Example 13)

$$\frac{\partial\mathscr{L}}{\partial\varphi_{,\mu}} = \partial^\mu\varphi$$

and hence

$$\begin{aligned} T^\mu{}_\nu &= \frac{\partial\mathscr{L}}{\partial(\partial_\mu\varphi)}\partial_\nu\varphi - g^\mu{}_\nu\mathscr{L} \\ &= (\partial^\mu\varphi)(\partial_\nu\varphi) - \frac{1}{2}\delta^\mu_\nu\big[(\partial_\alpha\varphi)(\partial^\alpha\varphi) - m^2\varphi^2\big] \end{aligned}$$

so the current becomes

$$j^\mu = -T^\mu{}_\nu\delta x^\nu = (\partial^\mu\varphi)(\partial_\nu\varphi)\varepsilon^\nu - \frac{1}{2}\big[(\partial_\alpha\varphi)(\partial^\alpha\varphi) - m^2\varphi^2\big]\varepsilon^\mu.$$

Note that the continuity equation (65) becomes

$$\partial_\mu j^\mu = -\partial_\mu T^\mu{}_\nu\varepsilon^\nu = 0.$$

But the displacements ε^ν are all independent, and hence we have

$$\partial_\mu T^{\mu\nu} = 0.$$

From the previous example, the canonical momentum density is given by

$$\pi(x) = \frac{\partial\mathscr{L}}{\partial\dot\varphi} = \frac{\partial\mathscr{L}}{\partial(\partial_0\varphi)} = \partial^0\varphi(x) = \dot\varphi(x).$$

Now consider an infinitesimal Lorentz transformation

$$x^\mu \to x'^\mu = \Lambda^\mu{}_\nu x^\nu = (g^\mu{}_\nu + \varepsilon^\mu{}_\nu)x^\nu = x^\mu + \varepsilon^\mu{}_\nu x^\nu.$$

For a Lorentz transformation we have $x'^\mu = \Lambda^\mu{}_\nu x^\nu$ along with $x'^\mu x'_\mu = x^\nu x_\nu$ and together these imply

$$\Lambda^\mu{}_\alpha\Lambda_{\mu\beta} = g_{\alpha\beta}. \tag{68}$$

This is equivalent to $(\Lambda^T)_\alpha{}^\mu g_{\mu\nu}\Lambda^\nu{}_\beta = g_{\alpha\beta}$ or simply

$$\Lambda^T g\Lambda = g.$$

(This is sometimes taken as the *definition* of a Lorentz transformation, i.e., it leaves the metric invariant.) Substituting the infinitesimal transformation $\Lambda^\mu{}_\nu = g^\mu{}_\nu + \varepsilon^\mu{}_\nu$ into (68) and expanding to first order yields $g_{\alpha\beta} = g_{\alpha\beta} + \varepsilon_{\alpha\beta} + \varepsilon_{\beta\alpha}$ and hence we have the important result that

$$\varepsilon_{\alpha\beta} = -\varepsilon_{\beta\alpha}. \tag{69}$$

So now we have $\delta x^\mu = \varepsilon^\mu{}_\nu x^\nu$ or, equivalently, $\delta x_\alpha = \varepsilon_{\alpha\beta}x^\beta$.

The current is given by

$$j^\mu = -T^{\mu\alpha}\delta x_\alpha = -T^{\mu\alpha}\varepsilon_{\alpha\beta}x^\beta$$

so using equation (69) we can antisymmetrize over α and β to write

$$j^\mu = -T^{\mu\alpha}x^\beta\varepsilon_{\alpha\beta} = \frac{1}{2}(-T^{\mu\alpha}x^\beta + T^{\mu\beta}x^\alpha)\varepsilon_{\alpha\beta} := \frac{1}{2}\varepsilon_{\alpha\beta}\mathscr{M}^{\mu\alpha\beta}$$

where we have defined

$$\mathscr{M}^{\mu\alpha\beta} = x^\alpha T^{\mu\beta} - x^\beta T^{\mu\alpha}.$$

Again, $\varepsilon_{\alpha\beta}$ is arbitrary so we are left with the continuity equation

$$\partial_\mu j^\mu = \partial_\mu \mathscr{M}^{\mu\alpha\beta} = 0.$$

The conserved 'charge' is now

$$M^{\alpha\beta} := \int d^3x\, j^0 = \int d^3x\, \mathscr{M}^{0\alpha\beta} = \int d^3x\,(x^\alpha T^{0\beta} - x^\beta T^{0\alpha})$$

where

$$M^{ij} = \int d^3x\,(x^i T^{0j} - x^j T^{0i})$$

represents the angular momentum of the field as the integral of an angular momentum density. (Remember that $P^i = \int d^3x\, T^{0i}$ is the ith component of the momentum of the field, so this is just like the classical expression $L_k = x_i p_j - x_j p_i$.)

Observe that the energy-momentum tensor for the Klein-Gordon field can be written as

$$T^{\mu\nu} = (\partial^\mu\varphi)(\partial^\nu\varphi) - g^{\mu\nu}\mathscr{L}$$

which is symmetric. While it is certainly not the case that $T^{\mu\nu}$ is symmetric in general, it is nonetheless possible to define a symmetric energy-momentum tensor closely related to $T^{\mu\nu}$. To see this, define

$$\Theta^{\mu\nu} = T^{\mu\nu} + \partial_\sigma f^{\sigma\mu\nu}$$

where $f^{\sigma\mu\nu}$ is any arbitrary function of the fields that is antisymmetric in its first two indices : $f^{\sigma\mu\nu} = -f^{\mu\sigma\nu}$. Then clearly

$$\partial_\mu \partial_\sigma f^{\sigma\mu\nu} = 0$$

so that we also have

$$\partial_\mu \Theta^{\mu\nu} = \partial_\mu T^{\mu\nu} = 0.$$

Furthermore, we see that

$$\begin{aligned}\int_R d^3x\, \Theta^{0\nu} &= \int_R d^3x\, (T^{0\nu} + \partial_\sigma f^{\sigma 0\nu}) = \int_R d^3x\, T^{0\nu} + \int_R d^3x\, \partial_i f^{i0\nu} \\ &= \int_R d^3x\, T^{0\nu} + \int_{\partial R} da_i f^{i0\nu} = \int_R d^3x\, T^{0\nu} \\ &= P^\nu\end{aligned}$$

where we used the fact that $f^{00\nu} = 0$, and we used the 4-dimensional divergence theorem to convert the volume integral to a surface integral where the fields are assumed to vanish. If we choose $f^{\sigma\mu\nu}$ in such a way as to make $\Theta^{\mu\nu}$ symmetric, then we see that while the energy-momentum tensor is not unique, we can still define a symmetric energy-momentum tensor that yields a unique 4-momentum, and thus the energy and momentum of the fields remain unique.

There are a number of reasons for wanting the energy-momentum tensor to be symmetric. One good reason is that this must be the case in order that the angular momentum of the field be conserved. For the Klein-Gordon field we have seen that the angular momentum density of the field is given by

$$\mathscr{M}^{\mu\alpha\beta} = x^\alpha T^{\mu\beta} - x^\beta T^{\mu\alpha}.$$

Then

$$\begin{aligned}\partial_\mu \mathscr{M}^{\mu\alpha\beta} &= \delta^\alpha_\mu T^{\mu\beta} + x^\alpha \partial_\mu T^{\mu\beta} - \delta^\beta_\mu T^{\mu\alpha} - x^\beta \partial_\mu T^{\mu\alpha} \\ &= T^{\alpha\beta} - T^{\beta\alpha}\end{aligned}$$

since $\partial_\mu T^{\mu\alpha} = 0$. But then $\partial_\mu \mathscr{M}^{\mu\alpha\beta} = 0$ implies that we must have $T^{\alpha\beta} = T^{\beta\alpha}$.

Another good reason comes from Einstein's equation of general relativity. This is

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = -\frac{8\pi G}{c^2} T_{\mu\nu}.$$

Since both the Ricci tensor $R_{\mu\nu}$ and the metric $g_{\mu\nu}$ are symmetric, we must also have a symmetric energy-momentum tensor.

So much for scalar fields and spinless particles. What about particles with spin? These are described by fields that have several components ϕ_r and definite transformation properties under Poincaré transformations. To motivate our next example, recall that a Poincaré transformation consists of a Lorentz transformation plus a displacement, and hence we may write a general variation in x as

$$x^\mu \to x'^\mu = x^\mu + \delta x^\mu = x^\mu + \varepsilon^\mu{}_\nu x^\nu + a^\mu \tag{70}$$

which represents an infinitesimal rotation $\varepsilon_{\mu\nu}$ (i.e., a Lorentz boost) plus a translation a^μ. To motivate a reasonable form for the transformation of the fields, recall (maybe, but the details here are unimportant anyway) from the theory of the Dirac equation that the operator corresponding to an infinitesimal Lorentz transformation $\Lambda^\mu{}_\nu = g^\mu{}_\nu + \varepsilon^\mu{}_\nu$ is given by

$$S(\Lambda) = I - \frac{i}{2}\varepsilon_{\mu\nu}\Sigma^{\mu\nu}$$

where $\Sigma^{\mu\nu} = (i/4)(\gamma^\mu\gamma^\nu - \gamma^\nu\gamma^\mu)$ is a 4×4 complex matrix. Hence in the infinitesimal case the Dirac spinors transform as (where $x' = \Lambda x$ and we use the summation convention on latin indices even though both are down)

$$\begin{aligned}\psi_r(x) \to \psi'_r(x') = S(\Lambda)_{rs}\psi_s(x) &= \left(1 - \frac{i}{2}\varepsilon_{\mu\nu}\Sigma^{\mu\nu}\right)_{rs}\psi_s(x) \\ &= \psi_r(x) - \frac{i}{2}\varepsilon_{\mu\nu}(\Sigma^{\mu\nu})_{rs}\psi_s(x).\end{aligned}$$

Example 20. Using the above discussion as a model, we assume that the transformation (70) induces a corresponding change in the fields of the form given by

$$\phi_r(x) \to \phi'_r(x') = \phi_r(x) + \Delta\phi_r(x) = \phi_r(x) + \frac{1}{2}\varepsilon_{\mu\nu}S^{\mu\nu}_{rs}\phi_s(x) \tag{71}$$

where $\varepsilon_{\mu\nu} = -\varepsilon_{\nu\mu}$.

We first consider a pure translation. This means $\varepsilon_{\mu\nu} = 0$ so equations (70) and (71) yield $\delta x^\mu = a^\mu$ and $\Delta\phi_r(x) = 0$. Then from (64) we have $j^\mu = -T^\mu{}_\alpha a^\alpha$ so that $\partial_\mu j^\mu = 0$ implies (since a^α is arbitrary)

$$\partial_\mu T^{\mu\alpha} = 0$$

where

$$T^{\mu\alpha} = \frac{\partial\mathscr{L}}{\partial\phi_{,\mu}}\frac{\partial\phi}{\partial x_\alpha} - g^{\mu\alpha}\mathscr{L}.$$

We now define

$$P^\alpha = \int d^3x\, T^{0\alpha} = \int d^3x\left(\frac{\partial\mathscr{L}}{\partial\phi_{,0}}\frac{\partial\phi}{\partial x_\alpha} - g^{0\alpha}\mathscr{L}\right)$$

where

$$\frac{\partial\mathscr{L}}{\partial\phi_{,0}} = \frac{\partial\mathscr{L}}{\partial\dot\phi} = \pi(x)$$

is the canonical momentum. Then

$$P^0 = \int d^3x\,(\pi\dot{\phi} - \mathscr{L}) = \int d^3x\,\mathscr{H} = H$$

which implies that T^{00} is the energy density, and

$$P^i = \int d^3x\,T^{0i} = \int d^3x\,\pi\,\partial^i\phi$$

which implies that T^{0i} is the momentum density. Note also that each P^α is conserved because applying the divergence theorem yields

$$\frac{dP^\alpha}{dt} = \int d^3x\,\partial_0 T^{0\alpha} = -\int d^3x\,\partial_i T^{i\alpha} = -\int dS_i\,T^{i\alpha}$$

and this also vanishes at infinity.

Now consider rotations. This means $a^\mu = 0$ so that $\delta x^\mu = \varepsilon^\mu{}_\nu x^\nu$ and $\Delta\phi_r(x) = \frac{1}{2}\varepsilon_{\mu\nu}S^{\mu\nu}_{rs}\phi_s(x)$. From (64) the current is given by

$$j^\mu = \frac{\partial\mathscr{L}}{\partial\phi_{,\mu}}\frac{1}{2}\varepsilon_{\alpha\beta}S^{\alpha\beta}_{rs}\phi_s - T^\mu{}_\alpha\varepsilon^\alpha{}_\beta x^\beta.$$

Using (69) we can write

$$T^{\mu\alpha}\varepsilon_{\alpha\beta}x^\beta = \frac{1}{2}(T^{\mu\alpha}\varepsilon_{\alpha\beta}x^\beta + T^{\mu\beta}\varepsilon_{\beta\alpha}x^\alpha) = \frac{1}{2}\varepsilon_{\alpha\beta}(x^\beta T^{\mu\alpha} - x^\alpha T^{\mu\beta})$$

and therefore

$$j^\mu = \frac{1}{2}\varepsilon_{\alpha\beta}\left[\frac{\partial\mathscr{L}}{\partial\phi_{r,\mu}}S^{\alpha\beta}_{rs}\phi_s + (x^\alpha T^{\mu\beta} - x^\beta T^{\mu\alpha})\right] := \frac{1}{2}\varepsilon_{\alpha\beta}\mathscr{M}^{\mu\alpha\beta}.$$

Define

$$M^{\alpha\beta} = \int d^3x\,\mathscr{M}^{0\alpha\beta} = \int d^3x\,[\pi_r S^{\alpha\beta}_{rs}\phi_s + (x^\alpha T^{0\beta} - x^\beta T^{0\alpha})]$$

and note

$$M^{ij} = \int d^3x\,[\pi_r S^{ij}_{rs}\phi_s + (x^i T^{0j} - x^j T^{0i})]$$

so that we interpret $x^i T^{0j} - x^j T^{0i}$ as the orbital angular momentum density and $\pi_r S^{ij}_{rs}\phi_s$ as the intrinsic spin angular momentum density of the field. Note this shows that the scalar field φ satisfying the Klein-Gordon equation indeed represents a spinless particle since its angular momentum tensor only contains an orbital part.

Now, the derivation of equation (65) (with equation (64)) specifically required that both the action *and* the Lagrangian density be invariant. However, using a slightly different approach we can arrive at the same result by requiring that the action alone be invariant.

So, with a somewhat different notation, let us consider the action

$$S = \int d^4x \, \mathscr{L}(\phi, \partial_\mu \phi)$$

and see what happens when we make the transformations

$$\begin{aligned} x &\to x' \\ \phi(x) &\to \phi'(x') := \mathcal{F}(\phi(x)). \end{aligned}$$

We now have a new action S' which we write out using the chain rule and Jacobi's formula for the change of variables:

$$\begin{aligned} S' &= \int d^4x' \, \mathscr{L}(\phi'(x'), \partial'_\mu \phi'(x')) \\ &= \int d^4x' \, \mathscr{L}(\mathcal{F}(\phi(x)), \partial'_\mu \mathcal{F}(\phi(x))) \\ &= \int d^4x \left| \frac{\partial x'}{\partial x} \right| \mathscr{L}(\mathcal{F}(\phi(x)), (\partial x^\nu / \partial x'^\mu) \partial_\nu \mathcal{F}(\phi(x))). \end{aligned} \tag{72}$$

Let us expand our transformations to first order in terms of a set of infinitesimal parameters ω_a as (using the summation convention on the index a over whatever range is necessary)

$$x'^\mu = x^\mu + \omega_a \frac{\delta x^\mu}{\delta \omega_a} := x^\mu + \omega_a X^\mu_a \tag{73a}$$

$$\phi'(x') = \mathcal{F}(\phi(x)) = \phi(x) + \omega_a \frac{\delta \mathcal{F}}{\delta \omega_a}(x) := \phi(x) + \omega_a \mathcal{F}_a(x). \tag{73b}$$

It is important to realize that so far we assume that the ω_a's can depend on x. As we also saw earlier, the Jacobian and its inverse may be written as

$$\frac{\partial x'^\mu}{\partial x^\nu} = \delta^\mu_\nu + \partial_\nu(\omega_a X^\mu_a) \tag{74a}$$

$$\frac{\partial x^\nu}{\partial x'^\mu} = \delta^\nu_\mu - \partial_\mu(\omega_a X^\nu_a) \tag{74b}$$

so that

$$\left| \frac{\partial x'}{\partial x} \right| = 1 + \partial_\mu(\omega_a X^\mu_a). \tag{74c}$$

Expanding to first order we have

$$\begin{aligned}\partial'_\mu\phi'(x') = \frac{\partial x^\nu}{\partial x'^\mu}\partial_\nu\mathcal{F}(\phi(x)) &= \frac{\partial x^\nu}{\partial x'^\mu}\partial_\nu(\phi+\omega_a\mathcal{F}_a)\\ &= [\delta^\nu_\mu - \partial_\mu(\omega_a X^\nu_a)][\partial_\nu\phi + \partial_\nu(\omega_a\mathcal{F}_a)]\\ &= \partial_\mu\phi + \partial_\mu(\omega_a\mathcal{F}_a) - \partial_\mu(\omega_a X^\nu_a)\partial_\nu\phi. \end{aligned} \tag{75}$$

Using equations (73b), (74c) and (75) in (72) we expand S' again to first order in ω_a:

$$\begin{aligned} S' &= \int d^4x \left|\frac{\partial x'}{\partial x}\right| \mathscr{L}(\phi+\omega_a\mathcal{F}_a, \partial_\mu\phi + \partial_\mu(\omega_a\mathcal{F}_a) - \partial_\mu(\omega_a X^\nu_a)\partial_\nu\phi)\\ &= \int d^4x\,[1+\partial_\mu(\omega_a X^\mu_a)]\big\{\mathscr{L}(\phi,\partial_\mu\phi) + \frac{\partial\mathscr{L}}{\partial\phi}\omega_a\mathcal{F}_a\\ &\qquad\qquad + \frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi)}[\partial_\mu(\omega_a\mathcal{F}_a) - \partial_\mu(\omega_a X^\nu_a)\partial_\nu\phi]\big\}\\ &= S + \int d^4x\left\{\frac{\partial\mathscr{L}}{\partial\phi}\omega_a\mathcal{F}_a + \frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi)}[\partial_\mu(\omega_a\mathcal{F}_a) - \partial_\mu(\omega_a X^\nu_a)\partial_\nu\phi]\right.\\ &\qquad\qquad \left. + \partial_\mu(\omega_a X^\mu_a)\mathscr{L}\right\}. \end{aligned}$$

Therefore

$$\begin{aligned}\delta S = S' - S = \int d^4x &\left\{\left[\frac{\partial\mathscr{L}}{\partial\phi} - \partial_\mu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi)}\right)\right]\omega_a\mathcal{F}_a + \partial_\mu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi)}\omega_a\mathcal{F}_a\right)\right.\\ &\left. - \frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi)}\partial_\mu(\omega_a X^\nu_a)\partial_\nu\phi + \partial_\mu(\omega_a X^\nu_a)g^\mu_\nu\mathscr{L}\right\}.\end{aligned}$$

But the term in square brackets vanishes by the Euler-Lagrange equation (which we assume that the fields satisfy) so we have

$$\begin{aligned}\delta S &= \int d^4x\left\{\partial_\mu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi)}\omega_a\mathcal{F}_a\right) - \left[\frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi)}\partial_\nu\phi - g^\mu_\nu\mathscr{L}\right]\partial_\mu(\omega_a X^\nu_a)\right\}\\ &= \int d^4x\left\{\partial_\mu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi)}\mathcal{F}_a\right)\omega_a + \frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi)}\mathcal{F}_a(\partial_\mu\omega_a)\right.\\ &\quad\left. - \left[\frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi)}\partial_\nu\phi - g^\mu_\nu\mathscr{L}\right]\omega_a(\partial_\mu X^\nu_a) - \left[\frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi)}\partial_\nu\phi - g^\mu_\nu\mathscr{L}\right]X^\nu_a(\partial_\mu\omega_a)\right\}\\ &= \int d^4x\left\{\partial_\mu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi)}\mathcal{F}_a\right) - \left[\frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi)}\partial_\nu\phi - g^\mu_\nu\mathscr{L}\right]\partial_\mu X^\nu_a\right\}\omega_a\\ &\qquad + \int d^4x\left\{\frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi)}\mathcal{F}_a - \left[\frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi)}\partial_\nu\phi - g^\mu_\nu\mathscr{L}\right]X^\nu_a\right\}\partial_\mu\omega_a. \end{aligned}\tag{76}$$

Next comes the clever argument. Suppose that our transformation is global, meaning that the parameters ω_a are constants. Then $\partial_\mu \omega_a = 0$ and the second integral in equation (76) vanishes, so if we are to have $\delta S = 0$, then the first integral must also vanish. But ω_a is an arbitrary *constant*, so it comes out of the integral, and we conclude that the first *integral* must vanish. Now suppose that we have the more general case where the transformation is arbitrary and local, i.e., it could be that $\partial_\mu \omega_a \neq 0$. Note that a global transformation is a special case of a local transformation. If the action is still to be invariant for *arbitrary* $\partial_\mu \omega_a$, then for those where ω_a is a constant we must have the first integral (without including the constant ω_a) vanish. But this integral is independent of ω_a, and since $\delta S = 0$ for *both* constant and non-constant ω_a's, it must vanish identically no matter which type of ω_a we consider. Then we are left with only the second integral, so for arbitrary $\partial_\mu \omega_a$ the rest of its *integrand* must vanish.

Let us write

$$\delta S = \int d^4x \, j_a^\mu \, \partial_\mu \omega_a$$

where

$$j_a^\mu = \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi)} \mathcal{F}_a - \left[\frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi)} \partial_\nu \phi - g_\nu^\mu \mathscr{L} \right] X_a^\nu \tag{77}$$

is called the **current** associated with the transformation. Integrating by parts and assuming that the fields vanish at infinity we obtain

$$\delta S = -\int d^4x \, (\partial_\mu j_a^\mu) \omega_a.$$

If we now assume that the fields obey the classical equations of motion (the Euler-Lagrange equation) and the action vanishes for arbitrary (continuous) position-dependent variations ω_a, then we have the conservation law

$$\partial_\mu j_a^\mu = 0$$

with the corresponding conserved charges

$$Q_a = \int j_a^0 \, d^3x.$$

Equation (77) is essentially identical to equation (64). (The connection is $\delta x^\mu = \omega_a X_a^\mu$ and $\Delta\phi = \omega_a \mathcal{F}_a$. What we wrote as j^μ in equations (64) and (65) would then apply to each individual and independent component a.)