Noether's Theorem

Symmetry and Conservation

By

Tristan Johnson

****

Submitted in partial fulfillment of the requirements for
Honors in the Department of Mathematics

Union College
May 17, 2016

Abstract

Johnson, Tristan   Noether's Theorem: Symmetry and Conservation.

Department of Mathematics, March 2016

Advisor: Jeffrey Jauregui

A common calculus problem is to find an input that optimizes (maximizes or minimizes) a function. An extension of this problem is to find a function that optimizes an expression depending on the function. This paper studies how small (differentiable) variations of functions give us more information about expressions dependent on these functions. Specifically, Noether's Theorem states that in a system of functions, each differential symmetry – or small variation where the system is invariant– constructs a conserved quantity. We will describe, interpret and prove Noether's Theorem using techniques from linear algebra, differential geometry, and the calculus of variations. Furthermore, we will apply these techniques and Noether's Theorem to physical examples such as the wave equation, the Schrödinger equation, and electromagnetism.

Contents

Figures

1 Introduction

1.1 The Brachistochrone

Our motivation in this study will come from a field of mathematics called the calculus of variations. A common problem that arises from the study of smooth functions is optimization. At the low level, we attempt to minimize or maximize some fixed quantity. However, we may extend this to the optimization of higher level objects, such as functions. In this study, we will be mainly motivated by physical applications of these optimization problems. Consider the following optimization problem:

Example 1.1. Given a point $p = (p_x, p_y) \in \mathbb{R}^2$ with $p_y < 0, p_x \neq 0$, find the plane curve $\gamma(t)$ that will allow a particle to travel from $(0,0)$ to p in minimal time only by gravity.

The solution to this problem is the brachistochrone (shortest time) curve. If s is the arc-length traveled by the particle, then the speed of the particle along $\gamma(t)$ is $v = \frac{ds}{dt}$. The total time taken is then

$$T = \int_{(0,0)}^{p} \frac{ds}{v}$$

If we assume the solution to this problem is a function $y(x)$, we can write

$$ds = \sqrt{1 + (y')^2}dx.$$

We are justified in doing so since if γ cannot be written as $y(x)$ then it crosses two points $(x, y_1), (x, y_2)$, so it either moves straight down or crosses back on itself. In either case, it is not headed toward p and so we would not expect a curve that is not a function $y(x)$ to be optimal. Suppose the particle has mass m. Since our acceleration is only due to gravity, the kinetic energy of the particle gives us

$$\frac{m}{2}v^2 = -gmy$$

where g is the gravitational acceleration constant. If units are chosen such that gravitational acceleration is $\frac{1}{2}$, then we get $v = \sqrt{-y}$ and so we can rewrite the integral in terms of y as

$$T = \int_0^{p_x} \sqrt{\frac{1 + (y')^2}{-y}}dx$$

If we define

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

Then our solution y is that which minimizes

$$\int_0^{p_x} \mathcal{L}(x, f, f')dx.$$

Solving for y is nontrivial, and may become even more complex in similar problems. We will develop tools that allow us to solve optimization problems of variable functions such as the brachistochrone.

1.2 Calculus of Variations

The name given to problems of optimization of variable functions is the Calculus of Variations. Normally, the expressions that occur involve coordinates x, and a function f and its first derivative. It is possible to study calculus of variations involving any order derivatives of f, but we will only consider the first derivative case.

Definition 1.1. Suppose $F : [x_0, x_1] \times \mathbb{R}^2 \to \mathbb{R}$ is smooth. A function

$$\mathcal{L} : [x_0, x_1] \times C^{\infty}([x_0, x_1])^1 \to \mathbb{R} \text{ by } \mathcal{L}(x, f) = F(x, f(x), f'(x))$$

is called a *Lagrangian* of f. We write $\mathcal{L}(x, f, f')$ to denote $\mathcal{L}(x, f)$. The function of f

$$I(f) = \int_{x_0}^{x_1} \mathcal{L}(x, f, f')dx$$

is known as a *functional*, and is the *action* of $\mathcal{L}$.

Notice that the derivatives

$$\frac{\partial \mathcal{L}}{\partial f}, \frac{\partial \mathcal{L}}{\partial f'}$$

are well-defined as

$$\frac{\partial \mathcal{L}}{\partial f} = \left.\frac{\partial F(x_1, x_2, x_3)}{\partial x_2}\right|_{x, f(x), f'(x)}$$
$$\frac{\partial \mathcal{L}}{\partial f'} = \left.\frac{\partial F(x_1, x_2, x_3)}{\partial x_3}\right|_{x, f(x), f'(x)}$$

Fix a Lagrangian $\mathcal{L}$ of $[x_0, x_1]$ and assume we want to extremize its action

$$I(f) = \int_{x_0}^{x_1} \mathcal{L}(x, f, f')dx$$

The approach to these problems is reminiscent of that of homotopies: we will create a family of curves $f_s, s \in (-\epsilon, \epsilon)$ that 'approximate' f and optimize over the family.

Definition 1.2. Suppose $f : [x_0, x_1] \to \mathbb{R}$ is smooth and let $\{f_s\}_{s \in (-\epsilon, \epsilon)}$ be a family of functions where

- $f_s(x_0) = f(x_0)$
- $f_s(x_1) = f(x_1)$
- $f_0 = f$
- $f_s(x)$ is smooth in x and s

Then the family f_s is called a *smooth variation* of f.

[1]For $X \subseteq \mathbb{R}$, $C^{\infty}(X)$ is the set of all smooth functions $f : X \to \mathbb{R}$

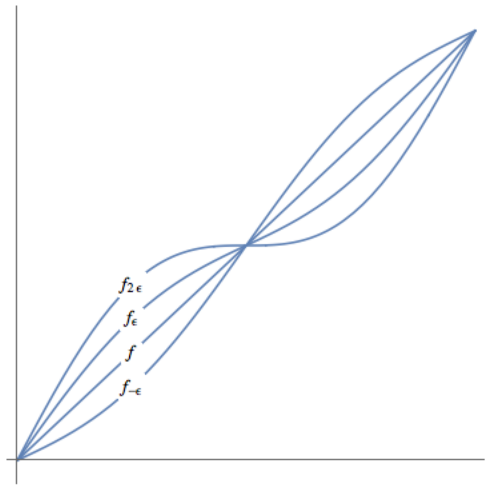

Figure 1: A smooth variation f_s of f

Explicitly, the last condition of a variation means that if we write a function

$$\bar{f} : [x_0, x_1] \times (-\epsilon, \epsilon) \to \mathbb{R} \text{ by } \bar{f}(x, s) = f_s(x),$$

then $\bar{f}$ is smooth. This concept allows us to compute the derivative of a smooth family, so that

$$\frac{df_s(x)}{ds} = \frac{\partial \bar{f}(x, s)}{\partial s}$$

A smooth variation may be thought of as slight perturbations of some base function f, as shown in Figure 1.

Definition 1.3. Suppose $\mathcal{L}(x, f, f')$ is a Lagrangian and I its action. Then f is a *solution* of $\mathcal{L}$ provided that for every smooth variation f_s of f, the function

$$I(f_s) = \int_{x_0}^{x_1} \mathcal{L}(x, f_s, f_s')dx$$

has

$$\frac{d}{ds} I(f_s)\bigg|_{s=0} = 0$$

We will now find a tool using variations that will be useful for the rest of our discussion.

Theorem 1.1. Suppose $\mathcal{L}(x, f, f')$ is a Lagrangian and f is a solution of $\mathcal{L}$. Then the (one-dimensional) *Euler–Lagrange* equation

$$\frac{\partial \mathcal{L}}{\partial f} - \frac{d}{dx}\frac{\partial \mathcal{L}}{\partial f'} = 0$$

is satisfied.

Before proving this statement, we need two lemmas from analysis that will allow us to remove the integral out of the functional I and only consider $\mathcal{L}$.

Lemma 1.1.1. Given smooth $f : [x_0, x_1] \to \mathbb{R}$ and $h : [x_0, x_1] \to \mathbb{R}$ with $h(x_0) = h(x_1) = 0$, there is a variation f_s of f satisfying

$$\left.\frac{df_s}{ds}\right|_{s=0} = h$$

Proof. Consider the variation

$$f_s(x) = sh(x) + f(x)$$

Notice this is a variation since

$$\begin{aligned}
f_s(x_0) &= sh(x_0) + f(x_0) \\
&= y_0 \\
f_s(x_0) &= sh(x_1) + f(x_1) \\
&= y_1 \\
f_0(x) &= 0 + f(x) \\
&= f(x) \\
f_s(x) &\text{ is smooth since } f \text{ and } h \text{ are smooth}
\end{aligned}$$

Then we get

$$\begin{aligned}
\left.\frac{df_s}{ds}\right|_{s=0} &= \left.\frac{d(sh+f)}{ds}\right|_{s=0} \\
&= h\big|_{s=0} \\
&= h
\end{aligned}$$

■

Lemma 1.1.2. Let $g : [x_0, x_1] \to \mathbb{R}$ be smooth and suppose

$$0 = \int_{x_0}^{x_1} g(x)h(x)dx$$

for every smooth $h : [x_0, x_1] \to \mathbb{R}$ with $h(x_0) = h(x_1) = 0$. Then $g \equiv 0$.

Proof. Assume, for contradiction, that there is some point c where $g(c) \neq 0$. Without loss of generality assume $g(c) > 0$. Since g is smooth we then know there is a neighborhood of c, say $(a, b) \subseteq [x_0, x_1]$ where $g(x) > 0$.

Define

$$B_a(x) = \begin{cases} e^{-\frac{1}{(x-a)^2}} & x > a \\ 0 & x \leqslant a \end{cases}$$

$$B_b(x) = \begin{cases} e^{-\frac{1}{(b-x)^2}} & x < b \\ 0 & x \geqslant b \end{cases}$$

Define[2]

$$h(x) = B_a(x) \cdot B_b(x)$$

[2] h is called a 'bump' function

h is displayed in Figure 2.

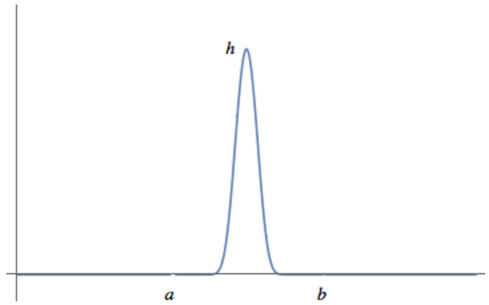

Figure 2: The bump function h

Notice that $h(x_0) = h(x_1) = 0$ and that $h(x) > 0$ for $x \in (a, b)$.
Claim: B_a is smooth on $[x_0, x_1]$.

Proof of claim. It is clear that B_a is smooth away from a. At a,

$$\begin{aligned} B_a'(x) &= \frac{2}{(x-a)^3} B_a(x) \\ B_a''(x) &= \left(\frac{2}{(x-a)^3}\right)^2 B_a(x) - \frac{6}{(x-a)^4} B_a \\ &\vdots \\ B_a^{(n)}(x) &= R_n(x) B_a(x) \end{aligned}$$

where R_n is some rational function of x. Then we can see

$$\lim_{x \to a^+} B_a^{(n)}(x) = 0$$

since

$$\lim_{x \to a^+} B_a(x) = 0$$

and

$$\lim_{x \to a^+} \frac{B_a'(x)}{R'(x)} = 0$$

Letting

$$y = \frac{1}{x-a}$$

We see that

$$\lim_{x \to a^+} B_a^{(n)}(x) = 0 = \lim_{y \to \infty} e^{-y^2} R_n\left(\frac{1+ay}{y}\right)$$

The exponential term will dominate the rational term as $y \to \infty$, so for any n

$$\lim_{x \to a^+} B_a^{(n)}(x) = 0$$

Furthermore, it is clear that for any n

$$\lim_{x \to a^-} B_a^{(n)}(x) = 0$$

since $B_a^{(n)}(x) = 0$ for $x \leqslant a$.

Thus $B^{(n)}(a)$ is continuous for any n, so B_a is smooth. ■

Similarly to the claim, B_b is smooth, and since h is their product, h is smooth. Then by assumption,

$$\begin{aligned} 0 &= \int_{x_0}^{x_1} g(x)h(x)dx \\ &= \int_a^b g(x)h(x)dx && h(x) = 0 \text{ outside (a,b)} \\ &> 0 && g(x) > 0, h(x) > 0 \ \forall x \in (a,b) \end{aligned}$$

is a contradiction.

■

Proof of Theorem 1.1. Let f_s be an arbitrary smooth variation. Notice that the first two conditions of the definition of a smooth variation imply

$$\left.\frac{df_s}{ds}\right|_{x=x_0} = 0 = \left.\frac{df_s}{ds}\right|_{x=x_1}$$

Since f is a solution of $\mathcal{L}$, writing the function of s

$$I(f_s) = \int_{x_0}^{x_1} \mathcal{L}(x, f_s, f_s')dx$$

we know that

$$\left.\frac{d}{ds} I(f_s)\right|_{s=0} = 0$$

Computing this derivative at $s = 0$, we find

$$
\begin{aligned}
0 &= \frac{d}{ds} I \\
&= \frac{d}{ds} \int_{x_0}^{x_1} \mathcal{L}(x, f_s, f_s') dx \\
&= \int_{x_0}^{x_1} \frac{\partial}{\partial s} \mathcal{L}(x, f_s, f_s') dx \\
&= \int_{x_0}^{x_1} \frac{\partial \mathcal{L}}{\partial f_s} \frac{df_s}{ds} + \frac{\partial \mathcal{L}}{\partial f_s'} \cdot \frac{d}{ds} \left(\frac{df_s}{dx} \right) dx && \text{chain rule} \\
&= \int_{x_0}^{x_1} \frac{\partial \mathcal{L}}{\partial f_s} \frac{df_s}{ds} + \frac{\partial \mathcal{L}}{\partial f_s'} \cdot \frac{d}{dx} \left(\frac{df_s}{ds} \right) dx && f_s(x) \text{ is smooth} \\
&= \int_{x_0}^{x_1} \frac{\partial \mathcal{L}}{\partial f_s} \frac{df_s}{ds} dx + \frac{df_s}{ds} \frac{\partial \mathcal{L}}{\partial f_s'} \bigg|_{x_0}^{x_1} - \int_{x_0}^{x_1} \frac{df_s}{ds} \frac{d}{dx} \left(\frac{\partial \mathcal{L}}{\partial f_s'} \right) dx && \text{integrate by parts} \\
&= \int_{x_0}^{x_1} \left(\frac{\partial \mathcal{L}}{\partial f_s} - \frac{d}{dx} \left(\frac{\partial \mathcal{L}}{\partial f_s'} \right) \right) \frac{df_s}{ds} dx && \frac{df_s}{ds} \bigg|_{x = x_0, x_1} = 0 \\
&= \int_{x_0}^{x_1} \left(\frac{\partial \mathcal{L}}{\partial f} - \frac{d}{dx} \left(\frac{\partial \mathcal{L}}{\partial f'} \right) \right) \frac{df_s}{ds} \bigg|_{s=0} dx && s = 0
\end{aligned}
$$

By Lemma 1.1.1 we know that

$$\frac{df_s}{ds} \bigg|_{s=0}$$

could be any smooth $h(x)$ where $h(x_0) = h(x_1) = 0$. Then by Lemma 1.1.2 we get

$$\frac{\partial \mathcal{L}}{\partial f} - \frac{d}{dx} \frac{\partial \mathcal{L}}{\partial f'} = 0$$

■

Corollary. If $\mathcal{L}(x, f, f')$ is a Lagrangian that does not depend explicitly on x and f is a solution of $\mathcal{L}$, then

$$f' \frac{\partial \mathcal{L}}{\partial f'} - \mathcal{L} = \text{const}$$

Proof. Since $\mathcal{L}$ does not depend explicitly on x, we know that $\frac{\partial \mathcal{L}}{\partial x} = 0$. Furthermore, the Euler–Lagrange equations are satisfied for $\mathcal{L}(x, f, f')$. Using this, we may compute

$$
\begin{aligned}
\frac{d}{dx} \left(f' \frac{\partial \mathcal{L}}{\partial f'} - \mathcal{L} \right) &= f'' \frac{\partial \mathcal{L}}{\partial f'} + f' \frac{d}{dx} \frac{\partial \mathcal{L}}{\partial f'} - \frac{\partial \mathcal{L}}{\partial x} - \frac{\partial \mathcal{L}}{\partial f} f' - \frac{\partial \mathcal{L}}{\partial f'} f'' \\
&= -f' \left(\frac{\partial \mathcal{L}}{\partial f} - \frac{d}{dx} \frac{\partial \mathcal{L}}{\partial f'} \right) - \frac{\partial \mathcal{L}}{\partial x} \\
&= 0 && \text{Euler–Lagrange, } \frac{\partial \mathcal{L}}{\partial x} = 0
\end{aligned}
$$

From here, integrating gives the result. ■

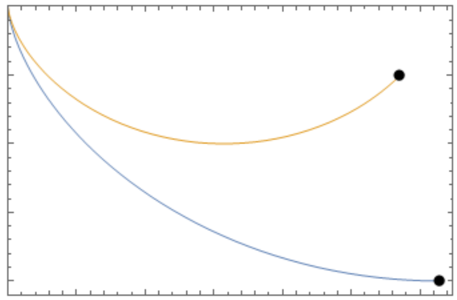

Figure 3: Brachistochrone curves

Corollary. The brachistochrone is given by the parametric curve

$$x(\theta) = a(\theta - \sin\theta) \quad y(\theta) = a(1 - \cos\theta) \quad \theta \in [0, \theta_f]$$

The constant a and θ_f may be chosen such that

$$(x(0), y(0)) = (0, 0) \quad (x(\theta_f), y(\theta_f)) = (p_x, p_y)$$

where $p_x \neq 0, p_y \leqslant 0$.

Proof. Recall the brachistochrone Lagrangian

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

Since $\mathcal{L}$ does not depend explicitly on x, the previous corollary gives us

$$\begin{aligned}
\text{const} &= y' \frac{\partial \mathcal{L}}{\partial y'} - \mathcal{L} \\
&= \frac{(y')^2}{\sqrt{-y(1 + (y')^2)}} - \sqrt{\frac{1 + (y')^2}{-y}} \\
&= \frac{(y')^2 - 1 - (y')^2}{\sqrt{-y(1 + (y')^2)}} \\
&= \frac{-1}{\sqrt{-y(1 + (y')^2)}}
\end{aligned}$$

Then we may assume for a constant a,

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

so

$$y' = \sqrt{\frac{2a + y}{-y}}$$

Then we know

$$x = \int \frac{dx}{y'} = \int \sqrt{\frac{-y}{2a+y}} dy$$

This integral may be solved with the substitution $y = -a(1 - \cos 2\theta) = -2a \sin^2 \theta$:

$$\begin{aligned}
\int \sqrt{\frac{-y}{2a+y}} dy &= \int \sqrt{\frac{a \sin^2 \theta}{2a - 2a \sin^2 \theta}} 4a \sin\theta \cos\theta d\theta \\
&= 4a \int \frac{\sin\theta}{\cos\theta} \sin\theta \cos\theta d\theta \\
&= 4a \int \sin^2 \theta d\theta \\
&= 2a \int 1 - \cos 2\theta d\theta \\
&= 2a\theta - a \sin 2\theta
\end{aligned}$$

Replacing 2θ with θ, we get the parametric equations

$$x = a(\theta - \sin\theta) \quad y = -a(1 - \cos\theta)$$

■

Some brachistochrone curves are plotted in Figure 3. An interesting fact of the brachistochrone is that although these curves minimize the time taken to travel the length of the curve, they may still bend back up away from the accelerating force to reach their destination.

Let us consider one more example involving the calculus of variations in one dimension.

Example 1.2. Suppose a particle's position is constrained to movement in one dimension. That is, let $x(t)$ describe the particle's position, where $t \in [t_0, t_1]$. Consider the Lagrangian

$$\int_{t_0}^{t_1} T - V(x) dt$$

where T, V are kinetic and potential energy of the particle.. Notice we are assuming V is only a function of x (this means that the potential energy is always the same for a given x, so we are in a closed system with conservative forces). We will find a solution x of the Lagrangian $T - V$.

Since T is kinetic energy, we know

$$T = \frac{1}{2} m \dot{x}^2$$

where m is the particle's mass and $\dot{x} = \frac{dx}{dt}$. Then our Lagrangian is

$$\mathcal{L} = \frac{1}{2} m \dot{x}^2 - V(x)$$

Since this does not depend explicitly on t, we can write

$$\begin{aligned}\text{const} &= \dot{x}\frac{\partial \mathcal{L}}{\partial \dot{x}} - \mathcal{L} \\ &= \dot{x}(m\dot{x}) - \left(\frac{1}{2}m\dot{x}^2 - V(x)\right) && V \text{ does not depend on } \dot{x} \\ &= \frac{1}{2}m\dot{x}^2 + V(x) \\ &= T + V\end{aligned}$$

$T + V$ is the total energy of the system, so we find that under these conditions, energy is conserved.

1.3 Generalized Euler–Lagrange Equations

We may easily generalize the Euler–Lagrange equation to a set of equations for k functions in n dimensions. To do so, we need to define a generalized Lagrangian, and the space it acts on.

Definition 1.4. A *smooth region* Ω of $\mathbb{R}^n$ is an bounded, connected, open subset of $\mathbb{R}^n$ such that the *boundary* $\partial\Omega$ is smooth.[3] The *closure* of Ω, denoted $\overline{\Omega}$, is $\Omega \cup \partial\Omega$.

Definition 1.5. Suppose Ω is a smooth region of $\mathbb{R}^n$. Let $\boldsymbol{x} \in \overline{\Omega} \subseteq \mathbb{R}^n$, and suppose $f^1, \ldots, f^k : \overline{\Omega} \to \mathbb{R}$ (f^i is the i^{th} function, not a derivative or exponent) are smooth. A smooth function

$$\mathcal{L}\left(\boldsymbol{x}, f^1, \ldots, f^k, \frac{\partial}{\partial x_1}f^1, \ldots, \frac{\partial}{\partial x_n}f^1, \frac{\partial}{\partial x_1}f^k, \ldots, \frac{\partial}{\partial x_n}f^k\right)$$

is a (generalized) *Lagrangian* of $f^1, \ldots, f^k$ and

$$I(f^1, \ldots, f^k) = \int_{\overline{\Omega}} \mathcal{L}\left(\boldsymbol{x}, f^1, \ldots, f^k, \frac{\partial}{\partial x_1}f^1, \ldots, \frac{\partial}{\partial x_n}f^1, \frac{\partial}{\partial x_1}f^k, \ldots, \frac{\partial}{\partial x_n}f^k\right) d\boldsymbol{x}$$

is a *functional*, the *action* of $\mathcal{L}$.

Definition 1.6. Suppose $f^1, \ldots, f^k : \overline{\Omega} \to \mathbb{R}$ are smooth, where Ω is a smooth region. Let $\{(f_s^1, \ldots, f_s^k)\}_{s \in (-\epsilon, \epsilon)}$ be a family of functions where for $1 \leqslant i \leqslant k$

- $f_s^i(\boldsymbol{x}) = f^i(\boldsymbol{x})$ for $\boldsymbol{x} \in \partial\Omega$.
- $f_0^i = f^i$.
- $f_s^i(\boldsymbol{x})$ is smooth in $\boldsymbol{x}$ and s.

Then the family f_s^i is called a (generalized) *smooth variation* of f^i, $1 \leqslant i \leqslant k$.

Smoothness in s is similar to the one-dimensional case, in that f_s^i is smooth provided

$$\bar{f}^i(\boldsymbol{x}, s) = f_s^i(\boldsymbol{x})$$

has arbitrary partials of s and $\boldsymbol{x}$.

A solution $f^1, \ldots, f^k$ is also defined in a similar way:

[3]By boundary we mean the set of limit points of Ω in $\mathbb{R}^n$ that are not themselves in Ω. The boundary is smooth if it is a differentiable manifold of dimension $n-1$. Think of a smooth surface bounding an open region of $\mathbb{R}^3$.

Definition 1.7. The $k-$tuple of functions $(f^1, \ldots, f^k)$ is a solution for the Lagrangian $\mathcal{L}$ provided that for every smooth variation f^i_s, the function

$$I(f^i_s) = \int_\Omega \mathcal{L}\left(\boldsymbol{x}, f^i_s, \frac{\partial}{\partial x_j} f^i_s\right) d\boldsymbol{x}, \quad 1 \leqslant i \leqslant k, 1 \leqslant j \leqslant n$$

has

$$\frac{d}{ds} I(f^i_s)\bigg|_{s=0} = 0$$

These definitions give the extension of the Euler–Lagrange equations.

Theorem 1.2. Suppose

$$\mathcal{L}\left(\boldsymbol{x}, f_1, \ldots, f_k, \frac{\partial}{\partial x_1} f_1, \ldots, \frac{\partial}{\partial x_n} f_1, \frac{\partial}{\partial x_1} f_k, \ldots, \frac{\partial}{\partial x_n} f_k\right)$$

is a Lagrangian of k functions over n coordinates and $f^1, \ldots, f^k$ is a solution of $\mathcal{L}$. Then the generalized system of Euler–Lagrange equations:

$$\frac{\partial \mathcal{L}}{\partial f_i} - \sum_{j=1}^{n} \frac{\partial}{\partial x_j} \frac{\partial \mathcal{L}}{\partial \left(\frac{\partial f^i}{\partial x_j}\right)} = 0, \quad 1 \leqslant i \leqslant k$$

are satisfied.

The proof of this theorem is similar to the one-dimensional case, using the divergence theorem to perform the integration by parts.

1.4 Physical Examples of Lagrangians

We will examine a few more applications of the Euler–Lagrange equations in physical spacetime that will serve as motivation for the rest of our work. Refer to *spacetime*[4] as $M = \mathbb{R}^{n+1}$, which will become notationally clear in later sections. M has coordinates $(t, x_1, \ldots, x_n)$, where t represents time and may also be referred to as x_0, while $x_1, \ldots, x_n$ represent space. Normally, 3-dimensional space has $n = 3$, so M is a 4-dimensional space. From here on out, when we refer to M we will assume it has dimension $n + 1$, without explicitly stating so. Choose units of M such that the speed of light $c = 1$, allowing us to ignore many constants. We will refer to the standard basis of M as $\{\boldsymbol{e}_0, \ldots, \boldsymbol{e}_n\}$ where $\boldsymbol{e}_i$ has the x_jth component 0 if $j \neq i$, and 1 if $j = i$.

Example 1.3. For the first of our examples of Lagrangians in M, we will examine systems that behave as waves.

Definition 1.8. Let $\varphi : M \to \mathbb{R}$ be smooth. The *Wave Lagrangian* is

$$\mathcal{W}(\varphi) = \|\nabla\varphi\|^2 - \dot{\varphi}^2$$

where $\dot{\varphi}$ is the time derivative and

$$\nabla\varphi = \frac{\partial\varphi}{\partial x_1}\boldsymbol{e}_1 + \ldots + \frac{\partial\varphi}{\partial x_n}\boldsymbol{e}_n$$

is the spatial gradient.

[4] M is actually a differentiable manifold of dimension $n + 1$

The name suggests that this Lagrangian is somehow related to the wave equation,

$$\ddot{\varphi} = \Delta\varphi$$

where

$$\Delta\varphi = \sum_{i=1}^{n} \frac{\partial^2\varphi}{\partial x_i^2}$$

is the Laplacian.

Theorem 1.3. A field φ satisfies the Wave Equation iff φ is a solution of the Wave Lagrangian $\mathcal{W}$.

Proof. Notice that $\mathcal{W}$ does not depend explicitly on φ, so

$$\frac{\partial\mathcal{W}}{\partial\varphi} = 0$$

Then the left-side of the Euler–Lagrange equation is

$$\begin{aligned}\sum_{j=0}^{n} \frac{\partial}{\partial x_j}\frac{\partial\mathcal{W}}{\partial(\varphi_{x_j})} &= \frac{\partial}{\partial t}2\dot{\varphi} - \sum_{i=1}^{n}\frac{\partial}{\partial x_i}\left(2\frac{\partial\varphi}{\partial x_i}\right) \\ &= 2\left(\frac{\partial^2\varphi}{\partial t^2} - \Delta\varphi\right)\end{aligned}$$

Now we see

$$\frac{\partial\mathcal{W}}{\partial\varphi} - \sum_{i=0}^{n}\frac{\partial}{\partial x_i}\frac{\partial\mathcal{W}}{\partial\left(\frac{\partial\varphi}{\partial x_i}\right)} = 0 \iff \frac{\partial^2\varphi}{\partial t^2} = \Delta\varphi$$

$\blacksquare$

Example 1.4.
Just as we found a Lagrangian $\mathcal{W}$ that represented the Wave Equation, we will find one for the Schrödinger Equation:

$$i\frac{\partial\varphi}{dt} = -\frac{1}{2m}\Delta\varphi + V\varphi$$

Physically, the Schrödinger Equation describes properties of a quantum particle with wave-function φ (traditionally, the wave-function is denoted by ψ, but we will reserve ψ for a later purpose), V is the potential energy of the particle at $\boldsymbol{x}$, and m is its mass (we choose units such that $\hbar = 1$). Although we have been discussing Lagrangians and functions over real spaces, the case for complex spaces is essentially identical, which is necessary for φ.

Definition 1.9. Suppose $V : \Omega \to \mathbb{R}$ and m is a non-negative real constant. The Schrödinger Lagrangian $\mathcal{S}$ is defined as:

$$\mathcal{S}\left(\boldsymbol{x}, \varphi, \varphi^*, \frac{d}{dx_j}\varphi, \frac{d}{dx_j}\varphi^*\right) = \mathcal{S}(\boldsymbol{x}, \varphi) = \frac{1}{2m}\nabla\varphi\cdot\nabla\varphi^* + V(\boldsymbol{x})\varphi\varphi^* - \frac{i}{2}(\varphi^*\dot{\varphi} - \varphi\dot{\varphi}^*)$$

Theorem 1.4. Let $V : \Omega \to \mathbb{R}, m \in \mathbb{R}$. Then a field φ satisfies the Schrödinger equation (and its complex conjugate) iff it is a solution of the Schrödinger Lagrangian $\mathcal{S}$.

Proof. We will again compute the left side of the Euler–Lagrange equations for φ^*, and the case for φ will give the complex conjugate.

We may find

$$\begin{aligned}
\frac{\partial \mathcal{S}}{\partial \varphi^*} &= V\varphi - \frac{i}{2}\dot{\varphi} \\
\sum_{i=0}^{n} \frac{\partial}{\partial x_i} \frac{\partial \mathcal{S}}{\partial \left(\frac{\partial \varphi^*}{\partial x_i}\right)} &= \frac{\partial}{\partial t}\left(\frac{i}{2}\varphi\right) + \sum_{i=1}^{n} \frac{\partial}{\partial x_i}\left(\frac{1}{2m}\frac{\partial \varphi}{\partial x_i}\right) \\
&= \frac{i}{2}\dot{\varphi} + \frac{1}{2m}\sum_{i=1}^{n}\frac{\partial^2 \varphi}{\partial x_i^2} \\
&= \frac{i}{2}\dot{\varphi} + \frac{1}{2m}\Delta\varphi
\end{aligned}$$

Then the Euler–Lagrange equations are satisfied iff

$$V\varphi - \frac{i}{2}\dot{\varphi} - \frac{i}{2}\dot{\varphi} - \frac{1}{2m}\Delta\varphi = 0 \iff i\frac{\partial \varphi}{\partial t} = -\frac{1}{2m}\Delta\varphi + V\varphi$$

■

1.5 Introduction to Noether's Theorem

Up to now, we have considered an arbitrary variation of f and the information this gives us about its Lagrangian. This section will begin to examine the information we get by choosing a specific 'variation.' However, we will first generalize the notion of variation to allow the coordinates of a Lagrangian to also transform.

Definition 1.10. Let Ω be a smooth region of M, and let $f^1, \ldots, f^k$ be smooth functions $f^i : \overline{\Omega} \to \mathbb{R}$. A *smooth transformation* of Ω and f^i is a family

$$\{(\boldsymbol{x}_s, f_s^i)\}_{s\in(-\epsilon,\epsilon)}$$

such that f_s^i is a smooth variation of f and $\boldsymbol{x}_s$ is a smooth variation of the identity function $\boldsymbol{x}$ on $\overline{\Omega}$. We write the transformation as $(\boldsymbol{x}_s, f_s^i)$.

For a smooth transformation $\{\boldsymbol{x}_s, f_s^i\}$, $\boldsymbol{x}_s = \boldsymbol{x}$ for $\boldsymbol{x} \in \partial\Omega$ by the definition of smooth variation. Since $\boldsymbol{x}_s$ is smooth in s and is fixed on the boundary of Ω, it must be that the image $\operatorname{im}(\boldsymbol{x}_s) \subseteq \Omega$ for small enough s. Then for a Lagrangian

$$\mathcal{L}\left(\boldsymbol{x}, f^i, \frac{\partial}{\partial x_j} f^i\right),$$

the expression

$$\mathcal{L}\left(\boldsymbol{x}_s, f_s^i, \frac{\partial}{\partial x_j} f_s^i\right)$$

is defined since $\boldsymbol{x}_s \in \Omega$ for every $\boldsymbol{x}, s$.

Intuitively, a symmetry is taken to be a transformation that leaves some quantity invariant. Since we have just shown that we may apply a smooth transformation to a Lagrangian, we will explicitly define a symmetry in terms of a transformation with a Lagrangian.

Definition 1.11. Let $\mathcal{L}$ be a Lagrangian of $f^1 \dots f^k$, where each f^i is defined over the closure of a smooth region Ω. Let $(\boldsymbol{x}_s, f^i_s)$ be a transformation where for $s \in (-\epsilon, \epsilon)$, $\mathrm{im}(\boldsymbol{x}_s) \subseteq \overline{\Omega}$. $(\boldsymbol{x}_s, f^i_s)$ is a *simple smooth symmetry* of $\mathcal{L}$ provided

$$I_\omega(f^i_s) = \int_\omega \mathcal{L}\left(\boldsymbol{x}_s, f^i_s, \frac{\partial}{\partial x_j} f^i_s\right) d\boldsymbol{x}_s$$

has

$$I_\omega(f^i_s) = I(f^i) = \int_\omega \mathcal{L}\left(\boldsymbol{x}, f^i, \frac{\partial}{\partial x_j} f^i\right) d\boldsymbol{x}$$

for all smooth regions $\omega \subseteq \mathrm{im}(\boldsymbol{x}_s)$.

We will describe the meaning of $d\boldsymbol{x}_s$ in Section 2.3. For now, recognize that since I is constant for every region, we see that over an arbitrary smooth region $\omega \subseteq \mathrm{im}(\boldsymbol{x}_s)$,

$$\begin{aligned} 0 &= \frac{\partial}{\partial s} \int_\omega \mathcal{L}\left(\boldsymbol{x}_s, f^i_s, \frac{\partial}{\partial x_j} f^i_s\right) d\boldsymbol{x}_s \bigg|_{s=0} \\ &= \int_\omega \frac{\partial}{\partial s} \mathcal{L}\left(\boldsymbol{x}_s, f^i_s, \frac{\partial}{\partial x_j} f^i_s\right) d\boldsymbol{x}_s \bigg|_{s=0} \end{aligned}$$

One may show that an integral is 0 over an arbitrary region iff the integrand itself must be 0. This gives us

$$0 = \frac{\partial}{\partial s} \mathcal{L}\left(\boldsymbol{x}_s, f^i_s, \frac{\partial}{\partial x_j} f^i_s\right) d\boldsymbol{x}_s \bigg|_{s=0}$$

We will take this last equation to be the definition of a more general form of symmetry:

Definition 1.12. Let $\mathcal{L}$ be a Lagrangian of $f^1 \dots f^k$, where each f^i is defined over the closure of a smooth region Ω. A transformation $(\boldsymbol{x}_s, f^i_s)$ is a *smooth symmetry*[5] of $\mathcal{L}$ provided

$$0 = \frac{\partial}{\partial s} \mathcal{L}\left(\boldsymbol{x}_s, f^i_s, \frac{\partial}{\partial x_j} f^i_s\right) d\boldsymbol{x}_s \bigg|_{s=0}$$

Our work leading to the definition of smooth symmetry gives the following statement:

Proposition. A simple smooth symmetry is a smooth symmetry.

Our (informal) statement of Noether's Theorem is now as follows:

Theorem. A smooth symmetry on a Lagrangian corresponds to a quantity Q conserved in time, called a conserved current.

We have already seen an example of Noether's Theorem in conservation of energy (Example 1.2). We found conservation of energy by recognizing that the Lagrangian $\mathcal{L}$ was not explicitly dependent on q. Then a transformation of $x(t)$ that leaves $x'(t)$ unchanged will be a symmetry of $\mathcal{L}$, for example, $x_s = x + s$. This symmetry corresponds to the conserved quantity of energy.

More generally, a special case of Noether's Theorem is found in the corollary to Theorem1.1. In one-dimension, if $\mathcal{L}$ does not depend explicitly on f, the symmetry that varies f but leaves f' unchanged results in the conserved quantity

$$f' \frac{\partial \mathcal{L}}{\partial f'} - \mathcal{L}$$

[5]There is an even more general definition of smooth symmetry in terms of gauge symmetries, which goes beyond the scope of this paper. In Section 2.4, the example for Electromagnetism requires this kind of symmetry.

Emmy Noether was primarily an algebraist, and was the founder of many algebraic ideas that motivated mathematical study in the 20$^{\text{th}}$ century. Whole fields such as algebraic topology or hypercomplex representations are attributed to her creativity, while also producing many important concepts within algebra such as ascending chains of ideals (Noetherian rings) or primary decomposition of ideals (Lasker-Noether Theorem). In 1915, her expertise of algebra was employed by two of her mentors (David Hilbert and Felix Klein, both equally prodigious mathematicians) to help them study the mathematics of relativity, as recently introduced by Albert Einstein. She published her nominal theorem in 1918.

As a female Jewish mathematician, Noether was frequently a victim of prejudice, but is now remembered as one of the greatest mathematicians in history. Despite the prejudice against women, she was granted formal admission as a lecturer at the University of Göttingen in 1919. After being dismissed from Göttingen by the Nazi regime for her heritage, she emigrated to the United States, teaching and studying for the remainder of her life. After her death, Einstein wrote 'Noether was the most significant creative mathematical genius thus far produced since the higher education of women began.'

2 Tensors and Other Tools

To begin discussing Noether's Theorem, we will need some tools that will allow us to describe the theorem and its proof more easily. These tools are called tensors, and are an extensions of simple linear algebra ideas.

2.1 Linear Algebra

Definition 2.1. Let V be a (finite dimensional) vector space over $\mathbb{R}$. The *dual space* V^* is the vector space of linear maps $\boldsymbol{\alpha} : V \to \mathbb{R}$, where vector addition and scalar multiplication are inherited from V. Members of the dual space are referred to as *covectors*. Notationally, we will refer to covectors with the bold greek alphabet and normal vectors with bold latin.

The dual space is itself a vector space generated directly from V. Although we may construct a basis for V^* directly from a basis for V, we would like to avoid using bases whenever possible so that the phenomena observed is coordinate independent. It may be easily shown beyond this that there is no natural (i.e. coordinate independent) isomorphism between V and V^* even though the spaces are in fact isomorphic. However, considering the 'double dual' $V^{**} = (V^*)^*$, there is a natural identification of V with V^{**}: given $\boldsymbol{v} \in V$, define $\boldsymbol{v}^{**}$ such that $\boldsymbol{v}^{**}(\boldsymbol{\alpha}) = \boldsymbol{\alpha}(\boldsymbol{v})$. This tells us that there are really only two distinct spaces that are generated from 'dualing' V, V itself and V^*. This notion of creating spaces from 'linear actions' on a space is where we derive tensors.

$$\begin{tikzcd} V \arrow[r, dashed, leftrightarrow] \arrow[rr, bend left, leftrightarrow] & V^* \arrow[r, dashed, leftrightarrow] \arrow[rr, bend right, leftrightarrow] & V^{**} \arrow[r, dashed, leftrightarrow] \arrow[rr, bend left, leftrightarrow] & V^{***} & \cdots \end{tikzcd}$$

Figure 4: A space naturally identifies with its double dual, not its dual

Definition 2.2. Let k, ℓ be non-negative integers, not both 0. A $(k, \ell)-$*tensor* over V is a function $\boldsymbol{T} : V^k \times (V^*)^l \to \mathbb{R}$ which is multi–linear. Explicitly, for any scalar c and

$1 \leqslant i \leqslant k, 1 \leqslant j \leqslant l$:

$$\begin{aligned}
&\boldsymbol{T}(\boldsymbol{v}_1, \ldots, c \cdot \boldsymbol{v}_i + \tilde{\boldsymbol{v}}_i, \ldots, \boldsymbol{v}_k, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_\ell) = \\
&c \cdot \boldsymbol{T}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_i, \ldots, \boldsymbol{v}_k, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_l) + \boldsymbol{T}(\boldsymbol{v}_1, \ldots, \tilde{\boldsymbol{v}}_i, \ldots, \boldsymbol{v}_k, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_l) \\
&\boldsymbol{T}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k, \boldsymbol{\alpha}_1, \ldots, c \cdot \boldsymbol{\alpha}_j + \tilde{\boldsymbol{\alpha}}_i, \ldots, \boldsymbol{\alpha}_\ell) = \\
&c \cdot \boldsymbol{T}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_j, \ldots, \boldsymbol{\alpha}_\ell) + \boldsymbol{T}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k, \boldsymbol{\alpha}_1, \ldots, \tilde{\boldsymbol{\alpha}}_i, \ldots, \boldsymbol{\alpha}_\ell)
\end{aligned}$$

The set of $(k, \ell)-$tensors over V is denoted $\mathcal{T}^{k,\ell}(V)$. We will emphasize that $\boldsymbol{T}$ is a tensor (as opposed to a scalar) by using bold letters.

Similarly to the dual space, $\mathcal{T}^{k,\ell}(V)$ forms a vector space with addition and multiplication inherited from V. A tensor need only take in vectors and covectors, since other spaces such as V^{**} or V^{***} naturally identify with V or V^*, as shown in Figure 4. This will be represented notationally in section 2.2. Some examples make this definition clearer:

Example 2.1. A covector $\boldsymbol{\alpha} \in V^*$ is a $(1,0)$–tensor. In other words, $V^* = \mathcal{T}^{1,0}(V)$. Similarly, $\boldsymbol{v} \in V$ may be viewed as a $(0,1)$–tensor once it is seen as a double dual vector. More generally, a $(k,0)$–tensor is called *covariant*, while a $(0,\ell)$–tensor is called *contravariant.* With this in mind, we can see that we may interchange the dual parts of a $(1,0)$ or $(0,1)$ tensor: $\boldsymbol{v}(\boldsymbol{\alpha}) = \boldsymbol{\alpha}(\boldsymbol{v})$.

Example 2.2. Given an inner product, $\boldsymbol{g}(\boldsymbol{v}, \boldsymbol{u}) = \langle \boldsymbol{v}, \boldsymbol{u} \rangle$ is a $(2,0)$–tensor. In particular, the usual dot product $\boldsymbol{\delta}$ on $\mathbb{R}^n$ is as well. In an inner product space, given $\boldsymbol{v} \in V$ we can define a $(1,0)$-tensor $\boldsymbol{v}^*(\boldsymbol{w}) = g(\boldsymbol{v}, \boldsymbol{w})$, which is an identification of vectors with covectors (which requires the additional structure of the inner product).

Example 2.3. In $\mathbb{R}^n$, the function det is a $(n,0)-$tensor where $\det(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ is the determinant of the matrix with column vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$.

2.2 Einstein Notation

Due to the complexity of tensors in the size of their input, we will implement a notational system to simplify their use. We will develop more properties of tensors in parallel with this notation.

For now, choose a basis $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\} = \mathcal{B}$ for V and $\{\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^n\} = \mathcal{B}^*$ as a basis for V^* where

$$\boldsymbol{\lambda}^i(\boldsymbol{e}_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

The basis $\mathcal{B}^*$ is called the *projections* of $\mathcal{B}$. Consider

$$\boldsymbol{v} = \sum_{\mu=1}^{n} v^\mu \boldsymbol{e}_\mu \quad \boldsymbol{\alpha} = \sum_{\mu=1}^{n} \alpha^\mu \boldsymbol{\lambda}_\mu$$

where v^μ, α^μ are scalars for $1 \leqslant \mu \leqslant n$. We may express $\boldsymbol{v}, \boldsymbol{\alpha}$ as vectors in $\mathbb{R}^n$ as

$$[\boldsymbol{v}]_{\mathcal{B}} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \quad [\boldsymbol{\alpha}]_{\mathcal{B}^*} = [\alpha_1 \ldots \alpha_n]$$

So that then

$$\boldsymbol{\alpha}(\boldsymbol{v}) = [\alpha_1 \ldots \alpha_n] \cdot \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \sum_{\mu=1}^{n} \alpha_\mu v^\mu$$

A $(1,1)$–tensor takes in a vector and a covector and outputs a scalar. Since it is linear, we may represent $\boldsymbol{T}(\boldsymbol{\alpha}, \boldsymbol{v})$ by using the bases $\mathcal{B}, \mathcal{B}^*$ through an $n \times n$ matrix where we write the μ^{th} column, ν^{th} row as T^ν_μ. Then

$$T(\boldsymbol{\alpha}, \boldsymbol{v}) = [\boldsymbol{\alpha}]_{\mathcal{B}*} \cdot [T]_{\mathcal{B},\mathcal{B}*} \cdot [\boldsymbol{v}]_{\mathcal{B}} = \sum_{\mu,\nu=1}^{n} \alpha_\nu T^\nu_\mu v^\mu$$

To avoid having to use Σ frequently, the notation (known as Einstein Notation) will imply summation over matched lower and upper indices. That is,

$$\boldsymbol{T}(\boldsymbol{v}, \boldsymbol{\alpha}) = \alpha_\nu T^\nu_\mu v^\mu$$

In general, a (k,l)–tensor has k lower indices and l upper indices to match the k upper indices of the vectors and l lower indices of the covectors. This allows us to recognize 'multiplication' of vectors or covectors with themselves (in a given basis), resulting in a tensor.

$$\boldsymbol{vu} = v^\mu u^\nu$$

We wrote $\boldsymbol{T}(\boldsymbol{v}, \boldsymbol{\alpha}) = \alpha_\nu T^\nu_\mu v^\mu$, but this equality only holds if the expression $\alpha_\nu T^\nu_\mu v^\mu$ is fixed under a change of basis. Normally, under a change of basis, the components of a tensor $T^{\nu_1 \ldots \nu_\ell}_{\mu_1 \ldots \mu_k}$ must change.

Proposition. Given a change of basis matrix P and its inverse Q that map between $\mathcal{B}$ and $\tilde{\mathcal{B}}$, a component of a tensor expressed in $\mathcal{B}$ with k vectors must be multiplied by Q^k and with l covectors is multiplied by P^l to be expressed in $\tilde{\mathcal{B}}$.

This is easily verified by using linearity of the tensor and applying it to $P[\boldsymbol{v}]_{\mathcal{B}}, [\boldsymbol{\alpha}]_{\mathcal{B}*}P$. However, notice that if all indices are matched, $k = l$ so the net change is $P^k Q^l = I$, the identity. That is, a tensor expression with matched indices is coordinate independent.

From here on out, we may refer to unbolded expressions with an upper index (v^μ) as a vector, unbolded expressions with a lower index (α_μ) as a covector, expressions with matched indices as a number, and any other numbers of indices as tensors, where we assume we have fixed a basis (in M, we fix $\{\boldsymbol{e}_0, \ldots, \boldsymbol{e}_n\}$). To work with tensor expressions, we will develop a method of 'raising' or 'lowering' an index.

Assume we have an inner product $\langle \quad \rangle$ and as in example 2.2, write the inner product as a $(2,0)$–tensor $\boldsymbol{g}$. Given an orthonormal basis $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ of V, define an orthonormal basis $\{\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^n\}$ for V^* as the projections. If $\boldsymbol{v}, \boldsymbol{w}$ have components v^μ, w^ν, we find

$$\begin{aligned} \boldsymbol{g}(\boldsymbol{v}, \boldsymbol{w}) &= \boldsymbol{g}\left(\sum_{\mu=1}^{n} v^\mu \boldsymbol{e}_\mu, \sum_{\nu=1}^{n} w^\nu \boldsymbol{e}_\nu\right) \\ &= \sum_{\mu,\nu=1}^{n} v^\mu w^\nu \boldsymbol{g}(\boldsymbol{e}_\mu, \boldsymbol{e}_\nu) \\ &= \sum_{\mu,\nu} v^\mu w^\nu \langle \boldsymbol{e}_\mu, \boldsymbol{e}_\nu \rangle \end{aligned}$$

Since we have that

$$\langle \boldsymbol{e}_\mu, \boldsymbol{e}_\nu \rangle = \begin{cases} 1 & \mu = \nu \\ 0 & \mu \neq \nu \end{cases}$$

we can see that $\boldsymbol{g}$ may be expressed as

$$g_{\mu\nu} = \begin{cases} 1 & \mu = \nu \\ 0 & \mu \neq \nu \end{cases}$$

since then

$$\boldsymbol{g}(\boldsymbol{v}, \boldsymbol{w}) = g_{\mu\nu} v^\mu w^\nu$$

Using our basis, we can define $g^{\mu\nu}$ as the equivalent inner product on V^* by

$$g^{\mu\nu} = \begin{cases} 1 & \mu = \nu \\ 0 & \mu \neq \nu \end{cases}$$

Then we may define the 'raised' and 'lowered' forms by

$$v_\mu = g_{\mu\nu} v^\nu, \quad v^\mu = g^{\mu\nu} v_\nu$$

And thus $\boldsymbol{v}$ may be identified with a covector. This is equivalent to defining the dual of $\boldsymbol{v}$ as

$$\boldsymbol{v}^*(\boldsymbol{w}) = \langle \boldsymbol{v}, \boldsymbol{w} \rangle$$

Raising and lowering indices always occurs using a special tensor which for us here was g. In general where an inner product is not necessarily defined, the tensor to raise or lower an index is called the *metric*. The metric allows us to 'measure' tensors and convert between types, in that we may define a norm for tensors.

Definition 2.3. Suppose T is a (k, ℓ)–tensor, $\boldsymbol{g}$ is a metric. The *norm* of T, $\|T\|$, is the scalar given by

$$\sqrt{T^{\mu_1 \ldots \mu_k}_{\nu_1 \ldots \nu_\ell} T^{\nu_1 \ldots \nu_\ell}_{\mu_1 \ldots \mu_k}}$$

where

$$T^{\mu_1 \ldots \mu_k}_{\nu_1 \ldots \nu_\ell} = g^{\mu_1 \rho_1} \ldots g^{\mu_k \rho_k} g_{\nu_1 \lambda_1} \ldots g_{\nu_\ell \lambda_\ell} T^{\lambda_1 \ldots \lambda_\ell}_{\rho_1 \ldots \rho_k}$$

Example 2.4. The dot product $\boldsymbol{\delta}$ is a metric on M,

$$\delta_{\mu\nu} = \begin{cases} 1 & \mu = \nu \\ 0 & \mu \neq \nu \end{cases}$$

$\boldsymbol{\delta}$ is called the *Kronecker Delta*.

Although we will use the dot product as a tensor, we will see this is not the metric we want for M in section 2.4.

Given our new method of writing vectors of M ($\boldsymbol{x} = x^\mu \boldsymbol{e}_\mu$), we will now change our notation for the coordinates of M from

$$x_0, x_1, \ldots, x_n \text{ to } x^0, x^1, \ldots, x^n$$

where $t = x^0$ and $x^1, \ldots, x^n$ are spatial coordinates. A point of M will be written as x^μ.

2.3 The Wedge Product and Differential Forms

As tensors form vector spaces themselves, we would expect to be able to place operations on them similar to the vectors they act on. This may be done naturally through the product in $\mathbb{R}$ by the standard tensor product.

Definition 2.4. For a (k_1, l_1)–tensor $\boldsymbol{T}$ with inputs I_1 and a (k_2, l_2)–tensor S with inputs I_2 the *tensor product* produces a $(k_1 + k_2, l_1 + l_2)$–tensor, $\boldsymbol{T} \otimes \boldsymbol{S}$, defined as

$$(\boldsymbol{T} \otimes \boldsymbol{S})(I_1, I_2) = \boldsymbol{T}(I_1)\boldsymbol{S}(I_2)$$

This is simply the tensor that multiplies the result of the two other tensors, which we have been implicitly using in our discussion of Einstein notation. A more complex product is the wedge product, which constructs antisymmetric tensors. The wedge product is roughly a generalization of the cross product in $\mathbb{R}^3$, and closely related to the determinant. Just as the triple product measures spatial volume, we will see the wedge product has a geometric interpretation of measuring volume.

To define the wedge product, we first need to define a type of tensor dependent on permutations of the inputs. We have already seen an example of a tensor that depends on permutations, det.

Definition 2.5. Let S_m be the permutations of $\{1, \ldots, m\}$. For $\sigma \in S_m$, denote

$$\mathrm{sgn} : S_m \to \{-1, 1\}$$

as the sign of the permutation.[6] The *antisymmetric* or *alternating form* of a (k, ℓ)–tensor T is the tensor:

$$\Lambda(\boldsymbol{T})(\boldsymbol{v}_1 \ldots, \boldsymbol{v}_k, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_\ell) = \frac{1}{k!\ell!} \sum_{\sigma_k \in S_k, \sigma_\ell \in S_\ell} \mathrm{sgn}\sigma_k \cdot \mathrm{sgn}\sigma_\ell \cdot \boldsymbol{T}\left(\boldsymbol{v}_{\sigma_k(1)}, \ldots, \boldsymbol{v}_{\sigma_k(k)}, \boldsymbol{\alpha}_{\sigma_\ell(1)}, \ldots, \boldsymbol{\alpha}_{\sigma_\ell(\ell)}\right)$$

Antisymmetry refers to the property of a tensor where if two of the inputs are swapped (that is, there is an inversion of the indices), the sign of the output inverts. This is clearly given by the addition of $\mathrm{sgn}\sigma_k \cdot \mathrm{sgn}\sigma_\ell$ in the definition. For a $(2, 0)$–tensor $\boldsymbol{T}$, antisymmetry is represented by $\boldsymbol{T}(\boldsymbol{v}, \boldsymbol{w}) = -\boldsymbol{T}(\boldsymbol{w}, \boldsymbol{v})$, or with indices $\boldsymbol{T}^{\mu\nu} = -\boldsymbol{T}^{\nu\mu}$.

[6] $\mathrm{sgn}(\sigma) = 1$ iff there is an even number of inversions, that is, pairs $x < y \in \{1, \ldots, m\}$ such that $\sigma(x) > \sigma(y)$

Notice that the coefficient, $\frac{1}{k!\ell!}$, is the number of the additions made in the sum, since $|S_k \times S_\ell| = k! \cdot \ell!$. This coefficient 'normalizes' the alternating form of $\boldsymbol{T}$.

Definition 2.6. Let $\boldsymbol{T}_1, \boldsymbol{T}_2$ be $(k_1, \ell_1), (k_2, \ell_2)-$tensors respectively. The wedge product is defined as

$$\boldsymbol{T}_1 \wedge \boldsymbol{T}_2 = \frac{(k_1 + k_2)!(\ell_1 + \ell_2)!}{k_1! \cdot k_2! \cdot \ell_1! \cdot \ell_2!} \Lambda(\boldsymbol{T}_1 \otimes \boldsymbol{T}_2)$$

The wedge product is associative and linear, which are inherited from $\otimes$ and Λ. That is,

$$\begin{aligned} \boldsymbol{R} \wedge (\boldsymbol{S} \wedge \boldsymbol{T}) &= (\boldsymbol{R} \wedge \boldsymbol{S}) \wedge \boldsymbol{T} && \text{associative} \\ \boldsymbol{R} \wedge (k\boldsymbol{S} + \boldsymbol{T}) &= k\boldsymbol{R} \wedge \boldsymbol{S} + \boldsymbol{R} \wedge \boldsymbol{T} && \text{linear} \end{aligned}$$

Again, the coefficient in the wedge product is meant to normalize $\boldsymbol{T}_1 \wedge \boldsymbol{T}_2$ to be related to $\boldsymbol{T}_1 \otimes \boldsymbol{T}_2$. However, the sign differences of the permutations lead to significant differences in their size. One example of this is seen in the following proposition.

Proposition. For any vectors $\boldsymbol{v}_1, \boldsymbol{v}_2$ or covectors $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$,

$$\boldsymbol{v}_1 \wedge \boldsymbol{v}_2 = -\boldsymbol{v}_2 \wedge \boldsymbol{v}_1 \text{ and } \boldsymbol{\alpha}_1 \wedge \boldsymbol{\alpha}_2 = -\boldsymbol{\alpha}_1 \wedge \boldsymbol{\alpha}_2$$

Proof. We will prove the result for a covector, and the proof for the other case is similar. Ignoring coefficients,

$$\begin{aligned} \boldsymbol{\alpha}_1 \wedge \boldsymbol{\alpha}_2(\boldsymbol{v}_1, \boldsymbol{v}_2) &= \Lambda(\boldsymbol{\alpha}_1 \otimes \boldsymbol{\alpha}_2(\boldsymbol{v}_1, \boldsymbol{v}_2) \\ &= \sum_{\sigma_1, \sigma_2 \in S_2} \text{sgn}\sigma_1 \cdot \text{sgn}\sigma_2 \cdot \boldsymbol{\alpha}_1(\boldsymbol{v}_{\sigma_1(1)}) \cdot \boldsymbol{\alpha}_2(\boldsymbol{v}_{\sigma_2(2)}) \end{aligned}$$

There are two permutations in S_2, the identity id and the inversion $\sigma : 1 \mapsto 2, 2 \mapsto 1$. Notice sgn id $= 1$, sgn$\sigma = -1$. Thus we get

$$\begin{aligned} \boldsymbol{\alpha}_1 \wedge \boldsymbol{\alpha}_2(\boldsymbol{v}_1, \boldsymbol{v}_2) &= \boldsymbol{\alpha}_1(\boldsymbol{v}_1) \cdot \boldsymbol{\alpha}_2(\boldsymbol{v}_2) - \boldsymbol{\alpha}_2(\boldsymbol{v}_2) \cdot \boldsymbol{\alpha}_1(\boldsymbol{v}_1) \\ &= -(\boldsymbol{\alpha}_2(\boldsymbol{v}_1) \cdot \boldsymbol{\alpha}_1(\boldsymbol{v}_2) - \boldsymbol{\alpha}_1(\boldsymbol{v}_2) \cdot \boldsymbol{\alpha}_2(\boldsymbol{v}_1)) \\ &= -\boldsymbol{\alpha}_2 \wedge \boldsymbol{\alpha}_1(\boldsymbol{v}_1, \boldsymbol{v}_2) \end{aligned}$$

■

Notice that the previous proposition gives us:

$$\boldsymbol{v} \wedge \boldsymbol{v} = 0 \ , \ \boldsymbol{\alpha} \wedge \boldsymbol{\alpha} = 0$$

for any vector $\boldsymbol{v}$ and covector $\boldsymbol{\alpha}$.

We will see the application of the wedge product in the following statement.

Proposition. For an orthonormal basis $\mathcal{B} = \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ of V, the projections $\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^n$ in $\mathcal{B}^*$ satisfy

$$\det = \boldsymbol{\lambda}^1 \wedge \ldots \wedge \boldsymbol{\lambda}^n$$

where det is the $(n, 0)-$tensor defined by

$$\det(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n) = \det\left[[\boldsymbol{v}_1]_{\mathcal{B}}, \ldots, [\boldsymbol{v}_1]_{\mathcal{B}}\right]$$

where $[\boldsymbol{v}_i]_{\mathcal{B}} \in \mathbb{R}^n$ is a column vector equal to $\boldsymbol{v}_i$ expressed in terms of $\mathcal{B}$.

Proof. We will prove the case for $\mathbb{R}^3$, from which the general case may be proved through induction and expression of vector spaces through bases.

We want to compute $\boldsymbol{\lambda}^1 \wedge \boldsymbol{\lambda}^2 \wedge \boldsymbol{\lambda}^3$. Notice that the permutations of S_2 are those that map $(1,2)$ to:

$$\begin{matrix} (1,2) & \text{sgn} : + \\ (2,1) & \text{sgn} : - \end{matrix}$$

And the permutations of S_3 are those that map $(1,2,3)$ to:

$$\begin{matrix} (1,2,3) & \text{sgn} : + \\ (1,3,2) & \text{sgn} : - \\ (2,1,3) & \text{sgn} : - \\ (2,3,1) & \text{sgn} : + \\ (3,1,2) & \text{sgn} : + \\ (3,2,1) & \text{sgn} : - \end{matrix}$$

For the wedge $\lambda^2 \wedge \lambda^3$, we get a coefficient of

$$\frac{(1+1)!}{1!1!}\frac{1}{2!} = 1$$

and for $\boldsymbol{\lambda}^1 \wedge (\boldsymbol{\lambda}^2 \wedge \boldsymbol{\lambda}^3)$, we get

$$\frac{(1+2)!}{\cdot 1! \cdot 2!}\frac{1}{3!} = \frac{1}{2}$$

Then (writing $\otimes$ implicitly)

$$\begin{aligned}
\boldsymbol{\lambda}^1 \wedge \boldsymbol{\lambda}^2 \wedge \boldsymbol{\lambda}^3(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}) &= \boldsymbol{\lambda}^1 \wedge \Lambda(\boldsymbol{\lambda}^2\boldsymbol{\lambda}^3) \\
&= \boldsymbol{\lambda}^1 \wedge \left(\sum_{\sigma\in S_2} \text{sgn}\sigma \cdot \boldsymbol{\lambda}^{\sigma(2)}\boldsymbol{\lambda}^{\sigma(3)}\right) \\
&= \boldsymbol{\lambda}^1 \wedge \left(\boldsymbol{\lambda}^2\boldsymbol{\lambda}^3 - \boldsymbol{\lambda}^3\boldsymbol{\lambda}^2\right) \\
&= \frac{1}{2}\sum_{\sigma\in S_3} \text{sgn}\sigma \cdot \boldsymbol{\lambda}^{\sigma(1)}\boldsymbol{\lambda}^{\sigma(2)}\boldsymbol{\lambda}^{\sigma(3)} - \text{sgn}\sigma\boldsymbol{\lambda}^{\sigma(1)}\boldsymbol{\lambda}^{\sigma(3)}\boldsymbol{\lambda}^{\sigma(2)} \\
&= \frac{1}{2}((\boldsymbol{\lambda}^1\boldsymbol{\lambda}^2\boldsymbol{\lambda}^3 - \boldsymbol{\lambda}^1\boldsymbol{\lambda}^3\boldsymbol{\lambda}^2 - \boldsymbol{\lambda}^2\boldsymbol{\lambda}^1\boldsymbol{\lambda}^3 + \boldsymbol{\lambda}^2\boldsymbol{\lambda}^3\boldsymbol{\lambda}^1 + \boldsymbol{\lambda}^3\boldsymbol{\lambda}^1\boldsymbol{\lambda}^2 - \boldsymbol{\lambda}^3\boldsymbol{\lambda}^2\boldsymbol{\lambda}^1) \\
&\quad - (\boldsymbol{\lambda}^1\boldsymbol{\lambda}^3\boldsymbol{\lambda}^2 - \boldsymbol{\lambda}^1\boldsymbol{\lambda}^2\boldsymbol{\lambda}^3 - \boldsymbol{\lambda}^3\boldsymbol{\lambda}^1\boldsymbol{\lambda}^2 + \boldsymbol{\lambda}^3\boldsymbol{\lambda}^2\boldsymbol{\lambda}^1 + \boldsymbol{\lambda}^2\boldsymbol{\lambda}^1\boldsymbol{\lambda}^3 - \boldsymbol{\lambda}^2\boldsymbol{\lambda}^3\boldsymbol{\lambda}^1)) \\
&= \boldsymbol{\lambda}^1\boldsymbol{\lambda}^2\boldsymbol{\lambda}^3 - \boldsymbol{\lambda}^1\boldsymbol{\lambda}^3\boldsymbol{\lambda}^2 - \boldsymbol{\lambda}^2\boldsymbol{\lambda}^1\boldsymbol{\lambda}^3 + \boldsymbol{\lambda}^2\boldsymbol{\lambda}^3\boldsymbol{\lambda}^1 + \boldsymbol{\lambda}^3\boldsymbol{\lambda}^1\boldsymbol{\lambda}^2 - \boldsymbol{\lambda}^3\boldsymbol{\lambda}^2\boldsymbol{\lambda}^1
\end{aligned}$$

Applying this to the column vectors of

$$A = \begin{bmatrix} a^1 & b^1 & c^1 \\ a^2 & b^2 & c^2 \\ a^3 & b^3 & c^3 \end{bmatrix} = [\boldsymbol{a}\ \boldsymbol{b}\ \boldsymbol{c}]$$

we get

$$a^1b^2c^3 - a^1b^3c^2 - a^2b^1c^3 + a^2b^3c^1 + a^3b^1c^2 - a^3b^2c^1$$

The determinant of A is

$$\det \begin{bmatrix} a^1 & b^1 & c^1 \\ a^2 & b^2 & c^2 \\ a^3 & b^3 & c^3 \end{bmatrix} = a^1 \det \begin{bmatrix} b^2 & c^2 \\ b^3 & c^3 \end{bmatrix} - a^2 \det \begin{bmatrix} b^1 & c^1 \\ b^3 & c^3 \end{bmatrix} + a^3 \det \begin{bmatrix} b^1 & c^1 \\ b^2 & c^2 \end{bmatrix}$$
$$= a^1 b^2 c^3 - a^1 b^3 c^2 - a^2 b^1 c^3 + a^2 b^3 c^1 + a^3 b^1 c^2 - a^3 b^2 c^1$$

Thus

$$\det A = \boldsymbol{\lambda}^1 \wedge \boldsymbol{\lambda}^2 \wedge \boldsymbol{\lambda}^3(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$$

■

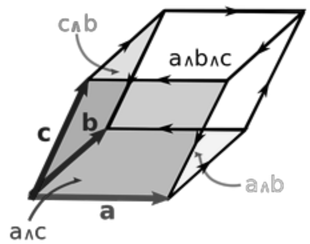

Figure 5: The wedge product as a volume unit
Image credit:
`https://commons.wikimedia.org/wiki/File:Exterior_calc_triple_product.svg`

In $\mathbb{R}^n$, the wedge product of n vectors viewed as $(0,1)$–tensors or n covectors is naturally identifiable with a scalar. For vectors $\boldsymbol{v}_1, \ldots \boldsymbol{v}_n$, the scalar is

$$(\boldsymbol{v}_1 \wedge \ldots \wedge \boldsymbol{v}_n)(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_n) = \boldsymbol{\lambda}_1 \wedge \ldots \wedge \boldsymbol{\lambda}_n(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n) = \det[\boldsymbol{v}_1 \ldots \boldsymbol{v}_n]$$

Geometrically, the determinant of a matrix of column vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ measures the space described by $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$. Using the relationship between det and the wedge product, define the volume of the hyper-parallelepiped described by n vectors to be the wedge product of the projections on these vectors. The three-dimensional parallelepiped is displayed in Figure 5. If our vector space is tangent to M at p ($V = T_pM$) then the hypercube $\boldsymbol{\lambda}^0 \wedge \ldots \wedge \boldsymbol{\lambda}^n(\boldsymbol{e}_0, \ldots, \boldsymbol{e}_n)$ may be thought of as an 'infinitesimal volume unit' for M. We write the volume unit as $d\boldsymbol{x} = dx^0 \wedge \ldots \wedge dx^n = \boldsymbol{\lambda}^0 \wedge \ldots \wedge \boldsymbol{\lambda}^n$ where $\boldsymbol{x}$ are the coordinates of M.

Definition 2.7. If $\boldsymbol{x}$ are the coordinates of M, then $d\boldsymbol{x}$ is called a *differential form*. For $p \in M$, the object $dx^\mu \in T_p^*M$[7] is the projection of $\boldsymbol{e}_\mu \in T_pM$.

We would like to extend this definition of 'd' to an arbitrary function $f : M \to \mathbb{R}$.

Definition 2.8. Let $f : M \to \mathbb{R}$ be a smooth function. The *exterior derivative* of f at p is the function $df \in T_p^*M$ (that is, $df : T_pM \to \mathbb{R}$ linear) where

$$df(\boldsymbol{v}) = \frac{d}{dt} f(p + t\boldsymbol{v})\bigg|_{t=0}$$

Notice that the definition of df is independent of the choice in coordinates. However, we will give a way to compute df given a coordinate system $\boldsymbol{x}$.

[7] $T_p^*M = (T_pM)^*$

Proposition. Let $f : M \to \mathbb{R}$ be smooth. Then

$$df = \partial_\mu f dx^\mu = \frac{\partial f}{\partial x^0} dx^0 + \ldots + \frac{\partial f}{\partial x^n} dx^n$$

Proof. Choose the coordinates $\boldsymbol{x} = (x^0, \ldots, x^n)$ for M. Then

$$\begin{aligned}
df(\boldsymbol{v}) &= \frac{d}{dt} f(p + t\boldsymbol{v}) \Big|_{t=0} \\
&= \frac{\partial f}{\partial x_0} \frac{d(p + tv^0)}{dt} \Big|_{t=0} + \ldots + \frac{\partial f}{\partial x_n} \frac{d(p + tv^n)}{dt} \Big|_{t=0} && \text{chain rule} \\
&= \frac{\partial f}{\partial x_0} v^0 + \ldots + \frac{\partial f}{\partial x_0} v^n \\
&= \frac{\partial f}{\partial x_0} dx^0(\boldsymbol{v}) + \ldots + \frac{\partial f}{\partial x_0} dx^n(\boldsymbol{v}) \\
&= \left(\frac{\partial f}{\partial x_0} dx^0 + \ldots + \frac{\partial f}{\partial x_0} dx^n \right) (\boldsymbol{v})
\end{aligned}$$

■

To be able to integrate a function $f(\boldsymbol{x})$, we need to multiply it by the differential form $d\boldsymbol{x}$. Notationally, we will drop the $\wedge$ when writing expanded differential forms as a part of an integral, such as $dxdydz$. As we will see later, differential forms as a part of integrands becomes important in computing the derivatives of integrals where not only the function, but also the differential form are varying.

2.4 Lagrangians involving Tensors

As a Lagrangian is some expression of functions and coordinates in a space, we would expect many Lagrangians to be expressible as tensors. We are specifically examining Lagrangians that express quantities in the physical world, and so must acknowledge some differences between physical spacetime and the traditional $\mathbb{R}^{n+1}$ vector space.

Definition 2.9. The Minkowski metric $\boldsymbol{\eta}$ is the $(2, 0)$–tensor that has the property

$$\eta^{\mu\nu} = \eta_{\mu\nu} = \begin{cases} -1 & \mu = \nu = 0 \\ 1 & \mu = \nu \neq 0 \\ 0 & \mu \neq \nu \end{cases}$$

in the standard basis $\{\boldsymbol{e}_0, \ldots, \boldsymbol{e}_n\}$.

Relativistic physical spacetime may be modeled as $\mathbb{R}^{n+1}$ equipped with $\boldsymbol{\eta}$. [8] This is called Minkowski space, and is the space we have been referring to as M. By 'equipped,' we mean that to raise or lower indices in M, we must use $\boldsymbol{\eta}$. Alternatively, one may view $\boldsymbol{\eta}$ as representing the intrinsic geometry of M. For expressions involving tensors in spacetime, the derivative should be compatible with raising or lowering these indices.

Since the derivative is a linear operator, it may be written as a tensor on the space of functions as ∂. That is,

$$\partial_\mu = \frac{\partial}{\partial x^\mu}$$

[8] In relativity, Lorentz Transforms are mappings between reference frames. These mappings should be isometric or geometry-preserving, but the metric they preserve is not the dot product $\boldsymbol{\delta}$, but instead $\boldsymbol{\eta}$.

We use this tensor to define a derivative on functions of covectors as

$$\partial^\nu = \frac{\partial}{\partial x_\nu} = \eta^{\mu\nu}\partial_\mu$$

Now that we have the derivative as an expression in tensors, we may rewrite the Euler-Lagrange equations as a tensor equation.

Proposition. The Euler-Lagrange Equations

$$\frac{\partial \mathcal{L}}{\partial f_i} - \sum_{j=1}^{n} \frac{\partial}{\partial x_j} \frac{\partial \mathcal{L}}{\partial \left(\frac{\partial f^i}{\partial x_j}\right)} = 0, \quad 1 \leqslant i \leqslant k$$

are equivalent to

$$\frac{\partial \mathcal{L}}{\partial f^a} + \partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu f^a)} = 0, \quad 1 \leqslant a \leqslant k$$

We use a in the above equations instead of i to avoid any potential confusion if we work with complex coordinates. These new tools allow us to express the problems of the previous section in a more generalizable format. Consider the same physical examples as the previous section: the Wave-Lagrangian $\mathcal{W}$, and the Schrödinger Lagrangian $\mathcal{S}$.

Example 2.5. For the Wave Lagrangian, we may write

$$\mathcal{W}(\varphi) = \|\nabla\varphi\|^2 - \dot{\varphi}^2$$

in the tensor form

$$\mathcal{W}(\varphi) = (\partial^\mu\varphi)(\partial_\mu\varphi)$$

To confirm this, compute

$$\begin{aligned} \partial^\mu\varphi \cdot \partial_\mu\varphi &= \eta_{\mu\nu}\partial^\mu\varphi \cdot \partial^\nu\varphi \\ &= -\left(\frac{\partial}{\partial t}\varphi\right)^2 + \sum_{\mu=1}^{n}\left(\frac{\partial}{\partial x^\mu}\varphi\right)^2 \\ &= \|\nabla\varphi\|^2 - \dot{\varphi}^2 \end{aligned}$$

Example 2.6. We will now introduce one more example of a physical Lagrangian which is difficult to handle without tensors. In 4-dimensional spacetime, consider the electric and magnetic fields $\boldsymbol{E}, \boldsymbol{B} : M \to \mathbb{R}^3$. Assume we may write these fields in terms of two functions:

$$\begin{aligned} \boldsymbol{E} &= -\nabla\phi - \frac{\partial \boldsymbol{A}}{\partial t} \\ \boldsymbol{B} &= \nabla \times \boldsymbol{A} \end{aligned}$$

where $\phi : M \to \mathbb{R}$ is called the electric potential, $\boldsymbol{A} : M \to \mathbb{R}^3$ is known as the vector potential, and $\nabla\times$ is the curl:

$$\nabla \times \boldsymbol{X} = \left(\frac{\partial X_z}{\partial y} - \frac{\partial X_y}{\partial z}\right)\boldsymbol{e}_1 - \left(\frac{\partial X_x}{\partial z} - \frac{\partial X_z}{\partial x}\right)\boldsymbol{e}_2 + \left(\frac{\partial X_y}{\partial x} - \frac{\partial X_x}{\partial y}\right)\boldsymbol{e}_3$$

Definition 2.10. The (vacuum) Electromagnetic Lagrangian is defined as

$$\mathcal{E} = \|\boldsymbol{B}\|^2 - \|\boldsymbol{E}\|^2$$

Although we are able to express this Lagrangian without tensors, it is difficult to find useful results without them. To find the equivalent statement in terms of tensors, define the 'complete potential' as the 4-vector

$$A^\mu = \begin{cases} \phi & \mu = 0 \\ A_{x^\mu} & \mu > 0 \end{cases}$$

Definition 2.11. The electromagnetic tensor $\boldsymbol{F}$ is defined as

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$$

Abusing notation, we could write this as $\partial \wedge A$.

Proposition. As a matrix where μ indexes the columns and ν,

$$F^{\mu\nu} = \begin{bmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -B_z & B_y \\ E_y & B_z & 0 & -B_x \\ E_z & -B_y & B_x & 0 \end{bmatrix}$$

where subscript denotes partial derivative. As expected, $F^{\mu\nu} = -F^{\nu\mu}$ since it is an anti-symmetric tensor.

To see the computation of this matrix, first consider the diagonal. On the diagonal, $\mu = \nu$ so

$$F^{\mu\nu} = F^{\mu\mu} = \partial^\mu A^\mu - \partial^\mu A^\mu = 0$$

We will compute F^{01} as an example, and the other components may be done similarly.

$$\begin{aligned} F^{01} &= \partial^0 A^1 - \partial^1 A^0 \\ &= \eta^{0\nu}\partial_0 A^1 - \eta^{1\nu}\partial_1 A^0 \\ &= -\frac{\partial}{\partial t}\boldsymbol{A}_x - \frac{\partial}{\partial x}\phi \\ &= \boldsymbol{E}_x \end{aligned}$$

To get the dual of $F^{\mu\nu}$, we use η so

$$F_{\mu\nu} = \eta_{\mu\rho}\eta_{\nu\lambda}F^{\rho\lambda}$$

Proposition. If $\boldsymbol{E}, \boldsymbol{B}$ may be written in terms of A^μ,

$$\|\boldsymbol{F}\|^2 = F^{\mu\nu}F_{\mu\nu} = 2\partial_\mu A_\nu F^{\mu\nu} = 2\mathcal{E}$$

Proof.

$$\begin{aligned}
F_{\mu\nu}F^{\mu\nu} &= \eta_{\mu\rho}\eta_{\nu\lambda}F^{\rho\lambda}F^{\mu\nu} \\
&= \eta_{\mu\rho}\eta_{\nu\lambda}(\partial^\rho A^\lambda - \partial^\lambda A^\rho)(\partial^\mu A^\nu - \partial^\nu A^\mu) \\
&= (\eta_{\mu\rho}\partial^\rho(\eta_{\nu\lambda}A^\lambda) - \eta_{\nu\lambda}\partial^\lambda(\eta_{\mu\rho}A^\rho))(\partial^\mu A^\nu - \partial^\nu A^\mu) \\
&= (\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu) \\
&= \partial_\mu A_\nu \partial^\mu A^\nu - \partial_\nu A_\mu \partial^\mu A^\nu - \partial_\mu A_\nu \partial^\nu A^\mu + \partial_\nu A_\mu \partial^\nu A^\mu \\
&= \partial_\mu A_\nu \partial^\mu A^\nu - \partial_\mu A_\nu \partial^\nu A^\mu - \partial_\mu A_\nu \partial^\nu A^\mu + \partial_\mu A_\nu \partial^\mu A^\nu \qquad \text{switch labels} \\
&= 2(\partial_\mu A_\nu \partial^\mu A^\nu - \partial_\mu A_\nu \partial^\nu A^\mu) \\
&= 2\partial_\mu A_\nu(\partial^\mu A^\nu - \partial^\nu A^\mu) \\
&= 2\partial_\mu A_\nu F^{\mu\nu}
\end{aligned}$$

So we have proved the second equality. For the third,

$$\begin{aligned}
\partial_\mu A_\nu F^{\mu\nu} &= \sum_{i=1}^{3} \left(\underbrace{\frac{\partial}{\partial t}A_{x_i}}_{\mu=0} - \underbrace{\frac{\partial}{\partial x_i}\phi}_{\nu=0}\right)\boldsymbol{E}_{x_i} + \underbrace{\boldsymbol{B}_x\left(\frac{\partial}{\partial y}A_z - \frac{\partial}{\partial y}A_z\right)}_{\mu,\nu=2,3} \\
&\quad \underbrace{-\boldsymbol{B}_y\left(\frac{\partial}{\partial x}A_z - \frac{\partial}{\partial z}A_x\right)}_{\mu,\nu=1,3} + \underbrace{\boldsymbol{B}_z\left(\frac{\partial}{\partial x}A_y - \frac{\partial}{\partial y}A_x\right)}_{\mu,\nu=1,2} \\
&= \sum_{i=1}^{3} -\boldsymbol{E}_{x_i}^2 + \boldsymbol{B}_{x_i}^2 \\
&= |\boldsymbol{B}|^2 - |\boldsymbol{E}|^2 \\
&= B^2 - E^2
\end{aligned}$$

■

Notice that for Lagrangians $\mathcal{L}$ and $\tilde{\mathcal{L}} = \lambda\mathcal{L}$ for some scalar λ, f is a solution for $\mathcal{L}$ iff f is a solution for $\tilde{\mathcal{L}}$ since we are setting these expressions equal to 0. Thus scalars are irrelevant, and so we may think of $\|F\|$ as being equivalent to $B^2 - E^2$.

Proposition. $F^{\mu\nu}$ satisfies the Bianchi Identity:

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

Proof.

$$\begin{aligned}
\partial^\rho F^{\mu\nu} + \partial^\mu F^{\nu\rho} + \partial^\nu F^{\rho\mu} &= \partial^\rho(\partial^\mu A^\nu - \partial^\nu A^\mu) + \partial^\mu(\partial^\nu A^\rho - \partial^\rho A^\nu) + \partial^\nu(\partial^\rho A^\mu - \partial^\mu A^\rho) \\
&= (\partial^\rho\partial^\mu A^\nu - \partial^\mu\partial^\rho A^\nu) + (\partial^\rho\partial^\nu A^\mu - \partial^\nu\partial^\rho A^\mu) + (\partial^\mu\partial^\nu A^\rho - \partial^\nu\partial^\mu A^\rho) \\
&= 0
\end{aligned}$$

■

We will use the Bianchi Identity in parallel with the Euler-Lagrange equations.

Now we may apply the Euler-Lagrange equations for tensors to our Lagrangian.

Theorem 2.1. If $\boldsymbol{E}, \boldsymbol{B}$ may be written in terms of potentials and is a solution for the Electromagnetic Lagrangian then they satisfy all of Maxwell's Equations (in vacuum):

$$\nabla \cdot \boldsymbol{B} = 0 \qquad M1$$
$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \qquad M2$$
$$\nabla \cdot \boldsymbol{E} = 0 \qquad M3$$
$$\nabla \times \boldsymbol{B} = \frac{\boldsymbol{E}}{\partial t} \qquad M4$$

Proof. Assume $\boldsymbol{E}, \boldsymbol{B}$ may be written in terms of potentials, that is,

$$\boldsymbol{E} = -\nabla \phi - \frac{\partial \boldsymbol{A}}{\partial t}$$
$$\boldsymbol{B} = \nabla \times \boldsymbol{A}$$

and are solutions for $\mathcal{E}$.

Consider $\rho = 1, \mu = 2, \nu = 3$. Then the Bianchi Identity yields

$$\begin{aligned} &\partial^1 F^{23} + \partial^2 F^{31} + \partial^3 F^{12} = 0 \\ \implies &\frac{\partial}{\partial x}\boldsymbol{B}_x + \frac{\partial}{\partial y}\boldsymbol{B}_y + \frac{\partial}{\partial z}\boldsymbol{B}_z = 0 \\ \implies &\nabla \cdot \mathbf{B} = 0 \qquad M1 \end{aligned}$$

Consider the Bianchi Identity over all choices of $\{0,1,2,3\}$ for indices ρ, μ, ν.For example, choosing $(0,2,3), (0,3,1), (0,1,2)$ respectively and summing we get

$$\begin{aligned} &\partial^0 F^{23} + \partial^2 F^{30} + \partial^3 F^{02} + \partial^0 F^{31} + \partial^1 F^{30} + \partial^3 F^{01} + \partial^0 F^{12} + \partial^1 F^{20} + \partial^2 F^{01} = 0 + 0 + 0 \\ \implies &\frac{\partial}{\partial t}B_x - \frac{\partial}{\partial y}E_z + \frac{\partial}{\partial z}E_y + \frac{\partial}{\partial t}B_y - \frac{\partial}{\partial x}E_z + \frac{\partial}{\partial z}E_x + \frac{\partial}{\partial t}B_z - \frac{\partial}{\partial x}E_y + \frac{\partial}{\partial y}E_x = 0 \\ \implies &\underbrace{\frac{\partial}{\partial t}B_x + \frac{\partial}{\partial t}B_y + \frac{\partial}{\partial t}B_z}_{\frac{\partial}{\partial t}B} + \underbrace{\left(\frac{\partial}{\partial z}E_y - \frac{\partial}{\partial y}E_z\right) + \left(\frac{\partial}{\partial z}E_x - \frac{\partial}{\partial x}E_z\right) + \left(\frac{\partial}{\partial y}E_x - \frac{\partial}{\partial x}E_y\right)}_{\nabla \times E} = 0 \end{aligned}$$

Keeping track of the components represented by the indices, we see that this expression is a $(0,3)$-tensor that projects down the x, y, z dual components in such a way that

$$\nabla \times \mathbf{E} = -\frac{\partial}{\partial t}\mathbf{B} \qquad M2$$

Now we will compute the Euler-Lagrange equations for $\mathcal{E}$. Notice that when taking a derivative with respect to a vector, we may write

$$\frac{\partial}{\partial(\partial_\sigma A_\tau)}(\partial_\mu A_\nu) = \delta^\sigma_\mu \delta^\tau_\nu$$

where $\boldsymbol{\delta}$ is the Kronecker Delta

$$\delta^\mu_\nu = \begin{cases} 1 & \mu = \nu \\ 0 & \mu \neq \nu \end{cases}$$

By the proposition, $\mathcal{E} = 2\partial_\mu A_\nu(\partial^\mu A^\nu - \partial^\nu A^\mu)$. Thus we may find:

$$\begin{aligned}
\frac{\partial \mathcal{E}}{\partial A^\tau} &= 0 \\
\partial_\sigma \frac{\mathcal{E}}{\partial(\partial_\sigma A_\tau)} &= \partial_\sigma \left(\frac{\partial}{\partial(\partial_\sigma A_\tau)} \left(2\partial_\mu A_\nu(\partial^\mu A^\nu - \partial^\nu A^\mu)\right)\right) \\
&= 2\partial_\sigma \left(\frac{\partial}{\partial(\partial_\sigma A_\tau)} \left(\partial_\mu A_\nu(\eta^{\mu\rho}\eta^{\nu\lambda}\partial_\rho A_\lambda - \eta^{\nu\lambda}\eta^{\mu\rho}\partial_\lambda A_\rho)\right)\right) \\
&= 2\eta^{\mu\rho}\eta^{\nu\lambda}\partial_\sigma \left(\frac{\partial}{\partial(\partial_\sigma A_\tau)} \left(\partial_\mu A_\nu \partial_\rho A_\lambda - \partial_\mu A_\nu \partial_\lambda A_\rho\right)\right) \\
&= 2\eta^{\mu\rho}\eta^{\nu\lambda}\partial_\sigma \left(\delta^\sigma_\mu \delta^\tau_\nu \partial_\rho A_\lambda + \delta^\sigma_\rho \delta^\tau_\lambda \partial_\mu A_\nu - (\delta^\sigma_\mu \delta^\tau_\nu \partial_\lambda A_\rho + \delta^\sigma_\lambda \delta^\tau_\rho \partial_\mu A_\lambda)\right) \\
&= 2\partial_\sigma \left(\eta^{\mu\rho}\eta^{\nu\lambda}\delta^\sigma_\mu \delta^\tau_\nu \partial_\rho A_\lambda + \eta^{\mu\rho}\eta^{\nu\lambda}\delta^\sigma_\rho \delta^\tau_\lambda \partial_\mu A_\nu - \eta^{\mu\rho}\eta^{\nu\lambda}\delta^\sigma_\mu \delta^\tau_\nu \partial_\lambda A_\rho - \eta^{\mu\rho}\eta^{\nu\lambda}\delta^\sigma_\lambda \delta^\tau_\rho \partial_\mu A_\lambda\right) \\
&= 2\partial_\sigma \left(\delta^\sigma_\mu \delta^\tau_\nu \partial^\mu A^\nu + \delta^\sigma_\rho \delta^\tau_\lambda \partial^\rho A^\lambda - \delta^\sigma_\mu \delta^\tau_\nu \partial^\nu A^\mu - \delta^\sigma_\lambda \delta^\tau_\rho \partial^\rho A^\lambda\right) \\
&= 2\partial_\sigma \left(\partial^\sigma A^\tau + \partial^\sigma A^\tau - \partial^\tau A^\sigma - \partial^\tau A^\sigma\right) \\
&= 4\partial_\sigma \left(\partial^\sigma A^\tau - \partial^\tau A^\sigma\right) \\
&= 4\partial_\sigma F^{\sigma\tau}
\end{aligned}$$

So the Euler-Lagrange equation is

$$\partial_\mu F^{\mu\nu} = 0$$

Thus applying $\nu = 0$ we get

$$\begin{aligned}
&-\frac{\partial}{\partial x}\mathbf{E}_x - \frac{\partial}{\partial y}\mathbf{E}_y - \frac{\partial}{\partial z}\mathbf{E}_z = 0 \\
&\implies \nabla \cdot \mathbf{E} = \mathbf{0} \qquad\qquad M3
\end{aligned}$$

For $\nu = 1, 2, 3$, we get

$$\frac{\partial}{\partial t}\mathbf{E}_x - \frac{\partial}{\partial y}\mathbf{B}_z + \frac{\partial}{\partial z}\mathbf{B}_y = 0 \tag{1}$$

$$\frac{\partial}{\partial t}\mathbf{E}_y + \frac{\partial}{\partial x}\mathbf{B}_z - \frac{\partial}{\partial z}\mathbf{B}_x = 0 \tag{2}$$

$$\frac{\partial}{\partial t}\mathbf{E}_z - \frac{\partial}{\partial x}\mathbf{B}_y + \frac{\partial}{\partial y}\mathbf{B}_x = 0 \tag{3}$$

Adding 1, 2, and 3 together, we get

$$0 = \underbrace{\left(\frac{\partial}{\partial t}\mathbf{E}_x + \frac{\partial}{\partial t}\mathbf{E}_y + \frac{\partial}{\partial t}\mathbf{E}_z\right)}_{\frac{\partial \mathbf{E}}{\partial t}} + - \underbrace{\left(\left(\frac{\partial \mathbf{B}_z}{\partial y} - \frac{\partial \mathbf{B}_y}{\partial z}\right) + \left(\frac{\partial \mathbf{B}_x}{\partial z} - \frac{\partial \mathbf{B}_z}{\partial x}\right) + \left(\frac{\partial \mathbf{B}_y}{\partial \mathbf{B}_x} - \frac{\partial \mathbf{B}_x}{\partial y}\right)\right)}_{\nabla \times \mathbf{B}}$$

Again, by the vector components represented by the indices we get

$$\nabla \times \mathbf{B} = \frac{\partial \mathbf{E}}{\partial t} \qquad M4$$

■

So far with the Electromagnetic Lagrangian $\mathcal{E}$, we have been considering the vacuum Lagrangian, that is, assuming there is no electric charge present. Now, consider $\rho : M \to \mathbb{R}$ as the charge density and $\boldsymbol{J} : M \to \mathbb{R}^3$ as the current density. Similarly as with A^μ, define

$$J^\mu = \begin{cases} \rho & \mu = 0 \\ \boldsymbol{J}_{x^\mu} & \mu > 0 \end{cases}$$

J^μ is called the *4-current*.

Now extend our Lagrangian:

Definition 2.12. The (complete) Electromagnetic Lagrangian is given by

$$\mathcal{E} = F^{\mu\nu} F_{\mu\nu} - 4A_\mu J^\mu$$

The coefficient of 4 in $\mathcal{E}$ comes from the definition of $F^{\mu\nu}$ as the antisymmetric tensor '$2\partial \wedge A$' not being normalized, so we are left with an extra factor of $2 \cdot 2 = 4$.

$M1, M2$ are the same since the properties of A^μ have not changed. By the work we did previously and since ∂ is linear, the Euler-Lagrange equations give us

$$\begin{aligned} 0 &= \frac{\partial \mathcal{E}}{\partial A^\nu} - \partial_\sigma \frac{\partial \mathcal{E}}{\partial(\partial_\sigma A_\nu)} \\ \partial_\mu F^{\mu\nu} &= \frac{\partial(A_\mu J^\mu)}{\partial A^\nu} - \partial_\sigma \frac{\partial A_\mu J^\mu}{\partial(\partial_\sigma A_\nu)} \\ \partial_\mu F^{\mu\nu} &= J^\nu \end{aligned}$$

Apply a similar derivation as for $M3, M4$ in vacuum applied to this equation to find (the non-vacuum) $M3, M4$ to be

$$\begin{aligned} \nabla \cdot \boldsymbol{E} &= \rho & M3 \\ \nabla \times \boldsymbol{B} &= \frac{\partial \boldsymbol{E}}{\partial t} + \boldsymbol{J} & M4 \end{aligned}$$

Theorem 2.2. If $\boldsymbol{E}, \boldsymbol{B}$ may be expressed in terms of potentials and are a solution for $\mathcal{E}$ for a fixed J^μ, then Maxwell's Equations hold:

$$\begin{aligned}
\nabla \cdot \boldsymbol{B} &= 0 && M1 \\
\nabla \times \boldsymbol{E} &= -\frac{\partial \boldsymbol{B}}{\partial t} && M2 \\
\nabla \cdot \boldsymbol{E} &= \rho && M3 \\
\nabla \times \boldsymbol{B} &= \frac{\partial \boldsymbol{E}}{\partial t} + \boldsymbol{J} && M4
\end{aligned}$$

Notice that the Euler-Lagrange equation for electromagnetism is a vector equation

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

Find a scalar equation (matching the indices) by taking the (spacetime) divergence:

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

Notice that if we switch the order of μ, ν in F, we also find

$$-\partial_\nu \partial_\mu F^{\nu\mu} = \partial_\nu J^\nu$$

This is only possible if

$$\partial_\nu \partial_\mu F^{\mu\nu} = 0$$

and so

$$\partial_\nu J^\nu = 0$$

This equation, that the spacetime divergence of a 4-current is 0, is an important notion that we will develop in the next section.

2.5 Conserved Currents

Consider the equation $\partial_\mu j^\mu = 0$. This is the divergence of the vector field j^μ on M. In terms of the components, this is

$$0 = \frac{\partial}{\partial t} j^0 + \sum_{k=1}^{n} \frac{\partial}{\partial x^k} j^k$$

Defining $\rho = j^0$ and $\boldsymbol{j} \in \mathbb{R}^n$ as the spatial components of j, we get

$$\frac{d}{dt}\rho = -\nabla \cdot \boldsymbol{j}$$

Think of ρ as a density of some quantity Q, that is,

$$Q = \int_\Omega \rho \, dV$$

Then it must be that $\boldsymbol{j}$ is the 'flow' of the Q. To see this, integrate both sides over space and apply the divergence theorem to see

$$\begin{aligned}\frac{d}{dt}Q &= \int\int\int_{\Omega} \nabla \cdot \boldsymbol{j} dV \\ &= -\int\int_{\partial\Omega} \boldsymbol{j} \cdot d\boldsymbol{S}\end{aligned}$$

That is, the change of Q over time exactly the matches the outward flux of $\boldsymbol{j}$. This means that Q must be conserved over time, as any gained or lost amounts must have been flowing into or out of the region. In general, Q is called a *Noether charge*, $\boldsymbol{j}$ is called a *Noether current*, and j^μ is called a *conserved current*.

Example 2.7. As we saw in the previous example for electromagnetism, we have a conserved current J^μ. The Noether charge is the standard electrical charge, and the Noether current is electromagnetic current. By our discussion above, we have the following:

Theorem 2.3. If $\boldsymbol{E}, \boldsymbol{B}$ may be written in terms of potentials and are a solution for the Electromagnetic Lagrangian $\mathcal{E}$ with a fixed 4-current J^μ, then electric charge is conserved.

3 Statement and Proof of Noether's Theorem

3.1 Proof of the Theorem

With the tools we have developed in the past two sections, we are now ready to state and prove Noether's theorem. Recall the statement of the theorem:

Theorem. A smooth symmetry on a Lagrangian corresponds to a quantity conserved in time, or a conserved current.

Let us rephrase this in terms of our new notation. This gives a more explicit and precise phrasing of Noether's Theorem.

Theorem 3.1. Suppose we have a Lagrangian $\mathcal{L} : M \to \mathbb{R}$ that depends on coordinates x^μ, functions φ^a of x^μ, and the first partials of φ:

$$\mathcal{L}\left(x^\mu, \varphi^a(x^\mu), \frac{\partial}{\partial x^\mu}\varphi^a(x^\mu)\right)$$

Furthermore, suppose we have a smooth symmetry x^μ_s, φ^a_s of $\mathcal{L}$ in s, and that φ^a is a solution of $\mathcal{L}$. Then there is a conserved current dependent on $\mathcal{L}$, j^u, in that

$$\partial_\mu j^\mu = 0$$

Before we prove the theorem, let us use our knowledge of tensors to understand symmetries better.

Definition 3.1. Suppose $\{x^\mu_s, \varphi^a_s\}$ is a smooth symmetry of the Lagrangian $\mathcal{L}$. We define the *coordinate shift*

$$X^\mu = \left.\frac{dx^\mu_s}{ds}\right|_{s=0}$$

the *field shift*

$$\bar{\psi}^a = \left.\frac{\partial \varphi_s^a(x^\mu)}{\partial s}\right|_{s=0}$$

and the total shift

$$\psi^a = \left.\frac{\partial \varphi_s^a(x_s^\mu)}{\partial s}\right|_{s=0}$$

Proposition. For a smooth symmetry $\{\boldsymbol{x}_s, \varphi_s^a\}$

$$\psi^a = \bar{\psi}^a + \partial_\mu \varphi^a X^\mu$$

Proof. Writing out $\varphi_s^a(x_s^\mu)$ explicitly, we see it is a function with input

$$\varphi^a\left(x^0(s), \ldots, x^n(s), s\right)$$

Thinking of s as an additional spatial coordinate, we may write $\varphi^a(x_s^\mu)$ as

$$\varphi^a\left(x^0(s), \ldots, x^n(s), x^{n+1}(s)\right)$$

where $x^{n+1}(s) = s$. Recall that the derivative of parametrized coordinates is given by

$$\frac{d}{ds} f(x^1(s), \ldots, x^n(s)) = \partial_\mu f \frac{dx^\mu(s)}{ds}$$

where $\mu = 0, \ldots, n$. Using this principle explicitly,

$$\frac{\partial}{\partial s}\varphi^a\left(x^0(s), \ldots, x^n(s), s\right) = \frac{d}{ds}\varphi^a\left(x^0(s), \ldots, x^n(s), x^{n+1}(s)\right) = \partial_\mu \varphi^a \frac{dx_s^\mu}{ds}$$

where $\mu = 0, \ldots, n+1$. Rewriting this where $\mu = n+1$ has $x^\mu = s$, we have the derivative as

$$\frac{\partial \varphi_s^a(x^\mu)}{\partial s} + \partial_\mu \varphi_s^a \frac{dx_s^\mu}{ds}$$

Thus we get

$$\begin{aligned}
\psi^a &= \left.\frac{\partial \varphi_s^a\left(x_s^\mu\right)}{\partial s}\right|_{s=0} \\
&= \left.\frac{\partial \varphi_s^a(x^\mu)}{\partial s} + \partial_\mu \varphi_s^a \frac{dx_s^\mu}{ds}\right|_{s=0} \\
&= \bar{\psi}^a + \partial_\mu \varphi^a X^\mu
\end{aligned}$$

■

Since a symmetry may have both a coordinate and field shift and each symmetry generates its own conserved current, we would expect the current to depend on ψ.

Now we will find some tools to help us prove the theorem.

Lemma 3.1.1. Assume x_s^μ is a smooth transformation of the coordinates x^μ of M. Recall the differential form $d\boldsymbol{x}_s = dx_s^1 \wedge \ldots \wedge dx_s^n$. Then

$$\left.\frac{d}{ds}\left(d\boldsymbol{x}_s\right)\right|_{s=0} = \partial_\mu X^\mu d\boldsymbol{x}$$

Proof. From the definition of the wedge product, we see that the derivative of a tensor product and consequently a wedge product should follow the same product rule as we are familiar with in $\mathbb{R}$, provided the order is unchanged.

First, notice that

$$\begin{aligned} d(X^\mu) &= d\left(\frac{d}{ds}x_s^\mu\bigg|_{s=0}\right) \\ \frac{\partial X^\mu}{\partial x^1}dx^1 + \ldots + \frac{\partial X^\mu}{\partial x^n}dx^n &= \frac{d}{ds}(dx_s^\mu)\bigg|_{s=0} \end{aligned}$$

Thus we get

$$\begin{aligned} \frac{d}{ds}\left(d\boldsymbol{x}_s\right)\bigg|_{s=0} &= \frac{d}{ds}\left(dx_s^0 \wedge \ldots \wedge dx_s^n\right)\bigg|_{s=0} \\ &= \frac{d}{ds}dx_s^0\bigg|_{s=0} \wedge \ldots \wedge dx^n + \ldots + dx^0 \wedge \ldots \wedge \frac{d}{ds}dx_s^n\bigg|_{s=0} \\ &= \left(\frac{\partial X^0}{\partial x^0}dx^0 + \ldots + \frac{\partial X^0}{\partial x^n}dx^n\right) \wedge \ldots \wedge dx^n + \ldots \\ &\quad + dx^0 \wedge \ldots \wedge \left(\frac{\partial X^n}{\partial x^0}dx^0 + \ldots + \frac{\partial X^n}{\partial x^n}dx^n\right) \\ &= \frac{\partial X^0}{\partial x^0}dx^0 \wedge \ldots \wedge dx^n + \ldots + dx^0 \wedge \ldots \wedge \frac{\partial X^n}{\partial x^n}dx^n \qquad (dx^i \wedge dx^j = 0 \iff i = j) \\ &= \sum_{\mu=0}^{n} \frac{\partial X^\mu}{\partial x^\mu}d\boldsymbol{x} \\ &= \partial_\mu X^\mu d\boldsymbol{x} \end{aligned}$$

■

Lemma 3.1.2. If f is a vector-valued function and X^μ is a vector field, then we have the product rule

$$\partial_\mu\left(fX^\mu\right) = \left(\partial_\mu f\right)X^\mu + f\partial_\mu X^\mu$$

We will omit this proof, as it follows directly from the definition of partial derivatives and the product rule for real derivatives. These lemmas provide the necessary tools for computing j^μ and proving Noether's Theorem.

Proof of Theorem. Since φ^a is a solution of $\mathcal{L}$, the Euler–Lagrange equations are satisfied, so

$$\frac{\partial \mathcal{L}}{\partial \varphi^a} + \partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \varphi^a)} = 0$$

Furthermore, since we have a symmetry in s, we know

$$0 = \frac{d}{ds}\mathcal{L}\left(x_s^\mu, \varphi_s^a, \frac{\partial}{\partial x_s^\mu}\varphi_s^a\right)d\boldsymbol{x}_s\bigg|_{s=0}$$

Applying the product rule to this derivative, we get

$$\frac{d}{ds}\left(\mathcal{L}\left(x_s^\mu, \varphi_s^a, \frac{\partial}{\partial x_s^\mu}\varphi_s^a\right)\right) d\boldsymbol{x}_s + \mathcal{L}\left(x_s^\mu, \varphi_s^a, \frac{\partial}{\partial x_s^\mu}\varphi_s^a\right)\frac{d}{ds} d\boldsymbol{x}_s\bigg|_{s=0}$$

Next, apply the definition of the total derivative to $\mathcal{L}$ and use Lemma 3.1.1 to get

$$\left(\frac{\partial \mathcal{L}}{\partial x_s^\mu}\frac{dx_s^\mu}{ds} + \frac{\partial \mathcal{L}}{\partial \varphi_s^a}\frac{d\varphi_s^a(x_s^\mu)}{ds} + \frac{\partial \mathcal{L}}{\partial\left(\partial_{\mu_s}\varphi_s^a\right)}\frac{d}{ds}\frac{\partial \varphi_s^a(x_s^\mu)}{\partial x_s^\mu}\right) d\boldsymbol{x}_s\bigg|_{s=0} + \mathcal{L}\partial_\mu X^\mu d\boldsymbol{x}$$

where

$$\partial_{\mu_s} = \frac{\partial}{\partial x_s^\mu}$$

Evaluating the expression at $s = 0$, using the definition of ψ and writing in index notation we get

$$\left(\partial_\mu \mathcal{L} X^\mu + \frac{\partial \mathcal{L}}{\partial \varphi^a}\psi^a + \frac{\partial \mathcal{L}}{\partial\left(\partial_\mu \varphi^a\right)}\partial_\mu \psi^a\right) d\boldsymbol{x} + \mathcal{L}\partial_\mu X^\mu d\boldsymbol{x}$$

Notice that the second term in this expression is the first term of the Euler–Lagrange Equations, so we may substitute

$$\frac{\partial \mathcal{L}}{\partial \varphi^a} = \partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \varphi^a)}$$

Performing this substitution and factoring out $d\boldsymbol{x}$ yields

$$\left(\partial_\mu \mathcal{L} X^\mu + \mathcal{L}\partial_\mu X^\mu + \left(\partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \varphi^a)}\right)\psi^a + \frac{\partial \mathcal{L}}{\partial\left(\partial_\mu \varphi^a\right)}\partial_\mu \psi^a\right) d\boldsymbol{x}$$

Next, we may apply the product rule (Lemma 3.1.2) on the two left terms and the two right terms independently to get

$$\left(\partial_\mu\left(\mathcal{L}X^\mu\right) + \partial_\mu\left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \varphi^a)}\psi^a\right)\right) d\boldsymbol{x}$$

Since the derivative is linear, we may factor it across addition.

$$\partial_\mu\left(\mathcal{L}X^\mu + \frac{\partial \mathcal{L}}{\partial\left(\partial_\mu \varphi^a\right)}\psi^a\right) d\boldsymbol{x}$$

Thus the integrand is the remaining term without $d\boldsymbol{x}$. Define

$$j^\mu = \mathcal{L}X^\mu + \frac{\partial \mathcal{L}}{\partial\left(\partial_\mu \varphi^a\right)}\psi^a$$

Since the integrand is 0, we know $\partial_\mu j^\mu = 0$ ■

By proposition 3.1, notice that if $X = \mathbf{0}$ ($x_s^\mu = x^\mu$) then

$$j^\mu = \frac{\partial \mathcal{L}}{\partial\left(\partial_\mu \varphi^a\right)}\bar{\psi}^a$$

If $\bar{\psi}^a = 0$ ($\varphi^a_s = \varphi^a$) then

$$j^\mu = \mathcal{L}X^\mu + \frac{\partial \mathcal{L}}{\partial\left(\partial_\mu \varphi^a\right)}\partial_\nu \varphi^a X^\nu$$

3.2 Applications

To begin exploring some conservation laws, we will first consider how the choice of symmetry affects the conserved quantity.

Example 3.1. Consider the Wave Lagrangian

$$\mathcal{L}(\varphi) = \|\nabla\varphi\|^2 - \dot{\varphi}^2 = \partial_\mu \varphi \partial^\mu \varphi$$

where φ is the spatial gradient. We will perform the simple symmetry $\varphi_s = \varphi + s$. Clearly this fixes $\mathcal{L}$ as $\mathcal{L}$ only depends on partials of φ, and $\partial^\mu \varphi_s = \partial^\mu \varphi$. Since there is no change in coordinates, our conserved current is

$$j^\mu = \frac{\partial \mathcal{L}}{\partial\left(\partial_\mu \varphi^a\right)} \cdot \psi$$

Compute

$$\begin{aligned} \psi(x^\mu) &= \left.\frac{\partial \varphi_s(x^\mu_s)}{\partial s}\right|_{s=0} \\ &= \left.\frac{\partial(\varphi(x^\mu_s) + s)}{\partial s}\right|_{s=0} \\ &= 1|s = 0 \\ &= 1 \end{aligned}$$

Thus we get

$$j^\mu = \frac{\partial \mathcal{L}}{\partial\left(\partial_\mu \varphi\right)} \cdot \psi = \frac{\partial \mathcal{L}}{\partial\left(\partial_\mu \varphi\right)}$$

But then

$$0 = \partial_\mu j^\mu = \partial_\mu \frac{\partial \mathcal{L}}{\partial\left(\partial_\mu \varphi\right)}$$

is exactly the statement of the Euler–Lagrange equations for φ since $\frac{\partial \mathcal{L}}{\partial \varphi} = 0$. Generalizing this, we may write the following proposition.

Proposition. If $\mathcal{L}$ does not depend explicitly on φ^a, then the symmetry $\varphi^a_s = \varphi^a + s$ fixes $\mathcal{L}$ for all s and results in only the Euler–Lagrange equations for $\mathcal{L}$.

This proposition completes our the discussion of a closed one-dimensional dynamic system in sections 1.2 and 1.3, and how the symmetry $x_s(t) = x(t) + s$ gives conservation of energy, as do the Euler-Lagrange equations.

Example 3.2. Recall the Schrödinger Lagrangian:

$$\mathcal{S}(\varphi) = \frac{1}{2m}\nabla\varphi \cdot \nabla\varphi^* + V\varphi\varphi^* - \frac{i}{2}(\varphi^*\dot{\varphi} - \varphi\dot{\varphi}^*)$$

where ∇ is the spatial gradient, $*$ denotes the complex conjugate, and m is a positive real number representing the mass of a particle. Traditionally, ψ is used where we have written φ, but we are already using ψ to represent the tangent of the variation. Consider the symmetry

$$\varphi_s = e^{is}\varphi, \varphi_s^* = e^{-is}\varphi^*$$

Since the coordinates are unchanged,

$$j^\mu = \frac{\partial \mathcal{L}}{\partial\left(\partial_\mu \varphi^a\right)} \cdot \psi^a$$

where $\varphi^a = \varphi, \varphi^*$. Compute

$$\begin{aligned}
\psi^0 &= \left.\frac{\partial \varphi_s^0\left(x_s^\mu\right)}{\partial s}\right|_{s=0} \\
&= i\varphi \\
\psi^1 &= \left.\frac{\partial \varphi_s^1\left(x_s^\mu\right)}{\partial s}\right|_{s=0} \\
&= -i\varphi^*
\end{aligned}$$

Thus the conserved current is

$$\begin{aligned}
j^0 &= \frac{\partial \mathcal{S}}{\partial(\partial_0\varphi^a)}\psi^a \\
&= \frac{\partial \mathcal{S}}{\partial(\dot{\varphi})} i\varphi - \frac{\partial \mathcal{S}}{\partial(\dot{\varphi}^*)} i\varphi^* \\
&= -\frac{i}{2}\varphi^* i\varphi - \frac{i}{2}\varphi i\varphi^* \\
&= \varphi^*\varphi \\
j^k &= \frac{\partial \mathcal{S}}{\partial(\partial_k\varphi^a)}\psi^a && k = 1, 2, 3 \\
&= \frac{\partial \mathcal{S}}{\partial(\partial_k\varphi)} i\varphi - \frac{\partial \mathcal{S}}{\partial(\partial_k\varphi^*)} i\varphi^* \\
&= \frac{1}{2m}\frac{\partial \varphi^*}{\partial x_k} i\varphi - \frac{1}{2m}\frac{\partial \varphi}{\partial x_k} i\varphi^* \\
&= \frac{i}{2m}\left(\frac{\partial \varphi^*}{\partial x_k}\varphi - \frac{\partial \varphi}{\partial x_k}\varphi^*\right)
\end{aligned}$$

As in our discussion above, $j^0 = \varphi\varphi^*$ represents a density of some quantity. As in Section 1.4, we may recover the Schrödinger Equation

$$i\frac{d\varphi}{dt} = -\frac{1}{2m}\Delta\varphi + V\varphi$$

from $\mathcal{S}$. If φ is a wave-function representing a field and V is the potential energy of the field, then $\varphi\varphi^*$ is taken to be the 'positional probability density'. In other words, its spatial integral

$$P = \int_{\Omega} \varphi \varphi^* dx^1 \dots dx^n$$

is the probability that the particle represented by the wave function is contained in the region Ω. The theorem then tells us that this quantity, the Noether charge, is conserved; if we are less likely to find the particle in our region, it is because the probability 'continuously flowed' out of the region and we are more likely to find it elsewhere, where the continuous flow is given by j^k. More importantly, this shows total probability is conserved in a closed system. If a particle is less likely to be observed at p at time $t + c$ than time t, it is because it is more probable to be found elsewhere and the probability moved smoothly away from p between t and $t + c$.

Theorem 3.2. A particle that obeys the Schrödinger equation for its wave equation φ has conserved positional probability.

References

[1] F.W. Byron and R.W. Fuller. *Mathematics of Classical and Quantum Physics*. Number v. 1-2 in Dover books on physics and chemistry. Dover Publications, 1992.

[2] Emmy Noether. *Encyclopædia Brittanica*. 2015.

[3] M. Spivak. *Calculus On Manifolds: A Modern Approach To Classical Theorems Of Advanced Calculus*. Mathematics monograph series. Westview Press, 1971.