

Class Notes

Math 505 - Classical Field Theory
Professor Kai Sun



Jinyan Miao

Fall 2023

Contents

1 | Classical Mechanics

Just as going from Newton's Law to Lagrangian mechanics for point-like particles, here we will develop a theory from wave equations to Lagrangian mechanics, referred as the classical field theory.

Action and Lagrangian

Example: For kinetic energy of a suspended mass by a massless string of length r . The mass has kinetic energy $T = m(r\dot{\theta})^2/2$, with potential energy $V = mgh = -mgr \cos(\theta)$. Hence we can write the Lagrangian

$$L = T - V = \frac{1}{2}mr^2\dot{\theta}^2 + mgr \cos(\theta).$$

In such a system, θ is the only canonical coordinate, with velocity $\dot{\theta}$.

In general, a Lagrangian is a function of n coordinates denoted as q_1, q_2, \dots, q_n and n velocities denoted as $\dot{q}_1, \dot{q}_2, \dots, \dot{q}_n$. That is, we can write

$$L(\dot{q}_1, \dot{q}_2, \dots, \dot{q}_n, q_1, q_2, \dots, q_n) \quad (1.1)$$

as the Lagrangian of the system, from which we can define the action

$$S := \int_{t_0}^{t_1} L(t) dt. \quad (1.2)$$

The **principle of least action** requires that $\delta S = 0$ for a physical path from t_0 to t_1 , that is S is stationary.

Example: For the pendulum problem, we here have

$$L(\dot{\theta}, \theta) = \frac{1}{2}mr^2\dot{\theta}^2 + mgr \cos(\theta),$$

hence we have

$$S = \int_{t_i}^{t_f} L(\dot{\theta}, \theta) dt = \int_{t_i}^{t_f} \left(\frac{1}{2}mr^2\dot{\theta}^2 + mgr \cos(\theta) \right) dt$$

we can apply small variation as $\theta(t) \rightarrow \theta(t) + \delta\theta(t)$, and correspondingly $\dot{\theta}(t) \rightarrow \dot{\theta}(t) + \delta\dot{\theta}(t)$. Here we write

$$\begin{aligned}\delta S = S' - S &= \int_{t_i}^{t_f} \left(L(\dot{\theta} + \delta\dot{\theta}, \theta + \delta\theta) - L(\dot{\theta}, \theta) \right) dt \\ &= \int_{t_i}^{t_f} \left(\frac{\partial L}{\partial \dot{\theta}} \frac{d(\delta\theta)}{dt} + \frac{\partial L}{\partial \theta} \delta\theta \right) dt.\end{aligned}$$

Utilizing integral by part and dropping the boundary term, we obtain

$$\delta S = \int_{t_i}^{t_f} \left(- \left(\frac{d}{dt} \frac{\partial L}{\partial \dot{\theta}} \right) \delta\theta + \frac{\partial L}{\partial \theta} \delta\theta \right) dt = \int_{t_i}^{t_f} \left(- \frac{d}{dt} \frac{\partial L}{\partial \dot{\theta}} + \frac{\partial L}{\partial \theta} \right) dt \delta\theta.$$

For requiring that $\delta S = 0$, we must have

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{\theta}} = \frac{\partial L}{\partial \theta} \quad (1.3)$$

holds for arbitrary $\delta\theta$.

For a general Lagrangian of the form given by (1.1), we have n equations, called the Euler Lagrangian Equations, each of the form

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} = \frac{\partial L}{\partial q_i}, \quad (1.4)$$

for $i \in \mathbb{N}_n$. Here Eq. (1.4) is called the equation of motion.

Example: For the pendulum example, we have (1.3) holds, and hence we can write

$$-mgr \sin(\theta) = \frac{\partial L}{\partial \theta} = \frac{d}{dt} \frac{\partial L}{\partial \dot{\theta}} = \frac{d}{dt} \left(\frac{mr^2}{2} \left(\frac{\partial \dot{\theta}}{\partial \dot{\theta}} \right)^2 \right) = mr^2 \ddot{\theta}.$$

Hence the equation of motion of such system is

$$mr\ddot{\theta} = -mg \sin(\theta).$$

For small angle $\theta \ll 1$, we have $\sin(\theta) \approx \theta$ and hence we can write

$$mr\ddot{\theta} = mg\theta,$$

from which we obtain the angular velocity $\omega = \sqrt{g/r}$ for such system.

1.1.1 Constraint and Lagrangian Multipliers

One can choose any coordinate system to describe the physics, as long as the the chosen coordinate components are independent. For the pendulum example, one cannot use the general Cartesian coordinate (x, y) as the only ingredient to describe the suspended mass as $x^2 + y^2 = r^2$ is a constraint to x and y , or in other words, x and y are not independent.

Example: For the pendulum example, if one uses the Cartesian coordinate (x, y) to describe the system, one can write

$$T = \frac{m\dot{x}^2}{2} + \frac{m\dot{y}^2}{2}, \quad V = mgy,$$

and hence the Lagrangian is given by

$$L = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}m\dot{y}^2 - mgy,$$

from which we can write the equation of motion of the system using (1.2), which yields

$$m\ddot{x} = 0, \quad m\ddot{y} = -mg.$$

which does not describe the full picture of the system as the condition $x^2 + y^2 = r^2$ is not included in the analysis.

However, one can define a constraint of the form

$$f(\dot{q}_1, \dot{q}_2, \dots, \dot{q}_n, q_1, q_2, \dots, q_n) = 0$$

and include f to the Lagrangian to restore the constraint on the coordinates.

Example: For the pendulum example, one can write

$$f(x, y) = \sqrt{x^2 + y^2} - r = 0$$

and define the new Lagrangian

$$\hat{L} = L - \lambda f = \hat{L}(\dot{q}_1, \dot{q}_2, \dots, \dot{q}_n, q_1, q_2, \dots, q_n, \lambda),$$

where $\lambda \in \mathbb{R}$ is called the Lagrange multiplier, and L is the Lagrangian obtained using the coordinates $(\dot{q}_1, \dot{q}_2, \dots, \dot{q}_n, q_1, q_2, \dots, q_n)$ only. Then by principle of least motion, we require that $\delta S = 0$, and obtain the equations of motion for the system

$$\frac{d}{dt} \frac{\partial \hat{L}}{\partial \dot{q}_i} = \frac{\partial \hat{L}}{\partial q_i}, \quad \frac{\delta S}{\delta \lambda} = f = 0.$$

If one is able to obtain a coordinate system such that no constraint is needed, or in other words, coordinate components are independent, the system is said to be conanical, and solving conanical systems is generally much easier than non-conanical systems.

If one has a system that requires more than one, say n , constraints f_i on the coordinate system, one needs to include n Lagrange multipliers λ_i such that the Lagrangian reads

$$\hat{L} = L - \left(\sum_{i \in \mathbb{N}_n} \lambda_i f_i \right)$$

and obtain $n + 1$ equations of motion by requiring that $\delta S = 0$.

Canonical Momenta and Hamiltonian

For a given Lagrangian and general coordinates, one can define the canonical momenta

$$p_i = \frac{\partial L}{\partial \dot{q}_i} \quad (1.5)$$

Example: For the pendulum example in the previous section, we have

$$p = \frac{\partial L}{\partial \dot{\theta}} = mr^2 \dot{\theta}$$

Note: Here if the coordinates are some linear coordinates, we can interpret the canonical momenta as linear momenta, while if the coordinates are some angular coordinates, we can interpret the canonical momenta as angular momenta. Also note here the canonical momenta and velocity of the system carry the same information, as for the pendulum example, we can write

$$\dot{\theta} = \frac{p}{mr^2}$$

Hence we can choose either (q_i, \dot{q}_i) or (q_i, p_i) to describe the system. The Lagrangian approach uses the the Lagrangian $L(q_i, \dot{q}_i)$ to describe the system, while the Hamiltonian approach uses Hamiltonian $H(p_i, q_i)$ to describe.

Example: In thermodynamics, one describe a gas using the pair (S, P) or the pair (T, P) , as $dU = T dS - P dV$ or $dF = -S dT - P dV$, where U is the internal energy and F is the Helmholtz free energy. In thermodynamics, the Legendre transformation gives

$$F = U - TS = U - \left(\frac{\partial U}{\partial S} \right)_V S.$$

In mechanics, a very similar idea applies.

The transformation from a Lagrangian $L(\dot{q}_1, \dot{q}_2, \dots, \dot{q}_n, q_1, q_2, \dots, q_n)$ to the correspondingly defined Hamiltonian $H(p_1, p_2, \dots, p_n, q_1, q_2, \dots, q_n)$ is given by

$$H = \left(\frac{\partial L}{\partial \dot{q}_i} \right) \dot{q}_i - L = p_i \dot{q}_i - L, \quad (1.6)$$

where we have used the Einstein notation that same index get summed.

Example: For the pendulum example, we can write that

$$p_\theta = \frac{\partial L}{\partial \dot{\theta}} = mr^2 \dot{\theta},$$

from which we can write

$$H = mr^2 \dot{\theta}^2 - \frac{1}{2} mr^2 \dot{\theta}^2 - mgr \cos(\theta) = \frac{1}{2} mr^2 \dot{\theta}^2 - mgr \cos(\theta) = \frac{p_\theta^2}{2mr^2} - mgr \cos(\theta)$$

Notice here that we have the Hamiltonian is the sum of kinetic energy and potential energy. While in general, in classical mechanics, $H = \text{KE} + \text{PE}$ only when H does not have explicit time dependence, and H does not necessarily represent the sum of kinetic

and potential energy in other fields of physics.

Note: In summary, to transform from Lagrangian to Hamiltonian, we need to define the canonical momenta via (1.5), then we need to apply the transformation via (1.6). Finally, we need to ensure that the Hamiltonian H is a function of p_i and q_i only, which does not contain the velocity components \dot{q}_i explicitly.

1.2.1 Hamiltonian Mechanics

For a Lagrangian $L(\dot{q}_i, q_i)$, we can write

$$dL = \frac{\partial L}{\partial \dot{q}_i} d\dot{q}_i + \frac{\partial L}{\partial q_i} dq_i. \quad (1.7)$$

For Hamiltonian $H(p_i, q_i)$, we can write

$$dH = \frac{\partial H}{\partial p_i} dp_i + \frac{\partial H}{\partial q_i} dq_i. \quad (1.8)$$

Note H is defined via (1.6), here we have

$$dH = d(p_i \dot{q}_i - L) = d(p_i \dot{q}_i) - dL = p_i d\dot{q}_i + \dot{q}_i dp_i - \frac{\partial L}{\partial \dot{q}_i} d\dot{q}_i - \frac{\partial L}{\partial q_i} dq_i$$

Note that H should not contain \dot{q} term explicitly, hence dH should not have $d\dot{q}_i$ here. Also note that by definition $p_i = \partial L / \partial \dot{q}_i$. Hence we have

$$dH = \dot{q}_i dp_i - \frac{\partial L}{\partial q_i} dq_i. \quad (1.9)$$

Also note that combining (1.4) and (1.5), we obtain

$$\frac{\partial L}{\partial q_i} = \frac{dp_i}{dt} = \dot{p}_i.$$

Hence (1.9) becomes

$$dH = \dot{q}_i dp_i - \dot{p}_i dq_i$$

for which comparing with (1.8) we conclude that we require

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q_i}. \quad (1.10)$$

As a result, for Lagrangian mechanics, we have, via (1.4), n equations of motion. While for Hamiltonian mechanics, we have, via (1.10), $2n$ equations of motion, where n is the number of canonical coordinate components. While (1.4) is a second order differential equation, (1.10) is a system of two first order differential equations.

Comparison between Lagrangian and Hamiltonian

In classical physics, the complexity of Lagrangian (n second order differential eq.) is the same as the Hamiltonian ($2n$ first order differential eq.). However, Hamiltonian has clearer physical meaning, that is the energy of the system, while the symmetry of the system can be more easily seen from the Lagrangian.

In the context of special relativity, the action S of a system is a 4-dimensional scalar, and is invariant under Lorentz boost, while the Hamiltonian of the system is neither a 4-dimensional scalar nor 4-dimensional vector, and is not invariant under Lorentz boost (changing reference frame would change the kinetic energy of the system described). Hence Lagrangian is preferred in this context.

In the context of quantum system, Hamiltonian has canonical quantization, directly fitting the Heisenberg picture of quantum systems. In quantum system, (1.10) still holds while viewing H , \dot{q} , q , \dot{p}_i , and p_i as operators, and enforcing $[\hat{q}_i, \hat{p}_j] = i\hbar\delta_{ij}$, from which one obtains

$$\frac{d\hat{q}_i}{dt} = \frac{\partial \hat{H}}{\partial \hat{p}_i} = \frac{i}{\hbar}[\hat{H}, \hat{q}_i], \quad \frac{d\hat{p}_i}{dt} = -\frac{\partial \hat{H}}{\partial \hat{q}_i} = -\frac{i}{\hbar}[\hat{H}, \hat{p}_i].$$

In quantum system, Lagrangian approach utilizes path integral formalism, which is very different from what is done using Hamiltonian.

Symmetry

There are two types of symmetry, discrete or continuous. By symmetry, we mean one does some operation to the system and the system obey the same laws of physics.

Example: For a mass suspended by massless string of length r , the angle between the string and the vertical is denoted as θ . Note such system $\theta \rightarrow -\theta$ is a symmetry, as the physics laws stay the same by mirroring the mass over the vertical axis. In particular, the equation of motion for the system

$$mr \frac{d^2\theta}{dt^2} = mg \sin(\theta)$$

is invariant under the mapping $\theta \rightarrow -\theta$.

For now we will focus on the Lagrangian formalism. Action, the time integral of Lagrangian, stays the same under symmetry. In the case of classical mechanics (and some other fields), for most textbooks, the Lagrangian of the system is unchanged under symmetry.

Example: For the pendulum example, under the map $\theta \rightarrow -\theta$, L is invariant

$$L = \frac{1}{2}mr^2\dot{\theta}^2 + mgr \cos(\theta).$$

Example: In the context of special relativity, the Lagrangian of the system changes under Lorentz boost, while the time interval dt in the time integral of Lagrangian also changes, and the resultant action of the system stays the same under Lorentz boost.

Remark: The symmetry $\theta \rightarrow -\theta$ in the pendulum example is a discrete symmetry. Another type of symmetry is the continuous symmetry, for instance a $\vec{r} \rightarrow \vec{r} + \vec{a}$ transformation where \vec{a} can be any real-valued vector. Note also that continuous symmetry can be broken down into smaller steps of continuous symmetry, as in the $\vec{r} \rightarrow \vec{r} + \vec{a}$ case, $\vec{r} \rightarrow \vec{r} + \vec{a}/N$ is also a continuously symmetric transformation and being performed N times will restore the $\vec{r} \rightarrow \vec{r} + \vec{a}$ transformation.

1.4.1 Noether's Theorem

The focus of this subsection is that each continuous symmetry gives a conservation law. Now suppose one has a Lagrangian $L(\dot{q}_1, \dot{q}_2, \dots, \dot{q}_n, q_1, q_2, \dots, q_n)$. One performs a translation at the j -th coordinate $q_j \rightarrow q_j + a$. Here we have immediately

$$\frac{dq_j}{dt} \rightarrow \frac{dq_j}{dt} + \frac{da}{dt} = \frac{dq_j}{dt}$$

as we have treated a as a constant. If the system is symmetric under such a translation $q_j \rightarrow q_j + a$, then L stays the same, hence we have

$$\frac{\partial L}{\partial q_j} = 0,$$

then from (1.4) we obtain

$$\dot{p}_j = \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_j} = \frac{\partial L}{\partial q_j} = 0,$$

so we see here p_j is a constant. Here we have present that the continuous symmetry with respect to one of the coordinate components q_j gives us a conservation law $p_j = \text{constant}$. In general, one can usually perform a change of coordinates to obtain the conservation law via one of the coordinate components for some given arbitrary continuous symmetry that is originally not along any given coordinate component.

1.4.2 Constraint on Action and Equations of Motion

A symmetry of the system in fact imposes constraint to the action and equation of motion of the system. For the pendulum example, as action of the system stays the same under symmetry $\theta \rightarrow -\theta$, S should be an even function of θ .

Example: For the pendulum example, here we *pretend* that we have no knowledge about Newtonian mechanics, and we would like to determine what the system should look like only based on the symmetry of the system. Here we some assumptions that we will enforce:

1. $\delta S = 0$ as for the principle of least action;
2. The pendulum is characterized by one angle θ only;
3. $t \rightarrow -t$ is a symmetry, called the time reversal symmetry;
4. $\theta \rightarrow -\theta$ is a symmetry, called the spatial mirror symmetry;
5. We are dealing with small oscillation and slow motion.

From which one knows that the Lagrangian takes the form

$$L(\theta, \dot{\theta}, \ddot{\theta}, \dots \text{ higher order derivatives of } \theta)$$

From small oscillation, the higher orders derivative of θ are negligible. For slow motion, we can apply Taylor expansion, hence we have Lagrangian to be simplified as

$$L(\theta, \dot{\theta}) = \kappa + \alpha\theta + \beta\dot{\theta} + \gamma\theta^2 + \delta\theta\dot{\theta} + \epsilon\dot{\theta}^2$$

where $\alpha, \beta, \gamma, \delta, \epsilon, \kappa \in \mathbb{R}$. Note that via (1.4), κ does not contribute to the equation of motion for the system. By symmetry requirement, $\alpha = \beta = 0$ by the spatial mirror symmetry, and $\beta = 0$ also by time symmetry. Furthermore, $\dot{\theta}$ is the total derivative term, here we write, suppose $\tilde{L} = b\dot{\theta}$

$$S = \int_{-\infty}^{\infty} \tilde{L} dt = b \int_{-\infty}^{\infty} \frac{d\theta}{dt} dt = \frac{\theta(t=\infty) - \theta(t=-\infty)}{2}$$

In physics, what happened at $t = -\infty$ and $t = \infty$ is a boundary condition and has nothing to do with the equation of motion. So as far as the equation of motion is concerned, this total derivative terms plays no role and thus can be ignored, that is $\beta = 0$. Furthermore, we also require $\delta = 0$ in the term $\delta\theta\dot{\theta}$ by time symmetry $\theta\dot{\theta} = -\theta\dot{\theta}$ and the fact that $\theta\dot{\theta} = (1/2)(d\theta^2/dt)$ gets integrated to 0 in periodic boundary conditions. Hence the Lagrangian gets simplified to the form

$$L = \gamma\theta^2 + \epsilon\dot{\theta}^2,$$

which gives an equation of motion

$$\epsilon\ddot{\theta} = \gamma\theta,$$

or in other words

$$\ddot{\theta} = \frac{\gamma}{\epsilon}\theta. \tag{1.11}$$

Note here γ/ϵ is not a universal quantity, but the form of (1.11) is universal under the given symmetries of the system.

2 | Classical Field Theory

Here we will shift our focus from classical mechanics to fields. Instead of rigid body or point mass, we will mostly focus on continuous medium or continuous fields.

Example: Let ρ denote the density of gas, or liquid, in a region. Here we define

$$\varphi(\vec{r}, t) = \rho(\vec{r}, t) - \bar{\rho},$$

where ρ is the mass density at position \vec{r} and time t and $\bar{\rho}$ is the average density of the whole region. Such description φ is a field. In the following we will explore how one can define the Lagrangian and action using φ and its derivatives.

The principle of locality states that an object can get direct influence only from its immediate neighbors.

Example: The principle of locality can be understood via an example of shaking a rope: One side of the rope is held by a person and the other side of the rope is held by a wall. If the person shakes the rope from one end, the movement of the rope takes time to propagate to the other end of the rope.

Remark: Note that there is no way to prove the principle of locality, while it is a general assumption for studying modern physics, as the standard model and theory of general relativity are both local.

Example: Maxwell's Equations respects the principle of locality as it is a set of differential equations, which describes how particles are affected by infinitesimal neighbors. While the Coulomb's Law and the gravitational laws do not respect the principle of locality, as they are just the approximation of the field theory derived from the principle of locality.

Note: Mathematically, *local* theory can be written in terms of differential equations, while a non-local theory requires integral equations. Local theory are easier to be dealt with as differential equations are easier to be solved than the integral equations, and also because we only need to worry about objects nearby for the dynamics of an object.

Note: Locality enforces strong constraint on the interaction potential between objects. We shall see later that the potential between particles can only take two possible forms $\sim 1/r$ or $\sim 1/re^{kr}$.

Principle of locality allows us to define the local Lagrangian density \mathcal{L} , that is the Lagrangian per volume, at each point (t, x, y, z) as a function of the field and the derivatives

of the field at (t, x, y, z) . Given a Lagrangian density $\mathcal{L}(t, x, y, z)$, we then can define the local action by

$$S := \iiint \mathcal{L} dt dx dy dz ,$$

and define the Lagrangian by

$$L = \iiint \mathcal{L} dx dy dz .$$

Note here in field theory, the Lagrangian is no longer a function of only coordinates and velocities, instead it is a function of the field and many of its derivatives, as for the gas example here we write

$$\mathcal{L}(t, x, y, z) = \mathcal{L}(\varphi(t, x, y, z), \partial\varphi(t, x, y, z), \partial^2\varphi(t, x, y, z) \dots) .$$

Via principle of locality, an action of the following form is never allowed:

$$S = \iiint \iiint \frac{\varphi(t, x, y, z)\varphi(t', x', y', z')}{\sqrt{(x-x')^2 + (y-y')^2 + (z-z')^2}} dx' dy' dz' dx dy dz ,$$

as it requires information at two separate points (t, x, y, z) and (t', x', y', z') .

Example: Now we want to obtain the form of Lagrangian density for the gas density φ example, with the following assumptions:

1. The density is small, $\varphi \ll 1$;
2. The system possess rotational symmetry, or being isotropic at each point;
3. Slow variation assumed, that is in the long wavelength and low frequency limits.

Based on the assumptions, the higher order derivatives $\partial_x^n \varphi \propto k_x^n \varphi$ can be ignored as k_x is small when having long wavelength. Similarly, the higher order derivatives $\partial_t^n \varphi \propto \omega^n \varphi$ can be ignored as the frequency ω is small. Thus we can approximate \mathcal{L} as a function $\mathcal{L}(\varphi, \partial\varphi, \partial^2\varphi)$, for which we can apply Taylor expansion

$$\mathcal{L}(\varphi, \partial\varphi, \partial^2\varphi) = a + a_0\varphi + \sum a_i \partial_i \varphi + (\text{nonlinear terms and second order terms})$$

where a can be ignored as a constant does not contribute to δS . a_0 does not contribute either as we can write by definition of $\bar{\rho}$,

$$\iiint \iiint (\rho(\vec{r}, t) - \bar{\rho}) dt dx dy dz = \left(\iiint \iiint \rho(\vec{r}, t) dt dx dy dz \right) - \left(\iiint \iiint \bar{\rho} dt dx dy dz \right) = 0 .$$

Furthermore, the linear term of $\partial\varphi$ and the terms like $\varphi \partial\varphi = \partial\varphi^2/2$ also can be ignored as integrating such terms gives the difference in boundaries, and boundaries conditions do not contribute to the equation of motion for the system. Thus in general, we can write

$$\mathcal{L} = a\varphi^2 + \sum a_{ij} \varphi \partial_j \partial_i \varphi + \sum b_{ij} \partial_i \varphi \partial_j \varphi .$$

Now for the $\varphi \partial_j \partial_i \varphi$ terms, if one performs integration by parts on those terms, one obtains the boundary terms and a $\partial_\varphi \partial_j \varphi$ terms, thus a_{ij} can also be assumed to be zero. Lastly, for the $\partial_i \varphi \partial_j \varphi$ terms, we utilize the spatial rotation symmetry. The action is unchanged under a rotation of the coordinate system, that is the action must be a

scalar rather than a vector, and thus the Lagrangian density is also a scalar instead of a vector. The only way we are able to write \mathcal{L} as a scalar is given by

$$\mathcal{L} = a\varphi^2 + b(\partial_t\varphi)^2 + c(\nabla\varphi \cdot \nabla\varphi),$$

where $\nabla\varphi := (\partial_x\varphi, \partial_y\varphi, \partial_z\varphi)$. Here we denote $\tilde{\varphi} := \sqrt{2b}\varphi$, from which we have

$$\mathcal{L} = \frac{1}{2} \left((\partial_t\tilde{\varphi})^2 + \frac{a}{b}(\tilde{\varphi})^2 + \frac{c}{b}|\nabla\tilde{\varphi}|^2 \right).$$

As we will use $\tilde{\varphi}$ instead of φ from now, we simply write

$$\mathcal{L} = \frac{1}{2} \left((\partial_t\varphi)^2 - v^2(\nabla\varphi \cdot \nabla\varphi) - \omega_0^2\varphi^2 \right), \quad (2.1)$$

where $c/b = -v^2$ and $a/b = -\omega_0^2$. Now we have obtained the form of Lagrangian density, we can compute for the action

$$S = \iiint \mathcal{L} dt dx dy dz.$$

Here we consider $\varphi'(\vec{r}, t) = \varphi(\vec{r}, t) + \delta\varphi(\vec{r}, t)$, with the action corresponds to φ denoted as S , action corresponds to φ' denoted as S' , the Lagrangian density corresponds to φ as \mathcal{L} , and Lagrangian density corresponds to φ' as \mathcal{L}' . Here we have

$$\begin{aligned} S' &= \iiint \frac{1}{2} \left((\partial_t(\varphi + \delta\varphi))^2 - (v^2(\nabla(\varphi + \delta\varphi) \cdot \nabla(\varphi + \delta\varphi)) - \omega_0^2(\varphi + \delta\varphi)^2 \right) dt dx dy dz \\ &= \iiint \frac{1}{2} \left((\partial_t\varphi)^2 + 2\partial_t\varphi\partial_t\delta\varphi - v^2(\nabla\varphi \cdot \nabla\varphi) - 2v^2(\nabla\varphi \cdot \nabla\delta\varphi) - \omega_0^2\varphi^2 - 2\omega_0^2\varphi\delta\varphi \right) dt dx dy dz \\ &= \iiint \mathcal{L} dt dx dy dz + \iiint (\partial_t\varphi\partial_t\delta\varphi - v^2(\nabla\varphi \cdot \nabla\delta\varphi) - \omega_0^2\varphi\delta\varphi) dt dx dy dz, \end{aligned}$$

where we have neglected some terms as $\delta\varphi$ is assumed to be small. Rewriting we have

$$\delta S = S' - S = \iiint (\partial_t\varphi\partial_t\delta\varphi - v^2(\nabla\varphi \cdot \nabla\delta\varphi) - \omega_0^2\varphi\delta\varphi) dt dx dy dz$$

Integration by parts, ignoring the boundary terms, yields the following

$$\delta S = \iiint \left(-(\partial_t^2\varphi) + v^2(\nabla^2\varphi) - \omega_0^2\varphi \right) \delta\varphi dt dx dy dz.$$

By principle of least action, holding for all $\delta\varphi$, we thus require

$$0 = (\partial_t^2\varphi) - v^2(\nabla^2\varphi) + \omega_0^2\varphi. \quad (2.2)$$

Thus experimentally we only need to measure v and ω_0 to obtain the precise form of (2.2), and here (2.1) is known as the linear wave equation, characterizing the equation of motion for the field φ .

For another approach, we can approximate that

$$\mathcal{L}' = \mathcal{L} + \frac{\partial\mathcal{L}}{\partial\varphi}\delta\varphi + \frac{\partial\mathcal{L}}{\partial(\partial_t\varphi)}\partial_t\delta\varphi + \frac{\partial\mathcal{L}}{\partial(\partial_x\varphi)}\partial_x\delta\varphi + \frac{\partial\mathcal{L}}{\partial(\partial_y\varphi)}\partial_y\delta\varphi + \frac{\partial\mathcal{L}}{\partial(\partial_z\varphi)}\partial_z\delta\varphi.$$

To avoid confusion, we rewrite ∂ as δ , and see

$$\mathcal{L}' = \mathcal{L} + \frac{\delta \mathcal{L}}{\delta \varphi} \delta \varphi + \frac{\delta \mathcal{L}}{\delta(\partial_t \varphi)} \partial_t \delta \varphi + \frac{\delta \mathcal{L}}{\delta(\partial_x \varphi)} \partial_x \delta \varphi + \frac{\delta \mathcal{L}}{\delta(\partial_y \varphi)} \partial_y \delta \varphi + \frac{\delta \mathcal{L}}{\delta(\partial_z \varphi)} \partial_z \delta \varphi.$$

Integrating over spacetime and applying integration by part we have

$$S' = S + \iiint \left(\frac{\delta \mathcal{L}}{\delta \varphi} \delta \varphi - \partial_t \frac{\delta \mathcal{L}}{\delta \partial_t \varphi} \delta \varphi - \partial_x \frac{\delta \mathcal{L}}{\delta \partial_x \varphi} \delta \varphi - \partial_y \frac{\delta \mathcal{L}}{\delta \partial_y \varphi} \delta \varphi - \partial_z \frac{\delta \mathcal{L}}{\delta \partial_z \varphi} \delta \varphi \right) dt dx dy dz,$$

Here we obtain another form of equation of motion

$$\partial_t \frac{\delta \mathcal{L}}{\delta \partial_t \varphi} + \partial_x \frac{\delta \mathcal{L}}{\delta \partial_x \varphi} + \partial_y \frac{\delta \mathcal{L}}{\delta \partial_y \varphi} + \partial_z \frac{\delta \mathcal{L}}{\delta \partial_z \varphi} = \frac{\delta \mathcal{L}}{\delta \varphi}, \quad (2.3)$$

which is universal as we have only employed $\delta S = 0$ in the derivation, whereas (2.2) is not universal as (2.2) requires further assumptions about the field. For notation, (2.3) is usually denoted as

$$\partial_\mu \frac{\delta \mathcal{L}}{\delta \partial_\mu \varphi} = \frac{\delta \mathcal{L}}{\delta \varphi}. \quad (2.4)$$

Example: For the gas density φ example, from (2.1), we can write

$$\mathcal{L} = \frac{1}{2} ((\partial_t \varphi)^2 - v^2 (\partial_x \varphi)^2 - v^2 (\partial_y \varphi)^2 - v^2 (\partial_z \varphi)^2 - \omega_0^2 \varphi^2)$$

with (2.3), we obtain

$$\partial_t^2 \varphi - v^2 \partial_x^2 \varphi - v^2 \partial_y^2 \varphi - v^2 \partial_z^2 \varphi = \omega_0^2 \varphi \quad (2.5)$$

as we have

$$\frac{\delta \mathcal{L}}{\delta \partial_t \varphi} = \frac{1}{2} \frac{\delta (\partial_t \varphi)^2}{\delta \partial_t \varphi} = \partial_t \varphi,$$

$$\frac{\delta \mathcal{L}}{\delta \partial_x \varphi} = -\frac{v^2}{2} \frac{\delta (\partial_x \varphi)^2}{\delta \partial_x \varphi} = -v^2 \partial_x \varphi,$$

$$\frac{\delta \mathcal{L}}{\delta \varphi} = -\frac{\omega_0^2}{2} \frac{\delta (\varphi^2)}{\delta \varphi} = -\omega_0^2 \varphi.$$

Hence simplifying (2.5) we obtain

$$\partial_t^2 \varphi - v^2 \nabla^2 \varphi = -\omega_0^2 \varphi, \quad (2.6)$$

which has a general plane wave solution

$$\varphi = A e^{i\vec{k} \cdot \vec{r} - i\omega t}. \quad (2.7)$$

With (2.7), note here we have

$$\partial_t \varphi = -i\omega A e^{i\vec{k} \cdot \vec{r} - i\omega t}, \quad \partial_t^2 \varphi = (i\omega)^2 \varphi = -\omega^2 \varphi, \quad \nabla \varphi = A \nabla e^{i\vec{k} \cdot \vec{r} - i\omega t} = i\vec{k} \varphi.$$

Thus combining with (2.6) we have

$$-\omega^2\varphi + v^2k^2\varphi = -\omega_0^2\varphi,$$

that is ω , v , k satisfy

$$\omega^2 - v^2k^2 = \omega_0^2,$$

or in other words

$$\omega = \pm\sqrt{v^2k^2 + \omega_0^2}. \quad (2.8)$$

Here (2.8) is called the dispersion relation.

In the case where $\omega_0 \neq 0$, there is a gap in the solution space for (2.8). The frequency ω can never have access to any value in the frequency range $-\omega_0 < \omega < \omega_0$, no matter what k is used, hence we say such solution space is gapped, or massive.

For the case where $\omega_0 = 0$, we have $\omega = \pm v|k|$, and thus the solution space is said to be gapless, or massless.

Example: From particle-wave duality, from (2.8) we have

$$\begin{aligned} \omega^2\hbar^2 &= v^2k^2\hbar^2 + \omega_0^2\hbar^2 \\ E^2 &= v^2p^2 + \omega_0^2\hbar^2, \end{aligned} \quad (*)$$

where $\omega_0^2\hbar^2$ is interpreted as $m_0^2c^4$ in special relativity, and thus (*) reduces to a conservation law in special relativity. Here we can also define

$$m_0 = \frac{\hbar\omega_0}{c^2}.$$

This explains why the solution space of (2.8) is said to be massive or massless.

Example: For the gas density example, one can interpret (2.7) as sound wave propagating through the gas, in which case we have $\omega_0 = 0$. In other words, acoustic sound waves are always gapless. Note here if $k = 0$, we have $\varphi = Ae^{-i\omega_0 t}$, that is φ is homogeneous. In other words, we are changing the density of the gas homogeneously, everywhere by the same amount, but in such case the system is still in a equilibrium state and there will be no mass flow in the system. Thus the density will not change as a function of t , and thus $\omega = 0$, enforcing $\omega_0 = 0$.

Note from (2.1), we have

$$\mathcal{L} = \frac{1}{2}(\partial_t\varphi)^2 - \frac{1}{2}v^2(\nabla\varphi)^2 - \frac{1}{2}\omega_0^2\varphi - \frac{1}{2}g\varphi^4 \quad (2.9)$$

where a higher order term $-(g/2)\varphi^4$ is appended and is allowed by symmetry. Such a form (2.9) is generic and is needed for certain topics, such as the Ising field theory and the Klein-Gordon Theory for Higgs.

Now we have, using an example of gas density, defined a field $\varphi(\vec{r}, t)$. This idea can be applied to characterize many other systems, such as the water surface wave, quantum particle, and Ising model. Once a field is being defined, then one can write down the Lagrangian density, as a function of the field itself and the derivatives of the field, as allowed by the principle of locality. Possibly with some further assumption of slow variation and small density, one then employs Taylor expansion and symmetry of the system to simplify the Lagrangian density and compute the equation of motion via the principle of least action.

We note that fields only carry information about certain length scale. For microscopic details that takes place at small length scale, fields would not capture those physics. For sound waves and water surface waves, the cut-off distance is roughly the typical distance between atoms, physics at smaller length scales cannot be capture by the description of field. Thus, the theory only work when wavelength is much longer than the cut-off length scale.

The equation of motion is in general a PDE. In the case for the gas density example, the PDE (2.2) has plane wave solution, from which we obtain a relation between ω , k , and v suggested by (2.4). There are cases there the equation of motion is a non-linear PDE, as in the case where \mathcal{L} contains a φ^4 term. The case of gas density example has linear PDE, which is solvable, but only an approximation and is not very interesting as everything is predictable. A non-linear PDE is much harder to be solved, and have much broader *possibilities*. In other words, if our universe behaves as a linear PDE system, then once we know the theory of fundamental principle of particles, then we know everything about the universe, while if our system behaves like a non-linear PDE system, then even we know the fundamental laws, it is hard to predict the outcomes.

Note further that linear system implies particles are not interacting, while non-linear system implies particle interact with each other. The form of Eq. (2.6) is an example of linear system. We will show later that for linear system, the total energy of the system is directly proportional to the number of particles, while for non-linear system, the total energy is not proportional to the total number of particles.

Noether's Theory

In previous section, we have shown that continuous symmetry gives rise to conserved quantity. In classical mechanics, a conserved quantity Q is defined by the statement $dQ/dt = 0$. While in classical theory, a conserved quantity is characterized by continuity equation, such as the charge conservation

$$\partial_t \rho + \nabla \cdot \vec{j} = 0.$$

Note that the continuity equation does not give conservation law directly, while it is a combination of conservation law and principle of locality.

Example: For a region Ω , the change of charge in Ω from time t to $t + \Delta t$ is given by

$$\Delta Q_\omega = I_{\partial\Omega} \Delta t + \Delta Q_{teleport},$$

while principle of locality requires that $\Delta Q_{teleport} = 0$, thus we have

$$\iiint_{\Omega} \partial_t \rho dV = \frac{d}{dt} \iiint_{\Omega} \rho dV = \frac{dQ}{dt} = - \iint_{\partial\Omega} \vec{j} \cdot d\vec{S} = - \iiint_{\Omega} \nabla \cdot \vec{j} dV,$$

where the negative sign comes from the orientation of $\partial\Omega$. Concluding we have

$$\partial_t \rho + \nabla \cdot \vec{j} = 0,$$

where ρ is the charge density, and \vec{j} is the current density.

2.1.1 Translation symmetry in space and time

Consider a time translation symmetry $t \rightarrow t' = t + \delta t$, and similarly a spatial translation symmetry $\vec{r} \rightarrow \vec{r}' = \vec{r} + \delta \vec{r}$. Thus a space-time translation can be denoted as

$$r^\nu \rightarrow r'^\nu = r^\nu + \epsilon^\nu,$$

where we assume ϵ^ν to be constant here, and here we makes use of the notation $\partial_\mu = (\partial_0, \partial_1, \partial_2, \partial_3) = (\partial_t, \partial_x, \partial_y, \partial_z)$. Note further here r is a 4-dimensional object, including time and spatial coordinates. Thus we can write

$$S = \iiint \mathcal{L}(\varphi(r), \partial\varphi(r)) dV, \quad S' = \iiint \mathcal{L}(\varphi'(r'), \partial'\varphi'(r')) dV'.$$

Note that the coordinate has been changed, and thus we have change of coordinate

$$\iiint f dV' = \iiint f \det \left(\begin{bmatrix} \frac{\partial t'}{\partial t} & \frac{\partial x'}{\partial t} & \frac{\partial y'}{\partial t} & \frac{\partial z'}{\partial t} \\ \frac{\partial t'}{\partial x} & \frac{\partial x'}{\partial x} & \frac{\partial y'}{\partial x} & \frac{\partial z'}{\partial x} \\ \frac{\partial t'}{\partial y} & \frac{\partial x'}{\partial y} & \frac{\partial y'}{\partial y} & \frac{\partial z'}{\partial y} \\ \frac{\partial t'}{\partial z} & \frac{\partial x'}{\partial z} & \frac{\partial y'}{\partial z} & \frac{\partial z'}{\partial z} \end{bmatrix} \right) dV,$$

where the Jacobian can be expanded as

$$\begin{aligned} \det \left(\frac{\partial r'^\nu}{\partial r^\mu} \right) &= \det \left(\frac{\partial r^\nu}{\partial r^\mu} + \frac{\partial \epsilon^\nu}{\partial r^\mu} \right) \\ &= \det (\delta^\nu_\mu + \partial_\mu \epsilon^\nu) \\ &= 1 + \partial_\mu \epsilon^\mu, \end{aligned}$$

that is we have

$$\iiint_{\Omega} f dV' = \iiint_{\Omega} f (1 + \partial_{\mu} \epsilon^{\mu}) dV.$$

Note that the field, under transformation, should satisfy

$$\varphi(r) \rightarrow \varphi'(r') = \varphi(r).$$

While the derivative of the field, under transformation, should satisfies

$$\partial_{\mu} \varphi(r) \rightarrow \partial'_{\mu} \varphi'(r') = \partial'_{\mu} \varphi(r) = \frac{\partial \varphi(r)}{\partial r^{\nu}} \frac{\partial r^{\nu}}{\partial r'^{\mu}}.$$

Note here $(\partial r / \partial r')$ is the inverse of the Jacobian matrix, that is we have

$$\frac{\partial r^{\nu}}{\partial r'^{\mu}} = \delta^{\nu}_{\mu} - \partial_{\mu} \epsilon^{\nu}.$$

Thus combining we have

$$\partial'_{\mu} \varphi'(r') = (\delta^{\nu}_{\mu} - \partial_{\mu} \epsilon^{\nu}) \frac{\partial \varphi}{\partial r^{\nu}} = \frac{\partial \varphi}{\partial r^{\mu}} - \frac{\partial \varphi}{\partial r^{\nu}} \partial_{\mu} \epsilon^{\nu},$$

and thus we have

$$\begin{aligned} S' &= \iiint \int dr (1 + \partial_{\mu} \epsilon^{\mu}) \mathcal{L}(\varphi(r), \partial_{\mu} \varphi \epsilon^{\nu} \partial_{\nu} \varphi) \\ &= \iiint \int dr (1 + \partial_{\mu} \epsilon^{\mu}) \left(\mathcal{L}(\varphi(r), \partial \varphi(r)) + \frac{\delta \mathcal{L}}{\delta \partial_{\mu} \varphi} (-\partial_{\mu} \epsilon^{\nu} \partial_{\nu} \varphi(x)) \right) \\ &= \iiint \int dr \mathcal{L}(\varphi, \partial \varphi) + \left(\partial_{\mu} \epsilon^{\mu} \mathcal{L} - \partial_{\mu} \epsilon^{\nu} \frac{\delta \mathcal{L}}{\delta \partial_{\mu} \varphi} \partial_{\nu} \varphi \right) + \text{higher order terms}. \end{aligned}$$

Here we will discard the higher order terms, and rewrite

$$\begin{aligned} S' &= S + \iiint \int dr \left(\mathcal{L} \delta^{\mu}_{\nu} - \frac{\delta \mathcal{L}}{\delta \partial_{\mu} \varphi} \partial_{\nu} \varphi \right) \partial_{\mu} \epsilon^{\nu} \\ &= S - \iiint \int dr \epsilon^{\nu} \partial_{\mu} \left(\mathcal{L} \delta^{\mu}_{\nu} - \frac{\delta \mathcal{L}}{\delta \partial_{\mu} \varphi} \partial_{\nu} \varphi \right), \end{aligned} \quad (*)$$

where we have employed integration by parts, and discarded the boundary terms. Thus the conclusion here, for $S' - S = 0$ by translation symmetry, we require

$$\partial_{\mu} \left(\frac{\delta \mathcal{L}}{\delta \partial_{\mu} \varphi} \partial_{\nu} \varphi - \mathcal{L} \delta^{\mu}_{\nu} \right) = 0, \quad (2.10)$$

which we will see later that (2.10) gives a conservation law. Note further that the integral in (*) gives zero does not necessarily require (2.10) to be zero everywhere, but by principle of locality, particle far away from each other cannot communicate, thus the only way to obtain zero after integrating is (2.10) holds. Expanding (2.10) for index $\nu = 0$, which corresponds to the time translation symmetry, and noting that $\delta^i_j = 0$ for $i \neq j$, we have

$$\underbrace{\partial_t \left(\frac{\delta \mathcal{L}}{\delta \partial_t \varphi} \partial_t \varphi - \mathcal{L} \right)}_w + \underbrace{\partial_x \left(\frac{\delta \mathcal{L}}{\delta \partial_x \varphi} \partial_t \varphi \right)}_{s_x} + \underbrace{\partial_y \left(\frac{\delta \mathcal{L}}{\delta \partial_y \varphi} \partial_t \varphi \right)}_{s_y} + \underbrace{\partial_z \left(\frac{\delta \mathcal{L}}{\delta \partial_z \varphi} \partial_t \varphi \right)}_{s_z} = 0.$$

In a compact form we write

$$\partial_t w + \partial_x s_x + \partial_y s_y + \partial_z s_z = 0,$$

or even better, we have

$$\partial_t w + \nabla \cdot \vec{s} = 0 \quad (2.11)$$

with $\vec{s} = (s_x, s_y, s_z)$ being defined. Notice that the form of (2.11) coincide with the form of charge conservation law $\partial_t \rho + \nabla \cdot \vec{j} = 0$. Here one can define

$$E = \iiint\!\!\!\int dr w$$

which is the energy of the system, as seen from the fact that

$$E = \iiint dr \frac{\delta \mathcal{L}}{\delta \partial_t \varphi} \partial_t \varphi - \iiint dr \mathcal{L} = \iiint dr \frac{\delta \mathcal{L}}{\delta \partial_t \varphi} \partial_t \varphi - L = H.$$

Note: In the derivation here we have assumed that the Lagrangian density has only one field φ . However, in general \mathcal{L} can contain multiple fields φ_i , in which case (w, s_x, s_y, s_z) each include additional similar terms for each of φ_i .

Now expanding (2.10) for index $\nu = 1$, which corresponds to x -spatial translation symmetry, we can write

$$\underbrace{\partial_t \left(-\frac{\delta \mathcal{L}}{\delta \partial_t \varphi} \partial_x \varphi \right)}_{g_x} + \underbrace{\partial_x \left(-\frac{\delta \mathcal{L}}{\delta_x \partial_x \varphi} \partial_x \varphi - \mathcal{L} \right)}_{t_{xx}} + \underbrace{\partial_y \left(-\frac{\delta \mathcal{L}}{\delta \partial_y \varphi} \partial_x \varphi \right)}_{t_{xy}} + \underbrace{\partial_z \left(-\frac{\delta \mathcal{L}}{\delta \partial_z \varphi} \partial_x \varphi \right)}_{t_{zx}} = 0,$$

which yields

$$\partial_t g_x + \nabla \cdot \vec{t}_x = 0 \quad (2.12)$$

for $\vec{t}_x = (t_{xx}, t_{yx}, t_{zx})$. Now defining

$$P_x = \iiint\!\!\!\int dr g_x$$

The physical interpretation of time translation is that no matter when to do the experiment, we should be able to find the same physics law, and the interpretation of spatial translation is that no matter where we do the experiment, we should be able to find the same physics law.

Energy Momentum

Previously we have found the the action of gas density, describing a sound wave, is given by

$$S = \int dt \iiint \left(\frac{1}{2} ((\partial_t \varphi)^2 - v^2 (\nabla \varphi)^2 - \omega_0^2 \varphi^2) \right) dx dy dz.$$

From expanding (2.10) with $\nu = 0$, we have

$$\begin{aligned} w &= \frac{\delta \mathcal{L}}{\delta \partial_t \varphi} \partial_t \varphi - \mathcal{L} = \frac{1}{2} \frac{\delta (\partial_t \varphi)^2}{\delta \partial_t \varphi} \partial_t \varphi - \mathcal{L} = \partial_t \varphi \partial_t \varphi - \mathcal{L} \\ &= (\partial_t \varphi)^2 - \frac{1}{2} (\partial_t \varphi)^2 + \frac{v^2}{2} (\nabla \varphi)^2 + \frac{\omega_0^2}{2} \varphi^2 \\ &= \underbrace{\frac{1}{2} (\partial_t \varphi)^2}_{\text{KE}} + \underbrace{\frac{v^2}{2} (\nabla \varphi)^2 + \frac{\omega_0^2}{2} \varphi^2}_{\text{PE}}, \end{aligned}$$

thus we have

$$\mathcal{L} = \frac{1}{2} (\partial_t \varphi)^2 - \frac{v^2}{2} (\nabla \varphi)^2 - \frac{1}{2} \omega_0^2 \varphi^2 = \text{KE} - \text{PE}.$$

Here we can write the Hamiltonian

$$H = \iiint \left(\frac{1}{2} (\partial_t \varphi)^2 + \frac{v^2}{2} (\nabla \varphi)^2 + \frac{\omega_0^2}{2} \varphi^2 \right) dx dy dz, \quad (2.13)$$

and the Lagrangian

$$L = \iiint \left(\frac{1}{2} (\partial_t \varphi)^2 - \frac{v^2}{2} (\nabla \varphi)^2 - \frac{\omega_0^2}{2} \varphi^2 \right) dx dy dz \quad (2.14)$$

Note here the requirement of setting $c/b = -v^2$ and $a/b = -\omega_0^2$ in (2.1) ensures that H is non-negative here. Furthermore, the expansion of (2.10) in the $\nu = 0$ component gives

$$s_x = \frac{\delta \mathcal{L}}{\delta \partial_x \varphi} \partial_t \varphi = \frac{\delta \mathcal{L}}{\delta \partial_x \varphi} = -\frac{v^2}{2} \frac{\delta (\partial_x \varphi)^2}{\delta \partial_x \varphi} = -v^2 \partial_x \varphi \partial_t \varphi,$$

and similarly we obtain

$$s_y = -v^2 \partial_y \varphi \partial_t \varphi, \quad s_z = -v^2 \partial_z \varphi \partial_t \varphi,$$

which can be combined to obtain

$$\vec{s} = -v^2 \nabla \varphi \partial_t \varphi.$$

Furthermore, expansion of (2.1) also gives

$$\begin{aligned} g_x &= -\frac{\delta \mathcal{L}}{\delta \partial_t \varphi} \partial_x \varphi = -\partial_t \varphi \partial_x \varphi, \\ g_y &= -\frac{\delta \mathcal{L}}{\delta \partial_t \varphi} \partial_y \varphi = -\partial_t \varphi \partial_y \varphi, \\ g_z &= -\frac{\delta \mathcal{L}}{\delta \partial_t \varphi} \partial_z \varphi = -\partial_t \varphi \partial_z \varphi. \end{aligned}$$

Combining here we have

$$\vec{g} = -\partial_t \varphi \nabla \varphi.$$

Comparing we have

$$\vec{s} = v^2 \vec{g}. \quad (2.15)$$

Note that (2.15) is a rather generic equation. In electrodynamics, \vec{s} is called the Poynting's vector

$$\vec{s} = \frac{\vec{E} \times \vec{B}}{\mu_0},$$

and if one defines

$$\vec{g} = \epsilon_0 (\vec{E} \times \vec{B}),$$

we have

$$\vec{s} = \frac{1}{\epsilon_0 \mu_0} \vec{g} = c^2 \vec{g}.$$

Now suppose one has a plane wave solution

$$\varphi(\vec{r}, t) = A \cos(\vec{k} \cdot \vec{r} - \omega t + \theta).$$

The energy of the wave is given by, using (2.13),

$$H = \iiint \frac{1}{2} \left(A^2 \omega^2 \sin^2(\vec{k} \cdot \vec{r} - \omega t + \theta) + v^2 k^2 A^2 \sin^2(\vec{k} \cdot \vec{r} - \omega t + \theta) + \omega_0^2 A^2 \cos^2(\vec{k} \cdot \vec{r} - \omega t + \theta) \right) dx dy dz.$$

One can get the time average of H , applying dispersion relation $\omega = \sqrt{v^2 k^2 + \omega_0^2}$,

$$\langle E \rangle = \iiint \frac{1}{4} (\omega^2 + v^2 k^2 + \omega_0^2) A^2 dx dy dz = \frac{V \omega^2 A^2}{2},$$

where V is the volume of region we are interested in, thus we see here $E \propto V$, $E \propto A^2$, and $E \propto \omega^2$. Furthermore, one can also compute

$$P_x = \iiint g_x dx dy dz = - \iiint \partial_t \varphi \partial_x \varphi dx dy dz = - \iiint -\omega A^2 k_x \sin(\vec{k} \cdot \vec{r} - \omega t + \theta) dx dy dz.$$

Taking the time average we obtain the time-averaged momentum

$$\langle P_x \rangle = \frac{1}{2} \omega k_x A^2 V.$$

Similarly, it is easy to see here

$$\langle \vec{P} \rangle = \frac{1}{2} \omega \vec{k} A^2 V.$$

That is, we see that $P \propto V$, $P \propto A^2$, $P \propto kw$, and \vec{P} is parallel to \vec{k} . Furthermore, we can compare

$$\frac{\langle E \rangle}{\langle p \rangle} = \frac{(1/2) \omega^2 A^2 V}{(1/2) \omega k A^2 V} = \frac{\omega}{k},$$

such equation can be interpreted using particle-wave duality from quantum mechanics

$$\frac{\langle E \rangle}{\langle P \rangle} = \frac{(\text{energy of one particle}) \cdot (\text{number of particles})}{(\text{momentum of one particle}) \cdot (\text{number of particles})} = \frac{\hbar \omega}{\hbar k} = \frac{\omega}{k}.$$

One can define the phase velocity as

$$v_p := \frac{\omega}{k}.$$

2.2.1 Velocity

For waves, one can define the phase velocity $v_p = \omega/k$ and the group velocity as

$$v_g := \frac{d\omega}{dk}.$$

Again, using dispersion relation, $\omega = \sqrt{v^2 k^2 + \omega_0^2}$, we have

$$v_p = \frac{\sqrt{v^2 k^2 + \omega_0^2}}{k}, \quad v_g = \frac{v^2 k}{\sqrt{v^2 k^2 + \omega_0^2}}.$$

Notice here we have

$$v_p \cdot v_g = v^2.$$

Note that v is a parameter that need to be *experimentally measured*.

For electromagnetic wave, one will find that we have

$$v_p \cdot v_g = c^2,$$

as only the group velocity is physically meaningful, by the requirement that nothing can propagate faster than the speed of light, we require that $v_g \leq c$ and $v_p \geq c$.

Note further that, for massless system, $\omega_0 = 0$ as mentioned previously, we have

$$v_p = \frac{\omega}{k} = v = \frac{d\omega}{dk} = v.$$

While for massive wave, we have found above that

$$v_g < v$$

as one must have $v_p > v$.

2.2.2 Speed of Energy Flow

We have shown that

$$\frac{\langle E \rangle}{\langle P \rangle} = \frac{\omega}{k} = v_p,$$

and we have mentioned that phase velocity v_p does not have physical interpretation. What we are truly interested in is the velocity of the flow of energy.

Example: Suppose we have charged particles moving in a wire. One can draw a cross-section of the wire to obtain the current density $\vec{j} = I/A$, where I is the current and A is the cross-section area. Here we can write

$$\vec{j} = \frac{I}{A} = \frac{Q/\delta t}{A} = \frac{(vA\rho\delta t)/\delta t}{A} = \rho v,$$

where v is the velocity of the charged particles, and δt is a unit time.

Similarly, we derive the speed of the energy flow, denoted as v^* ,

$$v^* = \frac{s}{w} = \frac{gv^2}{w} = \frac{g}{w} v^2 = \frac{v^2}{v_p} = v_g$$

where s is the current density and ρ is the energy density. Eq. (2.15) is used for $s = gv^2$.

Elasticity

In this section, we will discuss the difference between scalar field and vector field, and introduce the notion of tensor. Then we will define the notion of soft and hard, and the topic of elasticity.

Let \vec{r} be a point in some solid, the solid can deform, and the point \vec{r} moves the point \vec{r}' after the deformation. Define $\vec{u} = \vec{r}' - \vec{r}$. Here \vec{u} is a field that we will study. Note immediately that \vec{u} is a vector that has three components

$$\vec{u} = u_x \hat{e}_x + u_y \hat{e}_y + u_z \hat{e}_z .$$

Thus one would expect that we have a Lagrangian density of the form

$$\mathcal{L}(u_x, u_y, u_z, \partial u_x, \partial u_y, \partial u_z) ,$$

where we have omitted the higher order terms. Thus from previous section, we can write the equations of motion

$$\partial_\mu \frac{\delta \mathcal{L}}{\delta \partial_\mu u_x} = \frac{\delta \mathcal{L}}{\delta u_x} , \quad \partial_\mu \frac{\delta \mathcal{L}}{\delta \partial_\mu u_y} = \frac{\delta \mathcal{L}}{\delta u_y} , \quad \partial_\mu \frac{\delta \mathcal{L}}{\delta \partial_\mu u_z} = \frac{\delta \mathcal{L}}{\delta u_z} ,$$

where $\mu \in \{0, 1, 2, 3\}$.

A scalar, mathematically is a number, while physically we require scalar to be invariant under rotational transformation.

A vector, mathematically is an ordered set of three numbers, while physically we require the three numbers rotate the same way as the Cartesian coordinates (x, y, z) . If R is a rotation matrix, we require that $RR^T = R^T R = 1$, and thus $R^{-1} = R^T$, implying that R is orthogonal. We further require that $\det(R) = 1$ for rotation matrices R .

Note: Here we require, for a rotation matrix R , that $R^{-1} = R^T$, simply based on the fact that the length of a vector should not be changed via rotation, that is, if we have

$$\vec{r}' = R\vec{r} ,$$

then we should have

$$\vec{r}'^T \vec{r}' = \vec{r}^T \cdot \vec{r} = \vec{r}'^T \cdot \vec{r}' = (R\vec{r})^T (R\vec{r}) = \vec{r}^T R^T R \vec{r} ,$$

which implies $R^T R = \mathbb{I}$, so $R^{-1} = R^T$. Furthermore, we see that we have

$$\det(RR^T) = \det(R) \cdot \det(R^T) = (\det(R))^2 = \det(\mathbb{I}) = 1 ,$$

thus $\det(R) = \pm 1$. Intuitively, those transformations that satisfy $R^T = R^{-1}$ and $\det(R) = -1$ are the inversion, mirror reflection, or some compositions of them. Thus we focus on the rotation matrices that satisfy $R^T = R^{-1}$ and $\det(R) = 1$.

Example: In particular, in \mathbb{R}^3 , the rotation around the x -axis by an angle θ is given by

$$R_x(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{bmatrix} .$$

Similarly, the rotations around the y -axis by an angle θ , and around the z -axis by an angle θ , are given by

$$R_y(\theta) = \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix}, \quad R_z(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

If a quantity (A_x, A_y, A_z) satisfies exactly the same transformation as the coordinate vector, then we call it a vector. Thus all vectors rotate in the same way. As we assume here the physical universe is isotropic, we require that the laws of physics stays the same under rotation. $\text{Scalar}_1 = \text{Scalar}_2$ and $\text{Vector}_1 = \text{Vector}_2$ under any rotation as they all rotate in the same way.

Now we will introduce the notion of rank-2 tensor, which can be usually viewed as 3×3 matrices,

$$T = \begin{bmatrix} A_{xx} & A_{xy} & A_{xz} \\ A_{yx} & A_{yy} & A_{yz} \\ A_{zx} & A_{zy} & A_{zz} \end{bmatrix}.$$

Note here rank-1 tensors, which are the scalars, have only 1 component. The rank-1 tensors, which are the vectors, have 3 components. The rank-2 tensors have 9 components. Then it is expected that the rank-3 tensors have 12 components.

Furthermore, rank-2 tensors also need to follow some structure under rotation, namely all rank-2 tensors rotate in the same way

$$T = RTR^T.$$

In summary, under rotation R . Scalar A transforms as

$$A \rightarrow A' = A,$$

vector A_i transforms as

$$A_i \rightarrow A'_i = R_{ii'} A_{i'},$$

rank-2 tensor transforms as

$$A_{ij} \rightarrow A'_{ij} = R_{ii'} R_{jj'} A_{i'j'},$$

and rank-3 tensor transforms as

$$A_{ijk} \rightarrow A'_{ijk} = R_{ii'} R_{jj'} R_{kk'} A_{i'j'k'}.$$

We define the outer product of vectors \vec{A} and \vec{B} , denoted as $\vec{A} \otimes \vec{B}$, via

$$\vec{A} \otimes \vec{B} = \begin{bmatrix} A_x \\ A_y \\ A_z \end{bmatrix} \begin{bmatrix} B_x & B_y & B_z \end{bmatrix} = \begin{bmatrix} A_x B_x & A_x B_y & A_x B_z \\ A_y B_x & A_y B_y & A_y B_z \\ A_z B_x & A_z B_y & A_z B_z \end{bmatrix},$$

whose components are denoted as $(AB)_{ij} = A_i B_j$, from which we see that the outer product of vectors give is rank-2 tensor. Note further that $\vec{A} \otimes \vec{B}$ contains information

about the inner product, that is the trace of $\vec{A} \otimes \vec{B}$ gives the inner product $\vec{A} \cdot \vec{B}$. Furthermore, for a matrix M , we can always decompose M into a symmetric component and an antisymmetric component

$$M = M_S + M_T = \frac{M + M^T}{2} + \frac{M - M^T}{2}.$$

The Lagrangian density \mathcal{L} , Lagrangian L , and action S all need to be scalars. Thus immediately we know that \mathcal{L} does not contain any vector. For \vec{u} defined by $\vec{u} = \vec{r}' - \vec{r}$ where \vec{r} is a point in some solid, and \vec{r}' is the same point after the solid deforms. We can write

$$\mathcal{L}(\vec{u}, \partial \vec{u}) = c(\vec{u} \cdot \vec{u}) + a(\partial_t \vec{u} \cdot \partial_t \vec{u}) + b_1(\text{tr}(e))^2 + b_2(\text{tr}(ee)) \quad (2.16)$$

where e is the strain tensor defined in the following discussion, $c, a, b_1, b_2 \in \mathbb{R}$ are some constants. Note that higher order terms and constant term have been discarded here, the linear terms u_x, u_y, \dots are not allowed under symmetry, and the terms $\partial_x u, \partial_y u, \dots$ are discarded as they are total derivative which integrates to boundary conditions. The term $\vec{u} \cdot \partial_t \vec{u} = (1/2)\partial_t(\vec{u} \cdot \vec{u})$ is discarded as it integrates to boundary condition of the system. The term $\vec{u} \cdot (\nabla \times \vec{u})$, for i being spacial coordinate, is not allowed by argument of space inversion. The terms $\partial_t \vec{u} \cdot \partial_i \vec{u}$ is not allowed by time-reversal symmetry and space-inversion symmetry. Notice that $D\vec{u}$ is a Jacobian

$$\begin{aligned} D\vec{u} &= \begin{bmatrix} \partial_x u_x & \partial_x u_y & \partial_x u_z \\ \partial_y u_x & \partial_y u_y & \partial_y u_z \\ \partial_z u_x & \partial_z u_y & \partial_z u_z \end{bmatrix} := e + \eta \\ &= \begin{bmatrix} \partial_x u_x & \frac{\partial_x u_y + \partial_y u_x}{2} & \frac{\partial_x u_z + \partial_z u_x}{2} \\ \frac{\partial_x u_y + \partial_y u_x}{2} & \partial_y u_y & \frac{\partial_y u_z + \partial_z u_y}{2} \\ \frac{\partial_x u_z + \partial_z u_x}{2} & \frac{\partial_y u_z + \partial_z u_y}{2} & \partial_z u_z \end{bmatrix} + \begin{bmatrix} 0 & \frac{\partial_x u_y - \partial_y u_x}{2} & \frac{\partial_x u_z - \partial_z u_x}{2} \\ \frac{\partial_x u_y - \partial_y u_x}{2} & 0 & \frac{\partial_y u_z - \partial_z u_y}{2} \\ \frac{\partial_x u_z - \partial_z u_x}{2} & \frac{\partial_y u_z - \partial_z u_y}{2} & 0 \end{bmatrix}, \end{aligned}$$

where e is called the strain tensor, together with the anti-symmetric part η of $D\vec{u}$, we have $D\vec{u} = e + \eta$, and the components of e and η are characterized by

$$e_{ij} = \frac{\partial_i u_j + \partial_j u_i}{2}, \quad \eta_{ij} = \frac{\partial_i u_j - \partial_j u_i}{2}.$$

Also notice here the terms in the upper triangle on the right of η constitutes the curl of \vec{u}

$$\frac{\nabla \times \vec{u}}{2} = \left(\frac{\partial_y u_z - \partial_z u_y}{2}, -\frac{\partial_x u_z - \partial_z u_x}{2}, \frac{\partial_x u_y - \partial_y u_x}{2} \right).$$

If we consider rotating a crystal by a small angle $\delta\theta$ around an axis \vec{n} . Then we have, approximately,

$$\vec{r} \rightarrow \vec{r}' = \vec{r} + \delta\theta \vec{n} \times \vec{r},$$

The displacement $\vec{u} = \vec{r}' - \vec{r}$ is characterized by

$$\vec{u} = \delta\theta \vec{n} \times \vec{r}.$$

Furthermore, the crystal has the same energy before and after rotation, that is its Lagrangian stays the same, $L = L'$, before and after the rotation.

Back to the general case, we then can write

$$\nabla \times \vec{u} = \nabla \times (\vec{n} \times \vec{r}) \delta\theta = (\vec{n}(\nabla \cdot \vec{r}) - \vec{r}(\nabla \cdot \vec{n})) \delta\theta = 3\vec{n} \delta\theta,$$

suggesting that $\nabla \times \vec{u}$ is non-zero when an angle $\delta\theta$ exists. As argued above, the solid should have the same Lagrangian before and after rotational transformation, thus \mathcal{L} should not be dependent on η . Thus the only allowed terms are $(\text{tr}(e))^2$ and $\text{tr}(ee)$.

Consider two points $\vec{r} = (x, y, z)$ and $\vec{r} + d\vec{r} = (x + dx, y + dy, z + dz)$ in a solid, the squared distance between the two is given by

$$dr^2 = dx^2 + dy^2 + dz^2.$$

If one deforms the solid a little, such that \vec{r} goes to \vec{r}' , and $\vec{r} + d\vec{r}$ goes to $\vec{r}' + d\vec{r}'$, then the distance between the two now becomes

$$(d\vec{r}')^2 = d\vec{r}^2 + 2e_{ij} dr_i dr_j + (\text{higher order terms}).$$

Thus we see here e_{ij} describes the change of distance between the two points. Furthermore, for a spring, the elastic energy only depends on the change of length of the spring, thus we can write

$$U = \frac{1}{2} C \Delta L^2.$$

Thus for a solid, it is in nature to expect that the elastic energy depends on the change of length between two points

$$U = \int d\vec{r} \sum_{ijkl} \frac{1}{2} C_{ijkl} e_{ij}(\vec{r}) e_{kl}(\vec{r}).$$

While the components C_{ijkl} here should satisfy some symmetries

$$C_{ijkl} = C_{klij}, \quad C_{ijkl} = C_{jikl}, \quad C_{ijkl} = C_{ijlk}.$$

C_{ijkl} here is in fact a rank-4 tensor, called the elastic modulus tensor. Each component of C_{ijkl} is known as an elastic stiffness constant. Here C_{ijkl} has 81 components, but in reality only 21 of them are independent. As we know that $e_{ij} = e_{ji}$, one can show that $C_{ijkl} = C_{jikl} = C_{ijlk}$, and by definition of U , we also require $C_{ijkl} = C_{klij}$, which are the symmetry requirements shown. Note further that if one choose proper axes, one can make three of the 21 elastic constants 0, thus in reality we have only 18 independent constants.

It is not hard that one can calculate

$$\begin{aligned} (\text{tr}(e))^2 &= e_{xx}^2 + e_{yy}^2 + e_{zz}^2 + 2(e_{xx}e_{yy} + e_{xx}e_{zz} + e_{yy}e_{zz}) = (e_{ii})^2, \\ \text{tr}(ee) &= e_{xx}^2 + e_{yy}^2 + e_{zz}^2 + 2(e_{xy}e_{xy} + e_{xz}e_{xz} + e_{yz}e_{yz}) = e_{ij}e_{ji}. \end{aligned}$$

Furthermore, c in (2.16) should also vanish by translation symmetry and the fact that \mathcal{L} is closely related to the total energy of the system. Thus rewritten (2.16) here, we obtain

$$\mathcal{L} = \frac{\rho}{2} (\partial_t \vec{u} \cdot \partial_t \vec{u}) - \frac{\lambda}{2} (\text{tr}(e))^2 - G \text{tr}(ee)$$

where ρ is called the mass density. The energy density of the system can be described by

$$\begin{aligned} w &= \frac{\delta \mathcal{L}}{\delta \partial_t u_i} \partial_t u_i - \mathcal{L} = \frac{\delta \mathcal{L}}{\delta \partial_t u_x} + \frac{\delta \mathcal{L}}{\delta \partial_t u_x} + \frac{\delta \mathcal{L}}{\delta \partial_t u_x} - \mathcal{L} \\ &= \frac{\rho}{2} \partial_t \vec{u} \cdot \partial_t \vec{u} + \frac{\lambda}{2} (\text{tr}(e))^2 + G(\text{tr}(ee)). \end{aligned}$$

Comparing with \mathcal{L} we observe that, the kinetic energy term and elastic potential term in w is given by

$$\text{KE term} = \rho \partial_t \vec{u} \cdot \partial_t \vec{u} \quad \text{PE term} = \frac{\lambda}{2} (\text{tr}(e))^2 + G(\text{tr}(ee)).$$

Thus the total elastic energy can be obtained by computing

$$U = \iiint \left(\frac{\lambda}{2} (\text{tr}(e))^2 + G(\text{tr}(ee)) \right) dx dy dz.$$

Equivalently, one show that U is equivalent to

$$U = \iiint \left(\frac{1}{2} B (\text{tr}(e))^2 + G \text{tr} \left(\left(e - \frac{\text{tr}(e)}{3} \mathbb{I} \right) \left(e - \frac{\text{tr}(e)}{3} \mathbb{I} \right) \right) \right) dx dy dz \quad (2.17)$$

where \mathbb{I} is the 3×3 identity matrix, the form of U given by (2.17) is more often used.

2.3.1 Types of solid deformation

One type of deformation is the bulk deformation, which changes only the volume of the solid, but not the shape of the solid, then the change in energy in such type of deformation is governed by the coefficient B in (2.17).

Another type of deformation is the shear deformation, which changes the shape of the solid, but not the volume, then the change in energy in such type of deformation is governed by the coefficient G in (2.17).

For a spring, the deformation is governed by the spring constant k . When k is large, the spring is said to be hard as it takes a huge amount of energy to deform the spring. While when k is small, the spring is said to be soft. On the other hand, solid with a large B and large G is said to be hard. Solids with either large B and small G , small B and large G , or small G and small B , are all said to be soft.

Example: Liquid usually has $G \sim 0$. Rubber has $B \sim 1$ GPa, $G \sim 10^{-3}$ GPa. Solid crystal has $B \sim 10$ GPa, and $G \sim B$. There exists materials with small B and higher G , such as the auxetic materials.

One can introduce the Poisson's ratio. Suppose one stretch a cubic material along the x -axis, denote the change of length of the material along the x -axis as ΔL , and the change of the length of the material along the y -axis as $\Delta L'$, we define

$$\nu = \frac{\Delta L'}{\Delta L},$$

which is a quantity that is usually larger than 0 for most of the materials. For some materials that have $B \ll G$, ν will become negative, such as the cork wood used for wine stopper. Rigorously, the Possion's ratio is defined by

$$\nu = \frac{3B - 2G}{2(3B + G)}.$$

Previously we have derived the Lagrangian density for solid deformation (2.16), here we expand the Lagrangian density,

$$\begin{aligned} \mathcal{L} = & \frac{\rho}{2} (\partial_t u_x)^2 + \frac{\rho}{2} (\partial_t u_y)^2 + \frac{\rho}{2} (\partial_t u_z)^2 \\ & - \left(\frac{B}{2} + \frac{2}{3}G \right) ((\partial_x u_x)^2 + (\partial_y u_y)^2 + (\partial_z u_z)^2) \\ & - \left(B - \frac{2G}{3} \right) (\partial_x u_x \partial_y u_y + \partial_x u_x \partial_z u_z + \partial_y u_y \partial_z u_z) \\ & - 2G \left(\left(\frac{\partial_x u_y + \partial_y u_x}{2} \right)^2 + \left(\frac{\partial_y u_z + \partial_z u_y}{2} \right)^2 + \left(\frac{\partial_x u_z + \partial_z u_x}{2} \right)^2 \right). \end{aligned}$$

The equation of motion (2.4) thus reads

$$\partial_t \frac{\delta \mathcal{L}}{\delta \partial_t u_x} + \partial_x \frac{\delta \mathcal{L}}{\delta \partial_x u_x} + \partial_y \frac{\delta \mathcal{L}}{\delta \partial_y u_x} + \partial_z \frac{\delta \mathcal{L}}{\delta \partial_z u_x} = \frac{\delta \mathcal{L}}{\delta u_x}, \quad (2.18)$$

$$\partial_t \frac{\delta \mathcal{L}}{\delta \partial_t u_y} + \partial_x \frac{\delta \mathcal{L}}{\delta \partial_x u_y} + \partial_y \frac{\delta \mathcal{L}}{\delta \partial_y u_y} + \partial_z \frac{\delta \mathcal{L}}{\delta \partial_z u_y} = \frac{\delta \mathcal{L}}{\delta u_y}, \quad (2.19)$$

$$\partial_t \frac{\delta \mathcal{L}}{\delta \partial_t u_z} + \partial_x \frac{\delta \mathcal{L}}{\delta \partial_x u_z} + \partial_y \frac{\delta \mathcal{L}}{\delta \partial_y u_z} + \partial_z \frac{\delta \mathcal{L}}{\delta \partial_z u_z} = \frac{\delta \mathcal{L}}{\delta u_z}. \quad (2.20)$$

As the three equations are of the same form, we focus on (2.18). It is clear that the RHS of (2.18) vanishes, and the LHS of (2.18) gives

$$0 = \rho \partial_t^2 u_x - \left(B + \frac{4G}{3} \right) \partial_x^2 u_x - \left(B - \frac{2G}{3} \right) (\partial_x \partial_y u_y + \partial_x \partial_z u_z) - G (\partial_x \partial_y u_y + \partial_y^2 u_x) - G (\partial_x \partial_z u_z + \partial_z^2 u_x)$$

as we have

$$\frac{\delta \mathcal{L}}{\delta \partial_t u_x} = (\rho \partial_t u_x),$$

$$\begin{aligned} \frac{\delta \mathcal{L}}{\delta \partial_x u_x} &= \frac{\delta}{\delta \partial_x u_x} \left(\left(\frac{B}{2} + \frac{2G}{3} \right) (\partial_x u_x)^2 + \left(B - \frac{2}{3}G \right) (\partial_x u_x \partial_y u_y + \partial_x u_x \partial_z u_z) \right) \\ &= - \left(B + \frac{4}{3}G \right) \partial_x u_x - \left(B - \frac{2G}{3} \right) (\partial_y u_y + \partial_z u_z), \end{aligned}$$

$$\frac{\delta \mathcal{L}}{\delta \partial_y u_x} = -\frac{G}{2} \frac{\delta}{\delta \partial_x u_y} (\partial_x u_y + \partial_y u_x)^2 = -G (\partial_x u_y + \partial_y u_x).$$

$$\frac{\delta \mathcal{L}}{\delta \partial_z u_x} = -\frac{G}{2} \frac{\delta}{\delta \partial_x u_z} (\partial_x u_z + \partial_z u_x)^2 = -G (\partial_x u_z + \partial_z u_x).$$

Grouping the terms we obtain

$$\rho \partial_t^2 u_x = \left(\left(B + \frac{4G}{3} \right) \partial_x^2 u_x + G \partial_y^2 u_x + G \partial_z^2 u_x \right) + \left(B + \frac{G}{3} \right) \partial_x \partial_y u_y + \left(B + \frac{G}{3} \right) \partial_x \partial_z u_z. \quad (2.21)$$

Similarly, for (2.19) and (2.20), we obtain

$$\rho \partial_t^2 u_y = \left(B + \frac{G}{3} \right) \partial_x \partial_y u_x + \left(\left(B + \frac{4G}{3} \right) \partial_y^2 u_y + G \partial_x^2 u_y + G \partial_z^2 u_y \right) + \left(B + \frac{G}{3} \right) \partial_y \partial_z u_z, \quad (2.22)$$

$$\rho \partial_t^2 u_z = \left(B + \frac{G}{3} \right) \partial_x \partial_z u_x + \left(B + \frac{G}{3} \right) \partial_y \partial_z u_y + \left(\left(B + \frac{4G}{3} \right) \partial_z^2 u_z + G \partial_x^2 u_z + G \partial_y^2 u_z \right). \quad (2.23)$$

Intuitively, we believe that the solution to (2.21 - 2.23) is a plane wave

$$\vec{u} = \vec{A} e^{i\vec{k} \cdot \vec{r} - i\omega t}, \quad (2.24)$$

or in component form

$$u_x = A_x e^{i\vec{k} \cdot \vec{r} - i\omega t}, \quad u_y = A_y e^{i\vec{k} \cdot \vec{r} - i\omega t}, \quad u_z = A_z e^{i\vec{k} \cdot \vec{r} - i\omega t}.$$

It is not hard to see that (2.24) satisfies

$$\partial_x \vec{u} = i k_x \vec{u}, \quad \partial_y \vec{u} = i k_y \vec{u}, \quad \partial_z \vec{u} = i k_z \vec{u}, \quad \partial_t \vec{u} = -i\omega \vec{u}.$$

Thus (2.21) gives

$$\rho \omega^2 u_x = \left(\left(B + \frac{4G}{3} \right) k_x^2 + G k_y^2 + G k_z^2 \right) u_x + \left(B + \frac{G}{3} \right) k_x k_y u_y + \left(B + \frac{G}{3} \right) k_x k_z u_z.$$

Similarly, (2.22) and (2.23) give

$$\rho \omega^2 u_y = \left(B + \frac{G}{3} \right) k_x k_y u_x + \left(\left(B + \frac{4G}{3} \right) k_y^2 + G k_x^2 + G k_z^2 \right) u_y + \left(B + \frac{G}{3} \right) k_y k_z u_z.$$

$$\rho \omega^2 u_z = \left(B + \frac{G}{3} \right) k_x k_z u_x + \left(B + \frac{G}{3} \right) k_y k_z u_y + \left(\left(B + \frac{4G}{3} \right) k_z^2 + G k_x^2 + G k_y^2 \right) u_z.$$

In matrix form, the coupled differential equations can be written as

$$\rho \omega^2 \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix} = \begin{bmatrix} \left(B + \frac{4G}{3} \right) k_x^2 + G k_y^2 + G k_z^2 & \left(B + \frac{G}{3} \right) k_x k_y & \left(B + \frac{G}{3} \right) k_x k_z \\ \left(B + \frac{G}{3} \right) k_x k_y & G k_x^2 + \left(B + \frac{4G}{3} \right) k_y^2 + G k_z^2 & \left(B + \frac{G}{3} \right) k_y k_z \\ \left(B + \frac{G}{3} \right) k_x k_z & \left(B + \frac{G}{3} \right) k_y k_z & G k_x^2 + G k_y^2 + \left(B + \frac{4G}{3} \right) k_z^2 \end{bmatrix} \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix}.$$

We observe from (2.24) that the only *difference* between u_x , u_y , and u_z is characterized by A_x , A_y , and A_z , thus the system becomes

$$\rho \omega^2 \begin{bmatrix} A_x \\ A_y \\ A_z \end{bmatrix} = \begin{bmatrix} \left(B + \frac{4G}{3} \right) k_x^2 + G k_y^2 + G k_z^2 & \left(B + \frac{G}{3} \right) k_x k_y & \left(B + \frac{G}{3} \right) k_x k_z \\ \left(B + \frac{G}{3} \right) k_x k_y & G k_x^2 + \left(B + \frac{4G}{3} \right) k_y^2 + G k_z^2 & \left(B + \frac{G}{3} \right) k_y k_z \\ \left(B + \frac{G}{3} \right) k_x k_z & \left(B + \frac{G}{3} \right) k_y k_z & G k_x^2 + G k_y^2 + \left(B + \frac{4G}{3} \right) k_z^2 \end{bmatrix} \begin{bmatrix} A_x \\ A_y \\ A_z \end{bmatrix},$$

which can be written in a simple form

$$\rho\omega^2\vec{A} = D\vec{A}, \quad (2.25)$$

with matrix D referred as the dynamic matrix, which satisfies $D = D^\dagger$. We notice here (2.25) characterizes a eigenvalue problem. (2.25) can be solved via

$$(\rho\omega^2\mathbb{I} - D)\vec{A} = 0.$$

Solving the characteristic equation

$$\det(\rho\omega^2\mathbb{I} - D) = 0$$

gives the result

$$\omega_1 = \sqrt{\frac{G}{\rho}}k, \quad \omega_2 = \sqrt{\frac{G}{\rho}}k, \quad \omega_3 = \sqrt{\frac{B + 4G/3}{\rho}}k,$$

where $k = \sqrt{k_x^2 + k_y^2 + k_z^2}$, with their corresponding eigenvectors

$$\vec{A}_1 = \begin{bmatrix} -k_z \\ 0 \\ k_x \end{bmatrix}, \quad \vec{A}_2 = \begin{bmatrix} -k_y \\ k_x \\ 0 \end{bmatrix}, \quad \vec{A}_3 = \begin{bmatrix} k_x \\ k_y \\ k_z \end{bmatrix}.$$

Here \vec{A}_1 and \vec{A}_2 are called the transverse modes, or transverse phonons, and \vec{A}_3 is called the longitudinal mode, or longitudinal phonon. Here we observe that

$$\vec{A}_1 \cdot \vec{k} = \vec{A}_2 \cdot \vec{k} = 0$$

hence the name *transverse* for \vec{A}_1 and \vec{A}_2 . While we observe that

$$\vec{A}_3 \parallel \vec{k},$$

hence the name *longitudinal* for \vec{A}_3 . Note further that \vec{A}_3 mode always has higher frequency ω_3 than the transverse modes by the fact that $G, B > 0$. From that phase velocity $v_p = \omega/k$, and we have assumed here the wave is massless, thus group velocity $v_g = v_p = \omega/k$, we obtain

$$v_L = \sqrt{\frac{B + 4G/3}{\rho}}, \quad v_T = \sqrt{\frac{G}{\rho}}.$$

thus we see here $v_L > v_T$, the speed of longitudinal wave is always larger than that of the traverse wave.

Remark: Here we see that, for solid deformation, there can be transverse modes and longitudinal mode. However, for liquid and gas, we have $G = 0$, that is there is no transverse mode for liquid and gas deformation, only longitudinal wave survives.

2.3.2 Massive wave

Here we break the translation symmetry and add back the $u^2 = \vec{u} \cdot \vec{u}$ term to the Lagrangian density

$$\mathcal{L} = \frac{\rho}{2} \partial_t \vec{u} \partial_t \vec{u} - \frac{\lambda}{2} (\text{tr}(e))^2 - G \text{tr}(ee) - \frac{\rho}{2} \omega_0^2 u^2.$$

In this case, RHS of (2.18), (2.19), and (2.20) gives

$$\frac{\delta \mathcal{L}}{\delta u_x} = -\rho \omega_0^2 u_i,$$

and the system (2.25) thus become

$$\rho^2 \omega^2 \vec{A} = D \vec{A} + \rho \omega_0^2 \vec{A},$$

or rearranging we obtain

$$\rho(\omega^2 - \omega_0^2) \vec{A} = D \vec{A}. \quad (2.26)$$

Notice that the dynamic matrix D is unchanged, while the eigenvalue problem has been modified to include the $-\rho \omega_0^2$ term. The eigenvalues and their corresponding eigenvectors, in this case, are then given by

$$\begin{aligned} Gk^2 &= \rho(\omega_1^2 - \omega_0^2) & \text{with } \vec{A}_1 &= (-k_z, 0, k_x), \\ Gk^2 &= \rho(\omega_2^2 - \omega_0^2) & \text{with } \vec{A}_2 &= (-k_y, k_x, 0), \\ (B + 4G/3)k^2 &= \rho(\omega_3^2 - \omega_0^2) & \text{with } \vec{A}_3 &= (k_x, k_y, k_z). \end{aligned}$$

Solving we obtain the frequencies

$$\begin{aligned} \omega_1 &= \sqrt{\frac{G}{\rho} k^2 + \omega_0^2} = \sqrt{v_T^2 k^2 + \omega_0^2}, \\ \omega_2 &= \sqrt{\frac{G}{\rho} k^2 + \omega_0^2} = \sqrt{v_T^2 k^2 + \omega_0^2}, \\ \omega_3 &= \sqrt{\frac{B + 4G/3}{\rho} k^2 + \omega_0^2} = \sqrt{v_L^2 k^2 + \omega_0^2}. \end{aligned}$$

As derived above, we have

$$v_p v_g = v_T^2, \quad v_g < v_T \text{ for transverse wave,} \quad v_p v_g = v_L^2, \quad v_g < v_L \text{ for longitudinal wave.}$$

As derived before, appending the u^2 term, we have the energy density of the system

$$\begin{aligned}
w &= \frac{\rho}{2} \partial_t \vec{u} \cdot \partial_t \vec{u} + \frac{\lambda}{2} (\text{tr}(e))^2 + G \text{tr}(ee) + \frac{\rho}{2} \omega_0^2 u^2 \\
&= \frac{\rho}{2} ((\partial_t u_x)^2 + (\partial_t u_y)^2 + (\partial_t u_z)^2) \\
&\quad + \left(\frac{B}{2} + \frac{2G}{3} \right) \left(\left(\frac{\partial u_x}{\partial y} \right)^2 + \left(\frac{\partial u_y}{\partial y} \right)^2 + \left(\frac{\partial u_z}{\partial z} \right)^2 \right) \\
&\quad + \left(B + \frac{2G}{3} \right) \left(\frac{\partial u_x}{\partial x} \frac{\partial u_y}{\partial y} + \frac{\partial u_x}{\partial x} \frac{\partial u_z}{\partial z} + \frac{\partial u_y}{\partial y} \frac{\partial u_z}{\partial z} \right) \\
&\quad + \frac{G}{2} \left(\left(\frac{\partial u_x}{\partial y} \right)^2 + \left(\frac{\partial u_y}{\partial y} \right)^2 + \left(\frac{\partial u_x}{\partial z} \right)^2 + \left(\frac{\partial u_z}{\partial x} \right)^2 + \left(\frac{\partial u_y}{\partial z} \right)^2 + \left(\frac{\partial u_z}{\partial y} \right)^2 \right) \\
&\quad + G \left(\frac{\partial u_x}{\partial y} \frac{\partial u_y}{\partial x} + \frac{\partial u_x}{\partial z} \frac{\partial u_z}{\partial x} + \frac{\partial u_y}{\partial z} \frac{\partial u_z}{\partial y} \right) \\
&\quad + \frac{\rho}{2} \omega_0^2 (u_x^2 + u_y^2 + u_z^2) .
\end{aligned}$$

Here we assume here

$$\vec{u} = \Re \left(\vec{A} e^{i\vec{k} \cdot \vec{r} - i\omega t} \right) = \vec{A} \cos(\vec{k} \cdot \vec{r} - \omega t + \varphi) ,$$

where ω is given, as above, $\omega = \sqrt{v^2 k^2 + \omega_0^2}$ for the three modes with $v = v_T$ or $v = v_L$, and \vec{A} being the corresponding eigenvector. Combining all we obtain

$$\begin{aligned}
w &= G (k_y^2 A_x^2 + k_x^2 A_y^2 + k_y^2 A_z^2 + k_z^2 A_y^2 + k_x^2 A_z^2 + k_z^2 A_x^2 + 2k_x k_y A_x A_y + 2k_x k_z A_x A_z + 2k_y k_z A_y A_z) \sin^2(\mathcal{M})/2 \\
&\quad + (\rho/2) \omega^2 \sin^2(\mathcal{M}) (A_x^2 + A_y^2 + A_z^2) + (B/2 + 2G/3) (k_x^2 A_x^2 + k_y^2 A_y^2 + k_z^2 A_z^2) \sin^2(\mathcal{M}) \\
&\quad + (B - 2G/3) (k_x k_y A_x A_y + k_x k_z A_x A_z + k_y k_z A_y A_z) \sin^2(\mathcal{M}) \\
&\quad + (\rho/2) \omega_0^2 (A_x^2 + A_y^2 + A_z^2) \cos^2(\mathcal{M}) ,
\end{aligned}$$

where $\mathcal{M} = \vec{k} \cdot \vec{r} - \omega t + \varphi$. Now taking the time average, we obtain

$$2\langle w \rangle = \frac{\rho}{2} (\omega^2 + \omega_0^2) |\vec{A}|^2 + \frac{1}{2} \vec{A}^T D \vec{A} ,$$

where D is again the dynamic matrix. For the three eigenmodes of the system, we thus obtain that

$$2\langle w \rangle = \frac{\rho}{2} (\omega^2 + \omega_0^2) A^2 + \frac{\rho}{2} (\omega^2 - \omega_0^2) A^2 = \rho \omega^2 A^2 ,$$

or concluding here

$$\langle w \rangle = \frac{\rho \omega^2 A^2}{2} .$$

The total energy of the system is then

$$\langle E \rangle = \iiint dx dy dz \frac{\rho \omega^2 A^2}{2} = \frac{\rho \omega^2 A^2 V}{2} , \quad (2.27)$$

that is we observe that $E \propto V$, $E \propto A^2$, and $E \propto \omega^2$. As computed previously for scalar field, we have

$$\langle E \rangle = \frac{\omega^2 A^2 V}{2} \text{ for scalar field ,}$$

which differ from (2.27) by a factor of ρ , and that is because the ρ factor in the scalar field energy has been absorbed in the transformation that we used for obtaining (2.1).

Rotational Symmetry

Here we assume small angle rotation

$$\vec{r} \rightarrow \vec{r}' = R\vec{r} = \vec{r} + \vec{n} \times \vec{r} \delta\theta,$$

where \vec{n} is a unit vector representing the axis of rotation and $\delta\theta$ is the angle rotated. Here more conveniently, we can write

$$r^\nu \rightarrow r'^\nu = r^\nu + \frac{\delta\epsilon^\nu}{\delta\theta^j} \delta\theta^j. \quad (2.28)$$

Note that we have

$$(\vec{n} \times \vec{r} \delta\theta)_i = \epsilon_{ijk} \delta\theta^j r^k$$

where ϵ_{ijk} is the Levi-Civita symbol. Comparing we obtain

$$\frac{\delta\epsilon^0}{\delta\theta^j} = 0, \quad \frac{\delta\epsilon^i}{\delta\theta^j} = \epsilon_{ijk} r^k \text{ for } i \neq 0.$$

Before the rotation, we write the action as

$$S = \iiint \mathcal{L}(\phi_a(\vec{r}), \partial_\mu \phi_a(\vec{r})) dt dx dy dz,$$

and after the rotation, we write the action as

$$S' = \iiint \mathcal{L}(\phi'_a(\vec{r}'), \partial'_\mu \phi'_a(\vec{r}')) dt' dx' dy' dz'.$$

We have information about the transformation from r to r' , but not for the transformation from ϕ to ϕ' . For scalar field $\phi_a(r)$, after rotation we should have $\phi_a(r) = \phi'_a(r')$. While for vector field, we should write

$$\phi_a(\vec{r}) \rightarrow \phi'_a(\vec{r}') = \phi_a(\vec{r}) + \frac{\delta\phi_a}{\delta\theta^j} \Big|_{\vec{r}} \delta\theta^j,$$

and thus the derivative of ϕ_a transforms as

$$\partial_\mu \phi_a(\vec{r}) \rightarrow \partial'_\mu \phi'_a(\vec{r}') = \partial'_\mu \phi_a(\vec{r}) + \partial'_\mu \left(\frac{\delta\phi_a}{\delta\theta^j} \right) \Big|_{\vec{r}} \delta\theta^j.$$

Note here we can write

$$\begin{aligned} \partial'_\mu \phi_a(\vec{r}) &= \partial_\nu \phi_a(\vec{r}) \frac{\partial r^\nu}{\partial r'^\mu} \\ &= \partial_\nu \phi_a(\vec{r}) \left(\frac{\partial r'^\mu}{\partial r^\nu} \right)^{-1} \\ &= \partial_\nu \phi_a(\vec{r}) \left(\frac{\partial r^\mu}{\partial r^\nu} + \frac{\partial \epsilon^\mu}{\partial r^\nu} \right)^{-1} \\ &= \partial_\nu \phi_a(\vec{r}) \delta^\nu_\mu - \partial_\nu \phi_a(\vec{r}) \frac{\partial \epsilon^\nu}{\partial r^\mu} \\ &= \partial_\mu \phi_a(\vec{r}) - \partial_\nu \phi_a(r) \partial_\mu \epsilon^\nu \\ &= \partial_\mu \phi_a(\vec{r}) - \partial_\nu \phi_a \partial_\mu \epsilon^\nu + \partial_\mu \left(\frac{\delta\phi_a}{\delta\theta^j} \right) \delta\theta^j \\ &= \partial_\mu \phi_a(\vec{r}) - \partial_\nu \phi_a \partial_\mu \left(\frac{\delta\epsilon^\mu}{\delta\theta^j} \delta\theta^j \right) + \partial_\mu \left(\frac{\delta\phi_a}{\delta\theta^j} \delta\theta^j \right). \end{aligned}$$

Noting here the Jacobian is given by

$$\frac{dr}{dr'} = \left(1 + \partial_\mu \left(\frac{\delta \epsilon^\mu}{\delta \theta^j} \delta \theta^j \right) \right)$$

Now we obtain

$$\begin{aligned} S' &= \iiint\!\!\!\int dr' \mathcal{L}(\phi'(r'), \partial'_\mu \phi'(r')) \\ &= \iiint\!\!\!\int dr \left(1 + \partial_\mu \left(\frac{\delta \epsilon^\mu}{\delta \theta^j} \delta \theta^j \right) \right) \mathcal{L} \left(\phi_a(\vec{r}) + \frac{\delta \phi_a}{\delta \theta^j} \delta \theta^j, \partial_\mu \phi_a(\vec{r}) - \partial_\nu \phi_a \partial_\mu \left(\frac{\delta \epsilon^\mu}{\delta \theta^j} \delta \theta^j \right) + \partial_\mu \left(\frac{\delta \phi_a}{\delta \theta^j} \delta \theta^j \right) \right) \\ &= \iiint\!\!\!\int dr \mathcal{L}(\phi, \partial \phi) + \underbrace{\iiint\!\!\!\int dr (\text{terms of } \delta \theta^j) \delta \theta_j}_{\text{vanishes by } \delta S = 0} + \iiint\!\!\!\int dr (\text{terms of } \partial \delta \theta_j) \partial_\mu \delta \theta^j + \text{H.O.T.} \end{aligned}$$

where we have applied Taylor expansion. Neglecting the higher order terms, we have

$$S' = S + \iiint\!\!\!\int dr \left(\frac{\delta \epsilon^\mu}{\delta \theta^j} \mathcal{L}(\phi, \partial \phi) + \frac{\delta \mathcal{L}}{\delta \partial_\mu \phi} \left(-\partial_\nu \phi \frac{\delta \epsilon^\mu}{\delta \theta^j} + \frac{\delta \phi}{\delta \theta^j} \right) \right) \partial_\mu \theta^j.$$

Rotation symmetry states that we should have $S = S'$. Here we define

$$j^\mu := -\frac{\delta \epsilon^\mu}{\delta \theta^j} \mathcal{L}(\phi, \partial \phi) - \frac{\delta \mathcal{L}}{\delta \partial_\mu \phi} \left(\frac{\delta \phi}{\delta \theta^j} - \partial_\nu \phi \frac{\delta \epsilon^\mu}{\delta \theta^j} \right),$$

and thus rotational symmetry requires

$$0 = \iiint\!\!\!\int dr j^\mu \partial_\mu \theta^j = \iiint\!\!\!\int dr (\partial_\mu j^\mu) \theta^j.$$

Concluding we obtain the equation

$$\partial_t j^0 + \nabla \cdot \vec{j} = 0 \quad (2.29)$$

which characterizes the conservation law obtained from rotational symmetry.

2.4.1 General form of Noether's Theorem

Note here, in the discussion above, the procedure of deriving the conservation law (2.29) does not involve the rotational property of $\delta \theta^j$, that is, we can replace $\delta \theta^j$ by δw^j for arbitrary small deformation. In general, for a continuous transformation, such as translation or rotation discussed above, characterizes by $r \rightarrow r + \epsilon$, where ϵ is a small quantity, we can write

$$r^\mu \rightarrow r'^\mu = r^\mu + \frac{\delta \epsilon^\mu}{\delta w^b} \delta w^b, \quad \phi_a(r) \rightarrow \phi'_a(r') = \phi_a(r) + \frac{\delta \phi_a}{\delta w^b} \delta w^b.$$

Following a similar procedure as above for the rotational symmetry, we can obtain a conservation law

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

another form of writing (2.29), where j^μ is in general defined by

$$j^\mu := \left(\frac{\delta \mathcal{L}}{\delta \partial_\mu \phi_a} \partial_\nu \phi_a - \delta^\mu_\nu \mathcal{L} \right) \frac{\delta \epsilon^\nu}{\delta w^b} - \frac{\delta \mathcal{L}}{\delta \partial_\mu \phi_a} \frac{\delta \phi_a}{\delta w^b}.$$

Example: Revisit of time translation. We consider the case $\delta w^b = \delta t$. That is, we write

$$t \rightarrow t' = t + \delta t, \quad x \rightarrow x' = x, \quad y \rightarrow y' = y, \quad z \rightarrow z' = z.$$

Then it is immediate that we have

$$\delta^\mu_0 = \frac{\delta \epsilon^\mu}{\delta w^b}.$$

Here we have further that $\phi(r) \rightarrow \phi'(r') = \phi(r) + 0 \cdot \delta t$ as translation does not change the field, thus we see

$$\frac{\delta \phi}{\delta w^b} = 0.$$

Then the conserved quantity for time translation is given by

$$j^\mu = \left(\frac{\delta \mathcal{L}}{\delta \partial_\mu \phi} \partial_\nu \phi - \mathcal{L} \delta^\mu_\nu \right) \delta^\mu_0 = \frac{\delta \mathcal{L}}{\delta \partial_\mu \phi} \partial_0 \phi - \mathcal{L} \delta^\mu_0.$$

In particular, we have

$$j^0 = \frac{\delta \mathcal{L}}{\delta \partial_t \phi} \partial_t \phi - \mathcal{L},$$

and one can obtain

$$S_i = j^i = \frac{\delta \mathcal{L}}{\delta \partial_i \phi} \partial_t \phi$$

for i being one of the spatial coordinates.

Example: Going back to the rotation, consider $\phi = \rho - \bar{\rho}$ to be a scalar field. Let $r^\mu = (t, r^i)$ denote the general coordinate, then here we have

$$t \rightarrow t' = t + 0 \cdot \delta \theta^j, \quad r^i \rightarrow r'^i = r^i + \epsilon_{ijk} r^k \delta \theta^j.$$

That is, we have $\vec{r}' \rightarrow \vec{r}' = \vec{r} + \vec{n} \times \vec{r} \delta \theta$. Direct computation we see here

$$\frac{\delta \epsilon^\mu}{\delta \theta^j} = \begin{cases} 0 & \mu = 0 \\ \epsilon_{ijk} r^k & \mu = x, y, z \end{cases}.$$

Furthermore, we have $\phi'(r') = \phi(r) + 0 \cdot \delta \theta^j$ as ϕ is a scalar field, unaffected by rotation. Then the conserved quantity here is

$$j^\mu = \left(\frac{\delta \mathcal{L}}{\delta \partial_\mu \phi} \partial_\nu \phi - \delta^\mu_\nu \mathcal{L} \right) \frac{\delta \epsilon^\nu}{\delta \theta^j} = \left(\frac{\delta \mathcal{L}}{\delta \partial_\mu \phi} \partial_i \phi - \delta^\mu_i \mathcal{L} \right) \epsilon_{ijk} r^k,$$

where i is one of the spatial coordinates. In particular, we have

$$j^0 = \left(\frac{\delta \mathcal{L}}{\delta \partial_t \phi} \partial_i \phi \right) \epsilon_{ijk} r^k.$$

We showed previously that we have

$$j^0 = -g^i \epsilon_{ijk} r^k, \quad \text{where } g^i = \frac{\delta \mathcal{L}}{\delta \partial_t \phi} \partial_i \phi.$$

Here the rotation is along one of the spatial axes, denoted as j , thus j^0 is just the j -component of a more general conserved quantity $\vec{j}^0 := \vec{g} \times \vec{r}$. The orbital angular momentum of the system thus can be characterized by a conserved quantity

$$\vec{L} := \iiint dx dy dz \vec{g} \times \vec{r}.$$

Example: Now consider instead a vector field \vec{u} under a rotation transformation, the term $\delta \vec{u} / \delta w^b$ no longer vanishes. Note that all vectors rotate the same way, as vector \vec{r} rotates as

$$\vec{r} \rightarrow \vec{r}' = \vec{r} + \vec{n} \times \vec{r} \delta \theta,$$

then the field rotates in the same way

$$\vec{u}(\vec{r}) \rightarrow \vec{u}'(\vec{r}') = \vec{u}(\vec{r}) + \vec{n} \times \vec{u}(\vec{r}) \delta \theta.$$

Or in components, we can write

$$u^i(r) \rightarrow u'^i(r') = u^i(r) + \epsilon_{ijk} u^k \theta^j.$$

Thus we obtain

$$j_j^0 = -g^i \epsilon_{ijk} r^k - \frac{\delta \mathcal{L}}{\delta \partial_t u^l} \frac{\delta u^l}{\delta \theta^j} = \underbrace{-g^i \epsilon_{ijk} r^k}_{\text{orbital angular momentum}} - \underbrace{\frac{\delta \mathcal{L}}{\delta \partial_t u^l} \epsilon_{ijk} u^k}_{\text{spin angular momentum}}.$$

Example: For a sound wave, we have found that the Lagrangian density of the sound wave in solid is

$$\mathcal{L} = \frac{\rho}{2} \partial_t \vec{u} \partial_t \vec{u} - \frac{\lambda}{2} (\text{tr}(e))^2 - G \text{tr}(ee).$$

Here we denote \vec{S} as the spin angular momentum, then we can write

$$S_i = \iiint dx dy dz \epsilon_{ikl} u^k \frac{\delta \mathcal{L}}{\delta \partial_t u^l},$$

where we have

$$\frac{\delta \mathcal{L}}{\delta \partial_t u^l} = \frac{\delta}{\delta \partial_t u^l} \left(\frac{\rho}{2} (\partial_t u^l)^2 \right) = \rho \partial_t u^l,$$

thus we obtain

$$S_i = \iiint dx dy dz \rho \epsilon_{ikl} u^k \partial_t u^l.$$

In other words,

$$\vec{S} = \iiint dx dy dz \rho \vec{u} \times (\partial_t \vec{u}). \quad (2.30)$$

Note here (2.30) is generic, independent on the form of the strain tensor e . Furthermore, linearly polarized sound wave typically has the form given by

$$\vec{u}(\vec{r}, t) = \vec{A} \cos(\vec{k} \cdot \vec{r} - \omega t + \theta).$$

We observe here

$$\partial_t \vec{u} = \vec{A} \omega \sin(\vec{k} \cdot \vec{r} - \omega t + \theta),$$

thus we have $\vec{u} \times (\partial_t \vec{u}) = 0$, that is $\vec{S} = 0$. We conclude here linear polarized wave have spin zero. As sound wave can be decomposed into longitudinal and transverse modes, we see here immediately that the longitudinal mode has spin zero. For transverse sound wave, which typically takes the form

$$\vec{u}(\vec{r}, t) = \vec{A}_1 \cos(\vec{k} \cdot \vec{r} - \omega t + \theta_1) + \vec{A}_2 \cos(\vec{k} \cdot \vec{r} - \omega t + \theta_2).$$

If one has $\theta_1 = \theta_2$, we obtain

$$\vec{u} = (\vec{A}_1 + \vec{A}_2) \cos(\vec{k} \cdot \vec{r} - \omega t + \theta),$$

or if one has $\theta_1 = \theta_2 + n\pi$, then we can also write

$$\vec{u} = (\vec{A}_1 \pm \vec{A}_2) \cos(\vec{k} \cdot \vec{r} - \omega t + \theta).$$

In both cases, $\theta_1 = \theta_2 + n\pi$ or $\theta_1 = \theta_2$, we have linearly polarized wave, and thus $\vec{S} = 0$. While in general cases, $\theta_1 \neq \theta_2 + n\pi$, which gives elliptical polarized wave, with special case $\theta_1 - \theta_2 = \pi/2 + n\pi$ being circular polarized wave. We will see in the following that elliptical polarized waves have nonzero spin. Here we can write

$$\partial_t \vec{u} = \vec{A}_1 \omega \sin(\vec{k} \cdot \vec{r} - \omega t + \theta_1) + \vec{A}_2 \omega \sin(\vec{k} \cdot \vec{r} - \omega t + \theta_2)$$

Then we can see here

$$\begin{aligned} \vec{u} \times (\partial_t \vec{u}) &= \vec{A}_1 \times \vec{A}_2 \omega \cos(\mathcal{M}_1) \sin(\mathcal{M}_2) + \vec{A}_2 \times \vec{A}_1 \omega \cos(\mathcal{M}_2) \sin(\mathcal{M}_1) \\ &= \vec{A}_1 \times \vec{A}_2 \omega \cos(\mathcal{M}_1) \sin(\mathcal{M}_2) - \vec{A}_1 \times \vec{A}_2 \omega \cos(\mathcal{M}_2) \sin(\mathcal{M}_1) \\ &= (\vec{A}_1 \times \vec{A}_2) \omega \sin(\theta_2 - \theta_1), \end{aligned}$$

where we have $\mathcal{M}_1 = \vec{k} \cdot \vec{r} - \omega t + \theta_1$ and $\mathcal{M}_2 = \vec{k} \cdot \vec{r} - \omega t + \theta_2$. Thus we see here

$$\begin{aligned} \vec{S} &= \iiint dx dy dz \rho \vec{u} \times (\partial_t \vec{u}) \\ &= \iiint dx dy dz \rho (\vec{A}_1 \times \vec{A}_2) \omega \sin(\theta_2 - \theta_1) \\ &= \rho \omega (\vec{A}_1 \times \vec{A}_2) \sin(\theta_2 - \theta_1) V, \end{aligned}$$

where V is the volume of the solid. We observe that we have $\vec{S} \propto V$, $\vec{S} \propto (A_1 A_2) \sim A^2$. Furthermore, if $\theta_1 - \theta_2 = n\pi$, then $\vec{S} = 0$, and if $\theta_1 - \theta_2 \neq n\pi$, we must have $\vec{S} \neq 0$. Note also that $\vec{A}_1 \times \vec{A}_2$ is parallel to \vec{k} by definition of transverse wave, and thus $\vec{S} \parallel \vec{k}$, so here we can write

$$\vec{S} = \rho \omega A_1 A_2 \sin(\theta_2 - \theta_1) V \hat{k}.$$

To get an intuition of what \vec{S} represents, we compute the quantity

$$\frac{\vec{S} \cdot \hat{k}}{E} = \frac{\rho\omega A_1 A_2 \sin(\theta_2 - \theta_1)V}{\omega^2 \rho (A_1^2 + A_2^2)V/2} = \frac{2A_1 A_2}{\omega(A_1^2 + A_2^2)} \sin(\theta_2 - \theta_1). \quad (2.31)$$

where we have applied (2.27) for energy of a sound wave. To maximize (2.31), we can separately deal with the sin term and the amplitude term. One can see that we have

$$-\frac{1}{\omega} \leq \frac{\vec{S} \cdot \hat{k}}{E} \leq \frac{1}{\omega}.$$

The amplitude term in (2.31) can be maximized by choosing $A_1 = A_2$, that is, the ratio (2.31) is maximized when we have right-circular polarized mode

$$A_1 = A_2, \quad \text{with } \theta_2 - \theta_1 = \frac{\pi}{2} + 2n\pi;$$

and the ratio (2.31) is minimized when we have left-circular polarized mode

$$A_1 = A_2, \quad \text{with } \theta_2 - \theta_1 = -\frac{\pi}{2} + 2n\pi.$$

One can in general write a general wave with the basis of left- and right-circular polarized wave,

$$|T\rangle = \alpha_1 |R\rangle + \alpha_2 |L\rangle,$$

where $|R\rangle$ and $|L\rangle$ represent the right- and left-circular polarized wave, respectively, and here we have

$$\langle \vec{S} \cdot \vec{k} \rangle = (|\alpha_1|^2 - |\alpha_2|^2) \frac{E}{\omega}.$$

That is, one thing to clarify is that linearly polarized wave, with $\alpha_1 = \alpha_2$, does have left- and right-handed spin, but the sum of the two is zero, resulting a zero expectation value of $\vec{S} \cdot \hat{k}$. Here the quantity $\vec{S} \cdot \hat{k}$ is called the Helicity of the wave.

In conclusion here, for a sound wave in solid, we have

$$\frac{\vec{S} \cdot \hat{k}}{E} = \begin{cases} 0 & \text{longitudinal mode} \\ 1/\omega & \text{right-circular polarized mode} \\ -1/\omega & \text{left-circular polarized mode} \end{cases}.$$

Using particle-wave duality, we can interpret this result as

$$\frac{\vec{S} \cdot \hat{k}}{E} = \frac{(\vec{s} \cdot \hat{k}) \times \text{number of particles}}{(\hbar\omega) \times \text{number of particles}},$$

what we have found for the wave thus requires

$$\vec{s} \cdot \hat{k} = \begin{cases} 0 & \text{longitudinal mode} \\ +\hbar & \text{right-circular polarized mode} \\ -\hbar & \text{left-circular polarized mode} \end{cases},$$

if we set $\hat{k} = \hat{z}$, we recognize that vector fields correspond to spin-1 particles, with spin $\{-\hbar, 0, \hbar\}$. That is, for vector field, the spin is always 1, but its projection (the modes of the wave) on the z -axis takes integer values between -1 and 1 . It follows from our analysis that spin-0 particle corresponds to the scalar field, which always have spin zero. One can further show that rank-2 tensor fields correspond to spin-2 particles, rank-3 tensor fields correspond to spin-3 particles. However, in classical field theory, there is no field (scalar, vector, or tensor field) corresponds to spin-half particles. To obtain fields that corresponds to spin-half particles, one needs to introduce the notion of spinors.

3 | Special Relativity

Lorentz Transformation

By far, the speed of the waves, such as sound waves, are defined relative to the media that it is propagating in. However, electromagnetic waves (EM waves) can propagate in vacuum, which is not a *media*, thus we need to find a reference frame to describe the electromagnetic wave. Here there are three candidates to construct the reference frame for EM waves, (1) the source, (2) the medium, and (3) the observer. While in theory, by the principle of locality, the source cannot be used to define the reference frame as the motion of the source does not effect the wavefront when they are far away from each other. The medium cannot be used to define the reference frame either, as the EM wave can travel in vacuum, one would need to define a medium *ether* for which EM waves in general can travel through, and turns out to cause other problems in the theory and disagree with experimental observations, that is the Michelson-Morley experiment.

Therefore, we now can only use the observer itself as the reference frame. We need to ensure that the theory itself is self-consistent, and one can make experimental predictions from the theory. As long as some fundamental assumptions are adopted, the transformation between different reference frames $K = (t, x, y, z) \rightarrow K' = (t', x', y', z')$ will have the form

$$t' = \gamma \left(t - \frac{\gamma^2 - 1}{v\gamma^2} x \right), \quad x' = \gamma(x - vt), \quad y' = y, \quad z' = z. \quad (3.1)$$

To simplify our calculation, we choose the origin of the two reference frames to be the same point at $t = x = y = z = t' = x' = y' = z' = 0$, and conventionally we set up axes such that $x \parallel x'$, $y \parallel y'$, and $z \parallel z'$, with the unprimed frame moving relative to the primed frame in the xx' -direction.

Now we impose Assumption (1), that the two frames K and K' are related by a linear transformation,

$$\begin{bmatrix} t' \\ x' \\ y' \\ z' \end{bmatrix} = A \begin{bmatrix} t \\ x \\ y \\ z \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix},$$

with some matrix A . Note here the (b_1, b_2, b_3, b_4) vector vanishes as one of our conventions is that the two frames have the same origin. Here we denote the components of A as a_{ij} .

Furthermore, we impose Assumption (2), that the space is isotropic, that is, we can rotate our coordinates around the xx' -axis and everything shall remain unchanged. Note here if we rotate the coordinates in the K frame by an angle φ about the x -axis, by one of the conventions that the axes in the two frames are parallel pairwise, we shall also rotate the coordinates in the K' frame by the same angle φ .

Lastly, we impose Assumption (3) that we have mirror symmetry for the spacetime.

With the three Assumptions and the conventions imposed, the transformation law (3.1) and the elements of matrix A can be derived. The proof is left for the reader.

Now suppose in the K frame, a light signal is emitted at time $t = 0$, then the wavefront of the light has position $\sqrt{x^2 + y^2 + z^2}/c$ at time t , where c denotes the speed of the wavefront. Thus here we can write

$$ct = \sqrt{x^2 + y^2 + z^2}. \quad (3.2)$$

As K' and K have the same origin, then we can also write

$$ct' = \sqrt{x'^2 + y'^2 + z'^2}, \quad (3.3)$$

relating (3.2) and (3.3) via the transformation law (3.1), we can write

$$0 = c^2 \gamma^2 \left(t - \frac{\gamma^2 - 1}{v \gamma^2} x \right)^2 - \gamma^2 (x - vt)^2 + x^2 - c^2 t^2,$$

from which one obtain

$$0 = (c^2 \gamma^2 - v^2 \gamma^2 - c^2) t^2 + \left(c^2 \left(\frac{\gamma^2 - 1}{v \gamma} \right)^2 - \gamma^2 + 1 \right) x^2 - \left(c^2 \frac{\gamma^2 - 1}{v} - \gamma^2 v \right) tx,$$

which has a general solution given by

$$\gamma = \frac{c}{\sqrt{c^2 - v^2}} = \frac{1}{\sqrt{1 - v^2/c^2}}. \quad (3.4)$$

With γ given by (3.4), the transformation law (3.1) has the form

$$t' = \gamma(t - vx/c^2), \quad x' = \gamma(x - vt), \quad y' = y, \quad z' = z. \quad (3.5)$$

In Newton's theory, one can check that $\gamma = 1$ recovers the Galilean transformation. In Einstein's special theory of Relativity, $\gamma \geq 1$. We observe here, when $v \ll c$, we indeed obtain $\gamma \approx 1$, which ensures the consistency of Einstein's theory with the well-conducted experiments that agree with Newton's theory.

Causality

3.2.1 Relativity of simultaneity

Now with the new theory, not all intuition that we had in classical kinematic theories shall be kept. While, all those intuition can be kept as long as we stay in the same

reference frame. Otherwise *weird things* will happen. Here we will utilize the notion of K and K' frames defined in Section - Lorentz Transformation. In K frame, suppose we have event A and event B happening at the same time t . Classically, $t' = t$ in K' frame and the two events are simultaneous in both K and K' frames. However, this is not the case in the special theory of relativity. Denote that A has coordinate (t_A, x_A, y_A, z_A) in the K frame and with components primed in the K' frame. With similar definition for coordinate of B , one can derive

$$t'_A = \gamma \left(t_A - \frac{v}{c^2} x_A \right), \quad t'_B = \gamma \left(t_B - \frac{v}{c^2} x_B \right).$$

Thus we see here

$$t'_A - t'_B = \gamma(t_A - t_B) - \gamma \frac{v}{c^2} (x_A - x_B), \quad (3.6)$$

which does not vanish whenever $x_A \neq x_B$, thus even though the two events might happen at the same time $t_A = t_B$ in frame K , such notion of simultaneity does not hold in the frame K' . While if the two events happen at the same point $x_A = x_B$, the simultaneity in K frame will also hold in the K' frame. Further, denoting $T = t_A - t_B$, if we have $x_A = x_B$, then we obtain

$$T' = \gamma T \geq T$$

as $\gamma \geq 1$. This result can be intuitively understood via the example of a person A holding a flashlight in a train moving with speed v . Say the train has length L , and a person B is standing still outside of the train. Then the person B will observe that the light from the flashlight held by A reach the two ends of the train at different time as the train is moving, while the person A will observe that the light from his flashlight reach the two ends of the train simultaneously. Note here the assumption of light travels at the same speed in all inertial frames play an essential role.

The proper time is defined for an object. Suppose we have a moving object, and a clock is attached to the object, the proper time of the object is the reading of the clock. On the other hand, the coordinate time is the laboratory time defined for observer. Observers in different places in the laboratory might possess different coordinate time, but all observers will agree with the same proper time defined for an object.

Example: Suppose one has some object which has a proper decay time τ . The number of particles in the object is characterized by

$$N(t) = N_0 e^{-t/\tau}. \quad (3.7)$$

If the particle is stationary, a person in the laboratory will measure the number of particles in the object at time t given precisely by (3.7). However, if the particle is moving, we have instead

$$N(t) = N_0 e^{-t/(\gamma\tau)},$$

thus the faster the particle moves, the slower its decay measured in the laboratory. This is in fact one of the experimental verifications of special theory of relativity. One can measure the decay time τ^* of the particle,

$$N(t) = N_0 e^{-t/\tau^*},$$

as a function of the speed v of the particle in the laboratory. One in fact finds that

$$\tau^*(v) = \frac{\tau}{\sqrt{1 - v^2/c^2}}.$$

Example: Now suppose a clock C in frame S moves from point A to point B at constant velocity v in 15 seconds. Suppose further that $v = \sqrt{3}c/2$ such that $\gamma = 2$. In C 's own stationary frame, the clock itself ticks 30 seconds in its proper time. That is, one can picture the measurement in the frame S as having two timing clocks synchronized at point A and B , when the clock C moves from A to B , both timing clocks tick 15 seconds. While in the C 's own stationary frame, the simultaneity of the timing clock breaks, that is the two timing clocks are not at the synchronized. One can find that, in C 's frame, the timing clock at point A has time coordinate 0, while the one at point B has time coordinate 22.5 seconds, such that when C moves to B , the timing clock at B has time coordinate 30 seconds.

We note that, the proper time can be made much smaller than the coordinate time t , as $1 \geq \gamma > 0$ can be made arbitrarily small, thus no matter how far between two points, a person can travel from one to the other in an arbitrarily small amount of his proper time, as the person has its own coordinate time as proper time, provided that the person travel at a speed closed enough to the speed of light.

In fact, the proper time formula (3.6) reflects the conservation of spacetime. Note that in the object's own frame, the object does not move, thus we can write

$$c\tau = \sqrt{c^2(t_A - t_B)^2 - (x_A - x_B)^2 - (y_A - y_B)^2 - (z_A - z_B)^2},$$

where in the lab frame, we have

$$v^2(t_A - t_B)^2 = (x_A - x_B)^2 - (y_A - y_B)^2 - (z_A - z_B)^2$$

as the object is traveling at a velocity v , thus combining we have

$$c\tau = \sqrt{c^2(t_B - t_A)^2 - v^2(t_A - t_B)^2} = (t_A - t_B)\sqrt{c^2 - v^2},$$

denoting coordinate time $T = t_B - t_A$, we recover

$$\tau = T \frac{\sqrt{c^2 - v^2}}{c} = \frac{T}{\gamma}$$

3.2.2 Length contraction

Consider in the unprimed frame, we have two positions with coordinates

$$p = (t_B, x_B, 0, 0), \quad q = (t_A, x_A, 0, 0),$$

where x_A and x_B are constants independent on time. In the primed frame, which is moving at a speed v relative to the unprimed frame, we have the two positions with primed coordinates,

$$\begin{aligned} p' &= (\gamma(t_B - vx_B/c^2), \gamma(x_B - vt_B), 0, 0), \\ q' &= (\gamma(t_A - vx_A/c^2), \gamma(x_A - vt_A), 0, 0). \end{aligned}$$

The proper length between the two positions in the unprimed frame is thus given by $L = x_A - x_B$ measured at the same time in the unprimed frame. The length between the two positions in the primed frame is, on the other hand, $L' = x'_A - x'_B$ while measured at the same time in the primed frame. Thus we compute

$$\gamma(t_A - vx_A/c^2) = \gamma(t_B - vx_B/c^2), \quad \text{which gives } t_A - t_B = \frac{v}{c^2}(x_A - x_B).$$

Hence we obtain

$$L' = \gamma((x_A - x_B) - v(t_A - t_B)) = \gamma((x_A - x_B) - v^2(x_A - x_B)/c^2) = \frac{L}{\gamma}, \quad (3.8)$$

as we know that $\gamma \geq 1$, we see here $L' \leq L$, that is length contraction. This phenomenon, seemingly contradictive, can be in fact explained by the relativity of simultaneity. As one in the unprimed frame trying to measure the length of the object assuming that p and q have the same time coordinate in the unprimed frame), while the corresponding p' and q' coordinates do not have the same time coordinates in the unprimed frame.

Now suppose we have event A and event B , with coordinates (t_A, x_A) and (t_B, x_B) , respectively, such that $x_B > x_A$ and $t_B > t_A$. One can find a frame with relative moving speed $v > 0$, such that

$$t'_A = \gamma\left(t_A - \frac{v}{c^2}x_A\right), \quad t'_B = \gamma\left(t_B - \frac{v}{c^2}x_B\right).$$

One sees here

$$t'_A - t'_B = \gamma\left((t_A - t_B) - \frac{v}{c^2}(x_A - x_B)\right), \quad (3.9)$$

which is not guaranteed to be non-positive whenever $t_A - t_B < 0$. That is, seemingly, event B happens after event A has happened in one frame does not imply event B happens after event A has happened in all frames. However, if one also consider that, information propagates at some speed $u > 0$, one can ensure the causality, say $A \rightarrow B$, that is B happens after A has happened, if we have

$$t_B - t_A \geq \frac{x_B - x_A}{u}. \quad (3.10)$$

Via (3.9), one sees that, by imposing u as the speed of propagation of information, and if (3.10) holds, we then obtain

$$t'_B - t'_A \geq \gamma\left(\frac{x_B - x_A}{u} - \frac{v}{c^2}(x_B - x_A)\right) = \gamma\left(1 - \frac{uv}{c^2}\right)\left(\frac{x_B - x_A}{u}\right) > 0,$$

where we have imposed an extra requirement that $u \leq c$.

3.2.3 Light-cone

For two events A and B , with coordinates

$$A = (t_A, x_A, y_A, z_A), \quad B = (t_B, x_B, y_B, z_B);$$

The two events are said to be time-like separated provided that we have

$$c^2(t_B - t_A)^2 > (x_B - x_A)^2 + (y_B - y_A)^2 + (z_B - z_A)^2;$$

The two events are said to be space-like separated provided that we have

$$c^2(t_B - t_A)^2 < (x_B - x_A)^2 + (y_B - y_A)^2 + (z_B - z_A)^2;$$

The two events are said to be null separated provided that we have

$$c^2(t_B - t_A)^2 = (x_B - x_A)^2 + (y_B - y_A)^2 + (z_B - z_A)^2.$$

Note here, if one redefine A to be the origin of the chart, then the null separated points consist of the light-cone about A . Inside the light-cone, points are time-like related to A . Outside the light-cone, points are space-like related to B . Furthermore, the light-cone consists of two cones, the future cone and the past cone. The future cone lies completely in the positive t -hypersurface, and the past cone lies completely in the negative t -hypersurface. Points inside the light-cone stay inside the light-cone, while points outside the light-cone stay outside the light-cone, under all Lorentz transformation.

Types of Transformations

In 3-dimensional space \mathbb{R}^3 , we are interested in isometric transformation, that is the transformation that preserve the length

$$l^2 := \Delta x^2 + \Delta y^2 + \Delta z^2$$

between the two points A and B . Here $\Delta x := x_B - x_A$, and similarly for Δy and Δz . We denote the coordinate of A and B to be (x'_A, y'_A, z'_A) and (x'_B, y'_B, z'_B) , respectively, and the length between A and B to be l' , after the transformation. That is, for isometric transformation, we require

$$l^2 = l'^2 = \Delta x'^2 + \Delta y'^2 + \Delta z'^2.$$

There are four types of isometric transformation. The first type is the translation

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

The second type is the rotation

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = R \begin{bmatrix} x \\ y \\ z \end{bmatrix},$$

for some matrix R that satisfies $R^T = R^{-1}$ and $\det(R) = 1$. The third type is the space inversion transformation characterized by

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = - \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

The fourth type is any combination of the first three types of transformation. To show that these four types of transformation are the only isometric transformations, we write here

$$l^2 = \begin{bmatrix} \Delta x & \Delta y & \Delta z \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = \begin{bmatrix} \Delta x' & \Delta y' & \Delta z' \end{bmatrix} \begin{bmatrix} \Delta x' \\ \Delta y' \\ \Delta z' \end{bmatrix}.$$

Here we define

$$\begin{bmatrix} \Delta x' \\ \Delta y' \\ \Delta z' \end{bmatrix} = A \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix},$$

for some 3×3 matrix A , and elementary linear algebra result gives

$$\begin{bmatrix} \Delta x' & \Delta y' & \Delta z' \end{bmatrix} = \begin{bmatrix} \Delta x & \Delta y & \Delta z \end{bmatrix} A^T.$$

Thus combining we require

$$l'^2 = \begin{bmatrix} \Delta x & \Delta y & \Delta z \end{bmatrix} A^T A \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = l^2,$$

that is require here $A^T = A^{-1}$, the transformation A needs to be orthogonal. In particular, the rotation transformations constitute the special orthogonal group.

Lorentz boost and rotation transformation

Next we will show that Lorentz boost in 4-dimensional space are rotational isometric transformation. Here we first define the distance that we are interested in,

$$l^2 = c^2(t_B - t_A)^2 - (x_B - x_A)^2 - (y_B - y_A)^2 - (z_B - z_A)^2.$$

As Lorentz boosts preserve l^2 , we see immediately that Lorentz boosts are isometric transformation. Furthermore, note that in flat space, translation transformation has no fixed point. That is, all points in space are translated under a translation transformation. While in contrast, rotation transformation always have a fixed point, or a fixed axis. From the definition of Lorentz boost,

$$t' = \gamma(t - vx/c^2), \quad x' = \gamma(x - vt), \quad y' = y, \quad z' = z,$$

we see that the origin $(0,0,0,0)$ is a fixed point. Thus a Lorentz boost cannot be a translation transformation. Furthermore, it is obvious that the Lorentz boosts are not inversion either, as it is a continuous transformation. In conclude, we see that Lorentz boost is a rotation.

Note further that the l^2 being invariant under Lorentz boost implies a symmetry in space, which gives rise to the physical laws being invariant in 4-dimensional space. The *invariance* is in fact characterized by the defining all tensors *rotating in the same way*, as we shall discuss in the next section.

Here we introduce a chart $x^\mu = (x^0, x^1, x^2, x^3) = (ct, x, y, z)$, and another chart x'^μ . Here if x'^μ is related to x^μ via a Lorentz transformation, then we can write

$$\begin{bmatrix} x'^0 \\ x'^1 \\ x'^2 \\ x'^3 \end{bmatrix} = \begin{bmatrix} \gamma & -v\gamma/c & 0 & 0 \\ -v\gamma/2 & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{bmatrix} = \begin{bmatrix} \cosh(\xi) & \sinh(\xi) & 0 & 0 \\ \sinh(\xi) & \cosh(\xi) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{bmatrix},$$

and one can mathematically show that the two matrices here, being equivalent, are rotational transformation in 4-dimensional space. With the notion of rotation in 4-dimensional space, we are now able to define 4-dimensional vectors.

All types of isometric transformation in 4-dimensional space are the translation, spatial rotation, Lorentz boost, inversion of time, inversion of space, and any mix of these. In particular, translation can be compactly characterized by

$$x^\mu \rightarrow x'^\mu = x^\mu + b^\mu.$$

For the rotation transformation, which include both the spatial rotation and Lorentz boost, we can write

$$\Delta x'^\mu = R \Delta x^\mu,$$

with some matrix R . Let $g_{\mu\nu}$ denote the flat metric tensor, with matrix form

$$g = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix},$$

we thus can write

$$l^2 = g_{\mu\nu}(\Delta x^\mu)(\Delta x^\nu) = g_{\mu\nu}(\Delta x'^\mu)(\Delta x'^\nu) = l'^2.$$

Combining with rotation, and obtain

$$l^2 = \begin{bmatrix} \Delta t & \Delta x & \Delta y & \Delta z \end{bmatrix} g \begin{bmatrix} \Delta t \\ \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = \begin{bmatrix} \Delta t & \Delta x & \Delta y & \Delta z \end{bmatrix} R^T g R \begin{bmatrix} \Delta t \\ \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = l'^2.$$

That is, we require here the rotation matrix to satisfy

$$g = R^T g R.$$

One can also show that the determinant of the rotational matrix is $\det(R) = \pm 1$, as we can write

$$\det(R^T g R) = \det(R^T) \det(g) \det(R) = \det(g),$$

thus we obtain

$$\det(R) \cdot \det(R^T) = \det(R) \cdot \det(R) = 1,$$

which implies we have $\det(R) = \pm 1$. As the spacial inversion transformation,

$$I_s = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

has determinant -1 and is not considered to be a rotation, thus rotation transformation R should satisfy $\det(R) = 1$. Lastly, the total inversion transformation

$$I = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

has determinant 1 but is neither a rotation, thus rotation transformation is required to have $R_{00} > 0$. In conclude, a matrix transformation R is called a rotation in 4-dimensional spacetime provided that we have $R^T g R = g$, $\det(R) = 1$, and $R_{00} > 0$.

Vectors and Tensors in Spacetime

Now with rotation in spacetime, including both spatial rotations and Lorentz boost, well-defined, we can define the notion of vectors and tensors in spacetime. Due to rotational symmetry, we expect fundamental laws of physics to remain invariant no matter how we rotate the axes. Thus, these physics laws must take the form of

$$\text{Scalar}_1 = \text{Scalar}_2, \quad \text{Vector}_1 = \text{Vector}_2, \quad (\text{Rank-}n) \text{ Tensor}_1 = (\text{Rank-}n) \text{ Tensor}_2,$$

and we need to require all vectors (or tensors) rotate the same way.

The **contravariant vectors** are vectors that rotate the same way as coordinates. That is, the coordinates of spacetime are themselves contravariant vectors, and they rotate in the way characterized by

$$\begin{bmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{bmatrix} \rightarrow \begin{bmatrix} x'^0 \\ x'^1 \\ x'^2 \\ x'^3 \end{bmatrix} = R \begin{bmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{bmatrix},$$

where R is a rotation in spacetime, which can be either a boost,

$$R = \begin{bmatrix} \gamma & -v\gamma/c & 0 & 0 \\ -v\gamma/c & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

or a spatial rotation

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & R_{xx} & R_{xy} & R_{xz} \\ 0 & R_{yx} & R_{yy} & R_{yz} \\ 0 & R_{zx} & R_{zy} & R_{zz} \end{bmatrix}.$$

Contravariant vectors are denoted using superscript, such as x^μ for spacetime coordinates.

On the other hand, we will show that the derivatives of spacetime, ∂_μ , rotate in a way different from contravariant vectors. Here we note that, based on our convention $x_0 = ct$, we have

$$\partial_0 = \frac{\partial}{\partial x^0} = \frac{\partial t}{\partial x^0} \frac{\partial}{\partial t} = \frac{\partial(x^0/c)}{\partial x^0} \frac{\partial}{\partial t} = \frac{1}{c} \partial_t.$$

For boost transformation of coordinates, we have that

$$x^0 = \gamma x'^0 + \frac{v}{c} \gamma x'^1, \quad x^1 = \frac{v}{c} \gamma x'^0 + \gamma x'^1, \quad x^2 = x'^2, \quad x^3 = x'^3.$$

Thus chain rule gives

$$\partial'_0 = \frac{\partial}{\partial x'^0} = \frac{\partial x^0}{\partial x'^0} \frac{\partial}{\partial x^0} + \frac{\partial x^1}{\partial x'^0} \frac{\partial}{\partial x^1} + \frac{\partial x^2}{\partial x'^0} \frac{\partial}{\partial x^2} + \frac{\partial x^3}{\partial x'^0} \frac{\partial}{\partial x^3} = \gamma \partial_0 + \frac{v}{c} \gamma \partial_1.$$

Similarly for ∂'_1, ∂'_2 , and ∂'_3 , we obtain

$$\begin{bmatrix} \partial'_0 \\ \partial'_1 \\ \partial'_2 \\ \partial'_3 \end{bmatrix} = \begin{bmatrix} \gamma & v\gamma/c & 0 & 0 \\ v\gamma/c & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \partial_0 \\ \partial_1 \\ \partial_2 \\ \partial_3 \end{bmatrix}.$$

One can show similarly for spatial rotation to obtain

$$\begin{bmatrix} \partial'_0 \\ \partial'_1 \\ \partial'_2 \\ \partial'_3 \end{bmatrix} = (R^{-1})^T \begin{bmatrix} \partial_0 \\ \partial_1 \\ \partial_2 \\ \partial_3 \end{bmatrix}.$$

Here the derivative vector ∂_μ is called a covariant vector, and thus the **covariant vectors** are the objects that follow the inverse transpose transformation of coordinates. Note here we have used subscripts to denote covariant vectors.

It is not hard to observe that we have

$$\begin{bmatrix} \partial'_0 \\ -\partial'_1 \\ -\partial'_2 \\ -\partial'_3 \end{bmatrix} = \begin{bmatrix} \gamma & -v\gamma/c & 0 & 0 \\ -v\gamma/c & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \partial_0 \\ -\partial_1 \\ -\partial_2 \\ -\partial_3 \end{bmatrix},$$

thus the vector $(\partial_0, -\partial_1, -\partial_2, -\partial_3)$ is a contravariant vector.

Example: If we have a plane wave $\psi = Ae^{i\vec{k}\cdot\vec{r}-i\omega t}$, then combined with the contravariant derivative operator, $ic\partial^\mu\psi$ gives a contravariant vector of frequency and wave-vector $k^\mu = (\omega, ck_1, ck_2, ck_3)$, and the covariant derivative operator can be used to obtain a covariant vector $k_\mu = (\omega, -ck_1, -ck_2, -ck_3)$. From particle-wave duality, we know that $\hbar\omega$ is energy and $\hbar k$ is momentum, and thus the contravariant vector $(\omega, ck_1, ck_2, ck_3)$ tells us that $(\hbar\omega, \hbar ck_1, \hbar ck_2, \hbar ck_3)$ should be a contravariant vector, and this is in fact,

the energy-momentum 4-vector (E, cP_1, cP_2, cP_3) .

Remark: In 3-dimensional spatial space, that is \mathbb{R}^3 equipped with Euclidean metric, the metric tensor g is the identity matrix. In which case, if R is a rotation, then we require $R^T g R = g$, which implies $R^T R = I$, and thus $(R^{-1})^T = R$. That is, in this case, we do not need to distinguish contravariant vectors and covariant vectors, because they follow the same transformation under rotation R .

Now for the inner product between contravariant vector X^μ and Y^μ , denote by \vec{x} and \vec{y} in their usual R^n -vector form, after a rotation transformation, we can write

$$X'^\mu Y'_\mu = ((R^{-1})^T \vec{y})^T (R \vec{x}) = \vec{y}^T R^{-1} R \vec{x} = \vec{y}^T \cdot \vec{x} = X^\mu Y_\mu.$$

That is the inner product $X^\mu Y_\mu$ is a scalar invariant under rotation.

Let g denote the metric of spacetime. In special relativity, g takes the form

$$g = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

Next we would like to show that, given X^μ is a contravariant vector, denoted as \vec{x} in its \mathbb{R}^n -vector form, $g\vec{x}$ defines a covariant vector. Note that under rotation R , we have $\vec{x} \rightarrow R\vec{x}$, and thus $g\vec{x} \rightarrow gR(g^{-1}g)\vec{x}$. While rotation R satisfies $R^T g R = g$, thus $gR = (R^T)^{-1}g$, so we can write $gRg^{-1} = (R^T)^{-1}$, which implies $g\vec{x} \rightarrow (R^T)^{-1}g\vec{x}$, showing that $g\vec{x}$ is a covariant vector. Then in tensor notation, we thus write

$$X_\mu = g^{\mu\nu} X^\nu.$$

Similarly, one can show that $g^{-1}\vec{x}$ is a covariant vector if \vec{x} represents a contravariant vector, and thus

$$X^\mu = g^{\mu\nu} X_\nu,$$

where we have denoted that $g^{\mu\nu}$ is the inverse of $g_{\mu\nu}$. In general, one utilize the notation

$$g^{\mu\nu} g_{\nu\lambda} = \delta^\mu_\lambda = g^\mu{}_\lambda, \quad g_{\mu\lambda} g^{\lambda\nu} = \delta_\mu{}^\nu = g_\mu{}^\nu.$$

Now that inner product of X^μ and Y^μ can then be written in the way

$$X^\mu g_{\mu\nu} Y^\nu = X^\mu Y_\mu = X_\mu Y^\mu.$$

Example: Here we consider $x^\mu = (x^0, x^1, x^2, x^3)$ and $\partial_\mu = (\partial_0, \partial_1, \partial_2, \partial_3)$, we thus have $\partial_\mu x^\mu = 4$, and that

$$\partial_\mu \partial^\mu = \partial_0^2 - \partial_1^2 - \partial_2^2 - \partial_3^2 = \partial_0^2 - \nabla^2 := \square.$$

Example: The product between the covariant vector $(\omega, -ck_1, -ck_2, -ck_3)$ and the contravariant vector $(\omega, ck_1, ck_2, ck_3)$ is a Lorentz invariant scalar

$$\omega^2 - (ck_1)^2 - (ck_2)^2 - (ck_3)^2,$$

as multiplied by \hbar^2 , we obtain

$$E^2 - (cP_1)^2 - (cP_2)^2 - (cP_3)^2 = E^2 - cP^2 := m_0^2 c^4.$$

As we will discuss later, in special relativity, the energy of a particle is $E = \sqrt{P^2 c^2 + m_0^2 c^4}$, where m_0 is the rest mass of the particle.

For a rank-2 tensor, which can take one of the following forms

$$C_{\mu\nu}, \quad C_{\mu}{}^{\nu}, \quad C^{\mu}{}_{\nu}, \quad C^{\mu\nu},$$

one can lower or raise its indices to interchange between different forms, that is, for instance

$$C_{\mu\nu} = g_{\mu\mu'} g_{\nu\nu'} C^{\mu'\nu'}.$$

When we have two tensors A^{μ} and B^{ν} , we can form an outer product of the two by writing

$$C^{\mu\nu} = A^{\mu} B^{\nu} = \begin{bmatrix} A^0 B^0 & A^0 B^1 & A^0 B^2 & A^0 B^3 \\ A^1 B^0 & A^1 B^1 & A^1 B^2 & A^1 B^3 \\ A^2 B^0 & A^2 B^1 & A^2 B^2 & A^2 B^3 \\ A^3 B^0 & A^3 B^1 & A^3 B^2 & A^3 B^3 \end{bmatrix}.$$

Similarly, one can form

$$C^{\mu}{}_{\nu} = A^{\mu} B_{\nu}, \quad C_{\mu\nu} = A_{\mu} A_{\nu}, \quad W_{\mu\nu\lambda} = A_{\mu} C_{\nu\lambda},$$

and so on. When one has a tensor of rank higher than or equal to 2, such as T_{β}^{α} , one can contract the tensor by defining

$$T^{\alpha}{}_{\alpha} := T^0{}_0 + T^1{}_1 + T^2{}_2 + T^3{}_3,$$

which is the trace of T as a matrix. Further, one can define contraction

$$A^{\beta}{}_{\gamma} = C^{\alpha\beta}{}_{\alpha\gamma} = \sum_{\sigma} C^{\sigma\beta}{}_{\sigma\gamma}$$

and another contraction of the same tensor

$$B^{\beta}{}_{\gamma} = C^{\alpha\beta}{}_{\gamma\alpha} = \sum_{\sigma} C^{\sigma\beta}{}_{\gamma\sigma}.$$

Note that in general $A^{\beta}{}_{\gamma} \neq B^{\beta}{}_{\gamma}$, although they are both obtained by contracting $C^{\alpha\beta}{}_{\sigma\gamma}$. Note further here $C^{\alpha\beta}{}_{\alpha\beta}$ is a scalar, but is in general different from the scalar $C^{\alpha\beta}{}_{\beta\alpha}$, unless the tensor satisfies further symmetric property. If one would like to contract a tensor of the form $C^{\mu\nu}$, one would first need to lower one of the indices, that is,

$$C^{\mu}{}_{\mu} = C^{\mu\nu} g_{\mu\nu} = g^{\mu\nu} C_{\mu\nu}.$$

Contraction gives us a way to obtain a scalar from a tensor of any rank.

Next we will discuss the dynamics in special relativity. The principle is that, as we discussed in previous sections, all fundamental physics laws should take the form

$$\text{Scalar}_1 = \text{Scalar}_2, \quad \text{Vector}_1 = \text{Vector}_2, \quad (\text{Rank}-n) \text{ Tensor}_1 = (\text{Rank}-n) \text{ Tensor}_2,$$

such that they retain their forms under spacetime rotations, which include both spatial rotation and Lorentz boost.

Electromagnetic Theory in its Lorentz Invariant Form

The continuity equation for charge density and current is the natural consequence of charge conservation and principle of locality,

$$\partial_t \rho + \nabla \cdot \vec{j} = 0,$$

expanding we have the form

$$\frac{1}{c} \partial_t c\rho + \partial_x j_x + \partial_y j_y + \partial_z j_z = 0.$$

With our convention that $\partial_0 = (1/c) \partial_t$, we obtain

$$\partial_0(c\rho) + \partial_1 j_x + \partial_2 j_y + \partial_3 j_z = 0. \quad (*)$$

One might have noticed that, the RHS of (*) is a scalar, thus in order for (*) to retain its form under rotation, the LHS of (*) should also be a scalar. Thus, written in a more rigorous way, with the right subscripts and superscripts, we should have written

$$\partial_\mu j^\mu = \partial_0 j^0 + \partial_1 j^1 + \partial_2 j^2 + \partial_3 j^3 = 0,$$

where we define

$$j^\mu := (c\rho, j_x, j_y, j_z).$$

Written in matrix form, under a Lorentz boost, we can write

$$\begin{bmatrix} c\rho' \\ j'_x \\ j'_y \\ j'_z \end{bmatrix} = \begin{bmatrix} \gamma & -v\gamma/c & 0 & 0 \\ -v\gamma/c & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c\rho \\ j_x \\ j_y \\ j_z \end{bmatrix}.$$

For electric and magnetic field, we have the Maxwell's equations being satisfied

$$\nabla \cdot \vec{E} = \frac{\rho}{\epsilon_0}, \quad \nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}, \quad \nabla \cdot \vec{B} = 0, \quad \nabla \times \vec{B} = \mu_0 \vec{j} + \epsilon_0 \mu_0 \frac{\partial \vec{E}}{\partial t}.$$

Here we see six degrees of freedom here $E_x, E_y, E_z, B_x, B_y, B_z$. But we know that they are in fact not all independent. For instance, via $\nabla \cdot \vec{B} = 0$, we know that B_x is dependent on B_y and B_z , thus the system is seemingly too complicated. While notice that $\nabla \cdot \vec{B} = 0$ can be guaranteed if one write

$$\vec{B} = \nabla \times \vec{A}$$

for some vector field \vec{A} , and thus

$$\nabla \cdot \vec{B} = \nabla \cdot (\nabla \times \vec{A}) = 0.$$

We note that $\nabla \cdot \vec{B} = 0$ implies $\vec{B} = \nabla \times \vec{A}$ provided that the space of interest has no hole, while the other direction, $\vec{B} = \nabla \times \vec{A}$ implies $\nabla \cdot \vec{B} = 0$, always holds. Now with the assumption $\vec{B} = \nabla \times \vec{A}$, we can write

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} = -\frac{\partial}{\partial t} (\nabla \times \vec{A}) = -\nabla \times \left(\frac{\partial \vec{A}}{\partial t} \right).$$

Thus we obtain here

$$\nabla \times \left(\vec{E} + \frac{\partial \vec{A}}{\partial t} \right) = 0,$$

which implies

$$\vec{E} + \frac{\partial \vec{A}}{\partial t} = -\nabla \phi$$

for some scalar field ϕ . Now two of the Maxwell's equations have been used,

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}, \quad \nabla \cdot \vec{B} = 0,$$

and they lead us to

$$\vec{E} = -\nabla \phi - \frac{\partial \vec{A}}{\partial t}, \quad \vec{B} = \nabla \times \vec{A}.$$

In general, by the property of the two of the four Maxwell's equations that we have used, the choice of \vec{A} and ϕ are uniquely determined up to a gauge transformation

$$\vec{A} \rightarrow \vec{A} + \nabla \Lambda, \quad \phi \rightarrow \phi - \frac{\partial \Lambda}{\partial t},$$

where Λ is an arbitrary function of \vec{r} and t . Note that \vec{E} and \vec{B} are observables, while ϕ and \vec{A} are not. The values of ϕ and \vec{A} here are thus not unique, instead, they depend on the gauge choice. Note further that the quantum wavefunction of a quantum particle of charge q is also effected by the choice of gauge, up to a unobservable phase

$$\Psi \rightarrow \Psi e^{iq\Lambda/\hbar}.$$

As the gauge can be chosen arbitrarily, one will choose the gauge to simplify the calculation. For instance, one would use the transverse gauge such that

$$\nabla \cdot \vec{A} = 0,$$

and the Lorentz gauge such that

$$\nabla \cdot \vec{A} + \frac{1}{c^2} \frac{\partial \phi}{\partial t} = 0.$$

Note further that the Lorentz gauge is neither unique.

The other two Maxwell's equations lead us to

$$\begin{aligned}\nabla \times (\nabla \times \vec{A}) &= \nabla^2 \vec{A} + \nabla(\nabla \cdot \vec{A}) = \mu_0 \vec{j} - \epsilon_0 \mu_0 \frac{\partial}{\partial t} \nabla \phi - \epsilon_0 \mu_0 \frac{\partial^2 \vec{A}}{\partial t^2}, \\ \nabla \cdot \vec{E} &= -\nabla \cdot \nabla \phi - \nabla \cdot \frac{\partial \vec{A}}{\partial t} = \frac{\rho}{\epsilon_0}.\end{aligned}$$

Combining with $c^2 = (\epsilon_0 \mu_0)^{-1}$, we obtain

$$\nabla^2 \phi + \frac{\partial}{\partial t}(\nabla \cdot \vec{A}) = -\frac{\rho}{\epsilon_0}, \quad \nabla^2 \vec{A} - \frac{1}{c^2} \frac{\partial^2 \vec{A}}{\partial t^2} - \nabla \left(\nabla \cdot \vec{A} + \frac{1}{c^2} \frac{\partial \phi}{\partial t} \right) = -\mu_0 \vec{j}, \quad (3.11)$$

and that

$$\vec{B} = \nabla \times \vec{A}, \quad \vec{E} = -\nabla \phi - \frac{\partial \vec{A}}{\partial t}. \quad (3.12)$$

3.6.1 Transverse Gauge

Now suppose we have transverse gauge being satisfied, that is $\nabla \cdot \vec{A} = 0$, in which case the Maxwell's equations become

$$\nabla^2 \phi = -\frac{\rho}{\epsilon_0}, \quad \nabla^2 \vec{A} - \frac{1}{c^2} \frac{\partial^2 \vec{A}}{\partial t^2} - \frac{1}{c} \nabla \frac{\partial \phi}{\partial t} = -\mu_0 \vec{j},$$

where the first equation can be solved via

$$\phi(\vec{r}, t) = \frac{1}{4\pi\epsilon_0} \iiint \frac{\rho(\vec{r}', t)}{|\vec{r} - \vec{r}'|} dx' dy' dz'.$$

Note even though $\phi(\vec{r}, t)$ has a good form, but \vec{E} is still complicated as by definition

$$\vec{E} = -\frac{\partial \vec{A}}{\partial t} - \nabla \phi,$$

\vec{E} also depends on \vec{A} , and it is not easy to solve for \vec{A} . Furthermore, ϕ here is not local, as it depends on ρ everywhere in spacetime, which might also cause causality problem as $\rho(r', t)$ can effect $\phi(\vec{r}, t)$ no matter how far between \vec{r} and \vec{r}' is. While this does not cause problem to our theory as ϕ is not a physical observable quantity.

3.6.2 Lorentz Gauge

When Lorentz gauge is satisfied, then the second equation in (3.11) becomes

$$\nabla^2 \vec{A} - \frac{1}{c^2} \frac{\partial^2 \vec{A}}{\partial t^2} = -\mu_0 \vec{j}.$$

On the other hand, rewriting the Lorentz gauge as

$$\nabla \cdot \vec{A} = -\frac{1}{c^2} \frac{\partial \phi}{\partial t},$$

we see that the first equation in (3.11) becomes

$$\nabla^2 \phi - \frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} = -\frac{\rho}{\epsilon_0}.$$

That is, in Lorentz gauge, we have the Maxwell's equations become

$$\nabla^2 \vec{A} - \frac{1}{c^2} \frac{\partial^2 \vec{A}}{\partial t^2} = -\mu_0 \vec{j}, \quad \nabla^2 \phi - \frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} = -\frac{\rho}{\epsilon_0}, \quad (3.13)$$

and the corresponding definition of $\vec{B} = \nabla \times \vec{A}$ and $\vec{E} = -(\partial \vec{A} / \partial t) - \nabla \phi$. That is \vec{A} and ϕ follow the same form of equation, and these equations of motion are Lorentz invariant. In vacuum, $\rho = j = 0$, the solution to the form

$$\nabla^2 f - \frac{1}{c^2} \frac{\partial^2 f}{\partial t^2} = 0$$

is plane wave solution

$$f = \sum_{\vec{k}} A_{\vec{k}, \omega} e^{-\vec{k} \cdot \vec{r} - i\omega t},$$

where $\omega = c|\vec{k}|$, that gives a form of light as expected.

Here (3.13) written in a more covariant form, reads

$$\square \phi = \frac{\rho}{\epsilon_0}, \quad \square \vec{A} = \mu_0 \vec{j}.$$

In particular, we see that we have

$$\square \frac{\phi}{c} = \frac{\rho}{\mu_0 \epsilon_0} \frac{\mu_0}{c} = \rho c \mu_0.$$

Thus we can write

$$\square \begin{bmatrix} \phi/c \\ A_x \\ A_y \\ A_z \end{bmatrix} = \mu_0 \begin{bmatrix} \rho c \\ j_x \\ j_y \\ j_z \end{bmatrix}, \quad (3.14)$$

where we have seen that the RHS of (3.14) is a 4-vector in spacetime, thus forcing the LHS of (3.14) to be a 4-vector. As \square can be interpreted as a scalar, then we have $A^\mu = (\phi/c, A_x, A_y, A_z)$ being defined as a 4-vector. With A^μ defined, the Lorentz gauge condition becomes

$$\partial_\mu A^\mu = 0.$$

That is the Lorentz gauge condition is also Lorentz invariant. Note further that the transverse gauge condition

$$\nabla \cdot \vec{A} = 0$$

is not Lorentz invariant. Here (3.14), and thus (3.13), becomes, in its Lorentz invariant form,

$$\square A^\mu = \mu_0 j^\mu.$$

Under a gauge transformation, $\vec{A} \rightarrow \vec{A}' = \vec{A} + \nabla\Lambda$, $\phi \rightarrow \phi' = \phi - (\partial\Lambda\partial t)$, it follows immediately that we have

$$A'^\mu = A^\mu + \partial^\mu\Lambda.$$

From now on, we will mainly focus on Lorentz gauge, where the electromagnetic theory will be described by the equations

$$\square A^\mu = \mu_0 j^\mu, \quad \partial_\mu j^\mu = 0, \quad \partial_\mu A^\mu = 0. \quad (3.15)$$

4

With the essential tools developed in the previous chapter, we have the electromagnetic theory described by the equations (3.15),

$$\square A^\mu = \mu_0 j^\mu, \quad \partial_\mu j^\mu = 0, \quad \partial_\mu A^\mu = 0. \quad (4.1)$$

We first would like to recover \vec{B} and \vec{E} from (4.1), and one can expect we shall obtain them via outer products, instead of inner product.

If one has 4-vectors X^μ and Y^μ , with their usual \mathbb{R}^n -vector form denoted as X and Y , we have

$$X \otimes Y = X^\mu Y^\mu = \begin{bmatrix} X^0 Y^0 & X^0 Y^1 & X^0 Y^2 & X^0 Y^3 \\ X^1 Y^0 & X^1 Y^1 & X^1 Y^2 & X^1 Y^3 \\ X^2 Y^0 & X^2 Y^1 & X^2 Y^2 & X^2 Y^3 \\ X^3 Y^0 & X^3 Y^1 & X^3 Y^2 & X^3 Y^3 \end{bmatrix}.$$

Here we will introduce the exterior product of X^μ and Y^μ , which gets us an antisymmetric tensor,

$$X \wedge Y = \begin{bmatrix} 0 & \frac{X^0 Y^1 - X^1 Y^0}{2} & \frac{X^0 Y^2 - X^2 Y^0}{2} & \frac{X^0 Y^3 - X^3 Y^0}{2} \\ \frac{X^1 Y^0 - X^0 Y^1}{2} & 0 & \frac{X^1 Y^2 - X^2 Y^1}{2} & \frac{X^1 Y^3 - X^3 Y^1}{2} \\ \frac{X^2 Y^0 - X^0 Y^2}{2} & \frac{X^2 Y^1 - X^1 Y^2}{2} & 0 & \frac{X^2 Y^3 - X^3 Y^2}{2} \\ \frac{X^3 Y^0 - X^0 Y^3}{2} & \frac{X^3 Y^1 - X^1 Y^3}{2} & \frac{X^3 Y^2 - X^2 Y^3}{2} & 0 \end{bmatrix}.$$

Similar idea leads to

$$\partial^\mu \wedge A^\mu = \frac{\partial^\mu A^\mu - \partial^\mu A^\mu}{2} = \frac{1}{2} \begin{bmatrix} 0 & \partial^0 A^1 - \partial^1 A^0 & \partial^0 A^2 - \partial^2 A^0 & \partial^0 A^3 - \partial^3 A^0 \\ \partial^1 A^0 - \partial^0 A^1 & 0 & \partial^1 A^2 - \partial^2 A^1 & \partial^1 A^3 - \partial^3 A^1 \\ \partial^2 A^0 - \partial^0 A^2 & \partial^2 A^1 - \partial^1 A^2 & 0 & \partial^2 A^3 - \partial^3 A^2 \\ \partial^3 A^0 - \partial^0 A^3 & \partial^3 A^1 - \partial^1 A^3 & \partial^3 A^2 - \partial^2 A^3 & 0 \end{bmatrix}.$$

We notice that we have

$$\begin{aligned} \partial^0 A^1 - \partial^1 A^0 &= \frac{1}{c} \partial_t A_x + \partial_x \frac{\phi}{2} = \frac{1}{c} (\partial_t A_x + \partial_x \phi) = -E_x/c, \\ \partial^1 A^2 - \partial^2 A^1 &= -\partial_x A_y + \partial_y A_x = -(\nabla \times \vec{A})_z = -B_z. \end{aligned}$$

Thus similarly computed for other components, we obtain

$$\partial^\mu \wedge A^\mu = \frac{1}{2} \begin{bmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & -B_z & B_y \\ E_y/c & B_z & 0 & -B_x \\ E_z/c & -B_y & B_x & 0 \end{bmatrix}.$$

Expressed in a more compact way, using a single tensor, we can define

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu = 2(\partial^\mu \wedge A^\nu).$$

Under Lorentz boost $L^\mu{}_{\mu'}$, or in its matrix form L , we can write

$$F^{\mu\nu} \rightarrow F'^{\mu\nu} = L^\mu{}_{\mu'} L^\nu{}_{\nu'} F^{\mu'\nu'}.$$

or in its matrix form,

$$F \rightarrow F' = L F L^T,$$

where one can compute

$$\begin{aligned} & \begin{bmatrix} 0 & -E'_x/c & -E'_y/c & -E'_z/c \\ E'_x/c & 0 & -B'_z & B'_y \\ E'_y/c & B'_z & 0 & -B_x \\ E'_z/c & -B'_y & B'_x & 0 \end{bmatrix} \\ &= \begin{bmatrix} \gamma & -v\gamma/c & 0 & 0 \\ -v\gamma/c & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & -B_z & B_y \\ E_y/c & B_z & 0 & -B_x \\ E_z/c & -B_y & B_x & 0 \end{bmatrix} \begin{bmatrix} \gamma & -v\gamma/c & 0 & 0 \\ -v\gamma/c & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \end{aligned}$$

and obtain

$$\begin{aligned} E'_x &= E_x, & E'_y &= \gamma(E_y - vB_z), & E'_z &= \gamma(E_z + vB_y) \\ B'_x &= B_x, & B'_y &= \gamma(B_y + vE_z/c^2), & B'_z &= \gamma(B_z - vE_y/c^2). \end{aligned}$$

Furthermore, with the definition of $F^{\mu\nu}$, one finds that the Maxwell's equations become

$$\partial_\mu F^{\mu\nu} = \mu_0 j^\nu, \quad \epsilon_{\mu\nu\lambda\epsilon} \partial^\nu F^{\lambda\epsilon} = 0,$$

where the first equation generalizes the four equations (components) in

$$\nabla \cdot \vec{E} = \frac{\rho}{\epsilon_0}, \quad \nabla \times \vec{B} = \mu_0 \vec{j} + \epsilon_0 \mu_0 \frac{\partial \vec{E}}{\partial t},$$

and the second equation generalizes the other four equations (components) in

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}, \quad \nabla \cdot \vec{B} = 0.$$

Here $\epsilon_{\mu\nu\epsilon\lambda}$ is the 4-dimensional Levi-Civita symbol.

Conservation of Energy and Momentum

We have seen that energy and momentum can be combined into a 4-vector

$$p^\mu = (E/c, p_x, p_y, p_z),$$

and thus the conservation of energy and conservation of momentum become a single conservation law. Now we consider we have a decay of a particle A into substance B and substance C . We choose a reference frame where A is stationary. As a result, before the decay, we have

$$p_A^\mu = \begin{bmatrix} m_A c^2/c \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} m_A c \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

After the decay, we obtain two substances B and C , moving with velocity \vec{v}_B and \vec{v}_C respectively, then we can write

$$p_A^\mu = \begin{bmatrix} m_A c \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \gamma_B m_B c \\ m_B \gamma_B v_{B,x} \\ m_B \gamma_B v_{B,y} \\ m_B \gamma_B v_{B,z} \end{bmatrix} + \begin{bmatrix} \gamma_C m_C c \\ m_C \gamma_C v_{C,x} \\ m_C \gamma_C v_{C,y} \\ m_C \gamma_C v_{C,z} \end{bmatrix} = p_B^\mu + p_C^\mu.$$

where we have defined

$$\gamma_B = \frac{1}{(1 - v_B^2/c^2)^{1/2}}, \quad \gamma_C = \frac{1}{(1 - v_C^2/c^2)^{1/2}}.$$

Thus we have here the energy conservation

$$m_A = m_B \gamma_B + m_C \gamma_C, \tag{4.2}$$

and momentum conservation

$$m_B \gamma_B \vec{v}_B = -m_C \gamma_C \vec{v}_C,$$

from which we see that \vec{v}_B and \vec{v}_C are antiparallel, and that we have their amplitude being conserved, that is

$$m_B \gamma_B v_B = m_C \gamma_C v_C. \tag{4.3}$$

Assuming that m_A, m_B, m_C are all known, one can solve for v_B and v_C ,

$$v_B = c \left(1 - \frac{4m_A^2 m_B^2}{(m_A^2 + m_B^2 - m_C^2)^2} \right)^{1/2}, \quad v_C = c \left(1 - \frac{4m_A^2 m_C^2}{(m_A^2 + m_B^2 - m_C^2)^2} \right)^{1/2}. \tag{4.4}$$

From (4.2), we observe that we have

$$m_B + m_C < m_A,$$

and from (4.4) we observe that we have $v_B, v_C < c$ as expected. Here (4.2) can be generalized to multiparticles in a decay process,

$$M > m_1 + m_2 + m_3 + \cdots,$$

where we have particle of mass M decaying into particles with mass m_i .

Next we consider the Compton scattering, an incoming photon with energy $E = \hbar\omega$ is scattered by an electron. The electron, initially at rest in the inertial frame, has a recoiling velocity after scattering. Consider the electron has a wave-vector $\vec{k} = (k, 0, 0)$, with $\omega = ck$. Then before scattering, we can write

$$p_{\text{photon}}^\mu + p_{\text{electron}}^\mu = \begin{bmatrix} \hbar\omega/c \\ \hbar k \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} m_e c \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Suppose further that, after the scattering, the photon leaves with momentum $\hbar k'$, energy $\omega' = ck'$, making an angle θ with respect to the x -axis. On the other hand, the electron leaves with velocity magnitude v_e , making an angle ϕ with respect to the x -axis. Thus combining, after scattering, we can write

$$p_{\text{photon}}^\mu + p_{\text{electron}}^\mu = \begin{bmatrix} \hbar\omega'/c \\ \hbar k' \cos(\theta) \\ \hbar k' \sin(\theta) \\ 0 \end{bmatrix} + m_e \gamma_e \begin{bmatrix} c \\ v_e \cos(\phi) \\ -v_e \sin(\phi) \\ 0 \end{bmatrix},$$

where we have defined

$$\gamma_e = \frac{1}{(1 - v_e^2/c^2)^{1/2}}.$$

Thus the energy conservation gives

$$\hbar\omega + m_e c^2 = \hbar\omega' + m_e \gamma_e c^2,$$

and momentum conservation gives

$$\hbar\omega = \hbar\omega' \cos(\theta) + cm_e \gamma_e v_e \cos(\phi), \quad 0 = \hbar\omega' \sin(\theta) - cm_e \gamma_e v_e \sin(\phi).$$

Here we have four unknowns, ω' , θ , ϕ , and v_e , and thus we suppose further that θ is measured, and solve for the other three using the conservation laws. One obtains

$$\frac{\omega}{\omega'} = \frac{1}{1 + (1 - \cos(\theta))(\hbar\omega)/(m_e c^2)},$$

showing that we have

$$\omega' < \omega,$$

that is Compton scattering is in fact an inelastic scattering, the frequency of light decreases in the scattering process. Furthermore, when ω is small enough, such that $\hbar\omega \ll m_e c^2$, we have roughly $\omega' = \omega$, called the Thomson scattering, which is a form of elastic scattering. Note that the statement $\hbar\omega \ll m_e c^2$ is that the photon energy is much less than the electron scattering, in order to have Thomson scattering.

Derivation of Maxwell's Equations

In this section, we will derive Maxwell's equations using the special theory of relativity, together with principle of locality, principle of least action, and symmetries.

Here the action of our system of interest can be decomposed into three components

$$S = S_{\text{EM field}} + S_{\text{matter}} + S_{\text{coupling}},$$

where $S_{\text{EM field}}$ depends on only \vec{E} and \vec{B} , S_{matter} depends only on ρ and \vec{j} , and S_{coupling} depends on all \vec{E} , \vec{B} , ρ , and \vec{j} . We note further that, using \vec{E} and \vec{B} to derive the theory requires constraints such as

$$\nabla \cdot \vec{B} = 0,$$

which is not convenient and not desired. Thus we will instead use the scalar potential ϕ and vector potential \vec{A} to approach the theory. In this case, we thus can write

$$S = S_{\text{EM field}}(\phi, \vec{A}) + S_{\text{matter}}(\rho, \vec{j}) + S_{\text{coupling}}(\rho, \vec{j}, \phi, \vec{A}).$$

The principle of locality states further that, the Lagrangian density of the theory should involve only the terms $\phi, \vec{A}, \rho, \vec{j}$ and their products and derivatives, but should not involve the terms like $\rho(x)\rho(x')/(x-x')^2$.

The symmetry involved in our theory are:

- (1) **4-dimensional rotation symmetry**, both Lorentz boosts and spatial rotations,
- (2) **spatial parity symmetry** $\vec{r} \rightarrow -\vec{r}$,
- (3) **time reversal symmetry** $t \rightarrow -t$, and
- (4) **gauge symmetry**,

$$\begin{bmatrix} \phi/c \\ A_x \\ A_y \\ A_z \end{bmatrix} \rightarrow \begin{bmatrix} \phi/c \\ A_x \\ A_y \\ A_z \end{bmatrix} + \begin{bmatrix} \partial_t/c \\ -\partial_x \\ -\partial_y \\ -\partial_z \end{bmatrix} \Lambda. \quad (4.5)$$

Under transformation of all those symmetries, S and \mathcal{L} remain unchanged.

Another assumption that we impose here is that we are in the weak field limit, with slow variation. That is, we have ρ and \vec{j} being small, and thus \mathcal{L} is a quadratic form of fields ϕ, \vec{A}, ρ and \vec{j} , as well as their derivatives. Slow varying fields requires that all quantities to be slowly varying, and thus we only need to keep the lowest order derivative of a quantity, where higher order derivatives can be ignored.

To construct $\mathcal{L}_{\text{EM field}}$ which contributes to $S_{\text{EM field}}$, we start from $A^\mu = (\phi/c, \vec{A})$. We can make a scalar from A^μ by considering $A^\mu A_\mu$, that is, we first look at the term $\alpha = A^\mu A_\mu$, a straightforward computation leads to

$$\alpha A^\mu A_\mu = \alpha \left(\phi^2/c^2 - |\vec{A}|^2 \right).$$

Such a term obeys most of the symmetry requirements, but is not allowed by the gauge symmetry. Consider a gauge transformation characterized by (4.5), $A^\mu \rightarrow A'^\mu$, it is easy

to verify that $A_\mu A^\mu \neq A'^\mu A'_\mu$, and thus gauge symmetry prohibited the term $A_\mu A^\mu$. It is also easy to check that any terms like $A \partial A$ would also be prohibited by the gauge symmetry, while terms like $\partial A \partial A$ are gauge invariant. Furthermore, one can in fact obtain Lorentz invariant scalar from \vec{E} and \vec{B} . An easy exercise shows that the only Lorentz scalars that could be constructed from \vec{E} and \vec{B} are $E^2 - c^2 B^2$ and $\vec{E} \cdot \vec{B}$. That is, we can look at the terms

$$\alpha(E^2 - c^2 B^2) + \theta \vec{E} \cdot \vec{B},$$

with $\alpha, \theta \in \mathbb{R}$. Under spatial parity symmetry, $\vec{E} \rightarrow -\vec{E}$ and $\vec{B} \rightarrow \vec{B}$, thus $\vec{E} \cdot \vec{B} \rightarrow -\vec{E} \cdot \vec{B}$ under spatial parity. As we require S to remain invariant under spatial parity transformation, thus the term $\theta \vec{E} \cdot \vec{B}$ is prohibited by spatial parity symmetry. That is concluding we have

$$\mathcal{L}_{\text{EM field}} = \alpha(E^2 - c^2 B^2).$$

Now we construct $\mathcal{L}_{\text{coupling}}$ which contributes to S_{coupling} . We know here $\mathcal{L}_{\text{coupling}}$ is a function of A^μ and j^μ . Notice that $A^\mu A_\mu$ violates Gauge symmetry and $j^\mu j_\mu$ is considered to be matter contribution to S , thus we keep only the $A^\mu j_\mu$ term, and thus we write

$$\mathcal{L}_{\text{coupling}} = -\alpha' A^\mu j_\mu = -\alpha'(\phi\rho - \vec{j} \cdot \vec{A}).$$

Here for now we assume that $\mathcal{L}_{\text{matter}} = 0$ and thus $S_{\text{matter}} = 0$. Thus we can write, for now,

$$\begin{aligned} \mathcal{L} &= \alpha(E^2 - c^2 B^2) - \alpha'(\phi\rho - \vec{A} \cdot \vec{j}) \\ &= \alpha(\partial_t \vec{A} \cdot \partial_t \vec{A} + \nabla\phi \cdot \nabla\phi - c^2(\nabla \times \vec{A})(\nabla \times \vec{A})) - \alpha'(\phi\rho - \vec{A} \cdot \vec{j}) \\ &= \alpha((\partial_t A_x)^2 + (\partial_t A_y)^2 + (\partial_t A_z)^2 + (\partial_x \phi)^2 + (\partial_y \phi)^2 + (\partial_z \phi)^2) \\ &\quad + \alpha(2(\partial_t A_x)(\partial_x \phi) + 2(\partial_t A_y)(\partial_y \phi) + 2(\partial_t A_z)(\partial_z \phi)) \\ &\quad - c^2\alpha((\partial_x A_y - \partial_y A_x)^2 + (\partial_y A_z - \partial_z A_y)^2 + (\partial_z A_x - \partial_x A_z)^2) \\ &\quad - \alpha'(\phi\rho - A_x j_x - A_y j_y - A_z j_z) \end{aligned} \tag{4.6}$$

Applying the equation of motion

$$\partial_t \frac{\delta \mathcal{L}}{\delta \partial_t \phi} + \partial_x \frac{\delta \mathcal{L}}{\delta \partial_x \phi} + \partial_y \frac{\delta \mathcal{L}}{\delta \partial_y \phi} + \partial_z \frac{\delta \mathcal{L}}{\delta \partial_z \phi} = \frac{\delta \mathcal{L}}{\delta \phi},$$

one can show that we obtain

$$-2\alpha(\partial_x E_x + \partial_y E_y + \partial_z E_z) = -2\alpha(\nabla \cdot \vec{E}) = -\alpha' \rho.$$

Defining here $2\alpha/\alpha' = \epsilon_0$, we obtain

$$\nabla \cdot \vec{E} = \frac{\rho}{\epsilon_0}.$$

Now apply the equation of motion

$$\partial_t \frac{\delta \mathcal{L}}{\delta \partial_t A_x} + \partial_x \frac{\delta \mathcal{L}}{\delta \partial_x A_x} + \partial_y \frac{\delta \mathcal{L}}{\delta \partial_y A_x} + \partial_z \frac{\delta \mathcal{L}}{\delta \partial_z A_x} = \frac{\delta \mathcal{L}}{\delta A_x}.$$

For an example calculation we have here

$$\frac{\delta \mathcal{L}}{\delta \partial_t A_x} = \alpha \frac{\delta (\partial_t A_x)^2}{\delta \partial_t A_x} + 2\alpha \frac{\delta (\partial_t A_x \partial_x \phi)}{\delta \partial_t A_x} = 2\alpha \partial_t A_x + 2\alpha \partial_x \phi = 2\alpha (\partial_t A_x + \partial_x \phi) = -2\alpha E_x .$$

On the other hand, we have

$$\frac{\delta \mathcal{L}}{\delta A_x} = \frac{\delta (\alpha' j_x A_x)}{\delta A_x} = \alpha' j_x , \quad \frac{\delta \mathcal{L}}{\delta \partial_y A_x} = 2\alpha c^2 (\partial_x A_y - \partial_y A_x) = 2\alpha c^2 B_z .$$

Thus combining all we obtain

$$-2\alpha \partial_t E_x + 0 + 2\alpha c^2 \partial_y B_z - 2\alpha c^2 \partial_z B_y = \alpha' j_x .$$

Rearranging we obtain

$$(\partial_y B_z - \partial_z B_y) - \frac{1}{c^2} \partial_t E_x = \frac{1}{\epsilon_0 c^2} j_x .$$

We define here μ_0 to be the quantity that satisfies

$$c^2 = \frac{1}{\epsilon_0 \mu_0} ,$$

and we obtain

$$(\nabla \times B)_x = \partial_y B_z - \partial_z B_y = \mu_0 j_x + \epsilon_0 \mu_0 \partial_t E_x .$$

Similar calculation, using the equation of motion, on A_y and A_z leads to the final result

$$\nabla \cdot \vec{E} = \rho / \epsilon_0 , \quad \nabla \times \vec{B} = \epsilon_0 \mu_0 \partial_t \vec{E} + \mu_0 \vec{j} ,$$

recovering Maxwell's equations. Note that the other two equations in Maxwell's equations

$$\nabla \cdot \vec{B} = 0 , \quad \nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} ,$$

are automatically satisfied by defining

$$\vec{E} = -\partial_t \vec{A} - \nabla \phi , \quad \vec{B} = \nabla \times \vec{A} .$$

With the definition of ϵ_0 and μ_0 , now (4.6) becomes

$$\mathcal{L} = \alpha' \left(\frac{\epsilon_0}{2} E^2 - \frac{\epsilon_0 c^2}{2} B^2 - \phi \rho + \vec{A} \cdot \vec{j} \right) = \alpha' \left(\frac{\epsilon_0}{2} E^2 - \frac{B^2}{2\mu_0} - \phi \rho + \vec{A} \cdot \vec{j} \right) .$$

Note that renormalization of \mathcal{L} does not effect the equation of motion and thus has no effect on action, we take $\alpha' = 1$. Thus we write

$$\mathcal{L} = \frac{\epsilon_0}{2} E^2 - \frac{B^2}{2\mu_0} - \phi \rho + \vec{A} \cdot \vec{j} .$$

Point Charge

As we write the Lagrangian $L = \text{K.E.} - \text{P.E.}$, the potential energy contribution to L comes from the coupling Lagrangian density $\mathcal{L}_{\text{coupling}}$. Thus one would expect $\mathcal{L}_{\text{matter}}$, and thus L_{matter} , depends only on the velocity of matter.

One thing to notice is that, the total Lagrangian

$$L = \iiint \mathcal{L} \, dx \, dy \, dz$$

is not a scalar, as \mathcal{L} is a scalar while $dx \, dy \, dz$ is not a scalar due to length contraction. On the other hand,

$$S = \iiint \mathcal{L} \, dt \, dx \, dy \, dz$$

is a scalar as $dt \, dx \, dy \, dz$ is scalar (dt has time dilation, and $dx \, dy \, dz$ has length contraction). However, one can write

$$S = \int L \, dt = \int \gamma L \, d\tau,$$

and thus one expects that γL to be a scalar. To obtain a scalar from velocity, we recall that

$$u^\mu := \frac{dx^\mu}{d\tau} = \gamma \begin{bmatrix} c \\ v_x \\ v_y \\ v_z \end{bmatrix},$$

which gives us a trivial scalar

$$u^\mu u_\mu = \frac{c^2}{c^2 - v^2} (c^2 - v_x^2 - v_y^2 - v_z^2) = c^2.$$

Thus γL is a constant, that is

$$L = -\alpha \cdot \sqrt{1 - v^2/c^2},$$

for some $\alpha \in \mathbb{R}$. It turns out that we have $\alpha = m_0 c^2$. To see this, we note that the momentum can be written as

$$p_i = \frac{\delta L}{\delta v_i} = \frac{\alpha}{\sqrt{1 - v^2/c^2}} \frac{v_i}{c^2} = \alpha \gamma v_i / c^2,$$

by requiring $p_i = m_0 \gamma v_i$, we see here $\alpha = m_0 c^2$. Then we can compute the Hamiltonian,

$$H = \vec{p} \cdot \vec{v} - L = \frac{\alpha}{c^2} \gamma v^2 + \alpha \sqrt{1 - v^2/c^2} = \frac{m_0}{\sqrt{1 - v^2/c^2}} (v^2 + c^2(1 - v^2/c^2)) = \frac{m_0 c^2}{\sqrt{1 - v^2/c^2}},$$

as predicted. Concluding here, we see that

$$L = m_0 c^2 \sqrt{1 - v^2/c^2}, \quad \vec{p} = \frac{m_0}{\sqrt{1 - v^2/c^2}} \vec{v}, \quad H = \frac{m_0 c^2}{\sqrt{1 - v^2/c^2}}.$$

Notice here $L = \text{K.E.} - \text{P.E.}$ does not hold here.

4.3.1 A point charge coupled with E and B field

From previous result, we have that

$$\mathcal{L}_{\text{coupling}} = -\rho\phi + \vec{j} \cdot \vec{A}.$$

Thus for a point charge at \vec{r}_0 , $\rho = q\delta(\vec{r} - \vec{r}_0)$, and $\vec{j} = q\vec{v}\delta(\vec{r} - \vec{r}_0)$, and thus we can compute the Lagrangian,

$$\begin{aligned} L_{\text{coupling}} &= \iiint dx dy dz (-\rho\phi + \vec{j} \cdot \vec{A}) \\ &= \iiint dx dy dz \left(-q\phi(\vec{r})\delta(\vec{r} - \vec{r}_0) + q\vec{v} \cdot \vec{A}(\vec{r})\delta(\vec{r} - \vec{r}_0) \right) \\ &= -q\phi(\vec{r}_0) + q\vec{v} \cdot \vec{A}(\vec{r}_0). \end{aligned}$$

Thus for a point charge at \vec{r} , we obtain

$$\gamma L_{\text{coupling}} = -\gamma q\phi(\vec{r}) + \gamma q\vec{v} \cdot \vec{A}(\vec{r}). \quad (4.7)$$

By denoting

$$u^\mu = \gamma \begin{bmatrix} c \\ v_x \\ v_y \\ v_z \end{bmatrix},$$

we thus obtain a compact form of (4.7),

$$\gamma L_{\text{coupling}} = qu^\mu A_\mu. \quad (4.8)$$

In fact (4.8) can be deduced from the fact that, as $\gamma L_{\text{coupling}}$ is required to be a scalar, and should contain information about both the matter (moving with velocity \vec{v}) and the field (described by A_μ), we thus obtain $\gamma L_{\text{coupling}} = \beta u^\mu A_\mu$, where we define $\beta = -q$.

Here the evolution of the fields are not of our interest, thus focus on

$$L = L_{\text{matter}} + L_{\text{coupling}},$$

where we can write, from the result above,

$$L(\vec{r}, \vec{v}) = -m_0 c^2 \sqrt{1 - v^2/c^2} - q\phi(\vec{r}) + q\vec{v} \cdot \vec{A}(\vec{r}), \quad (4.9)$$

where \vec{r} is the position of the charge. Then we can compute

$$p_x = \frac{\delta L}{\delta v_x} = \frac{m_0}{\sqrt{1 - v^2/c^2}} v_x + qA_x = mv_x + qA_x,$$

where we have defined $m := m_0\gamma$. Thus here we define

$$\vec{P} = m\vec{v} + q\vec{A}$$

to be the momentum of the particle, and $m\vec{v} = \gamma m_0 \vec{v}$ is the momentum of the particle without coupling to the E and B field. For notation here, we will denote $\vec{p} = m\vec{v}$ to be the momentum of the charged particle without coupling to the fields, and \vec{P} to be the

momentum of the charged particle coupling to the field.

Next we can compute

$$H = \vec{P} \cdot \vec{v} - L = \left(\frac{m_0}{\sqrt{1 - v^2/c^2}} \vec{v} + q \vec{A} \right) \cdot \vec{v} - L = \frac{m_0 c^2}{\sqrt{1 - v^2/c^2}} + q \phi(\vec{r}),$$

where the second term $q \phi(\vec{r})$ is interpreted as the electric potential energy. Note that H should instead be a function of \vec{r} and \vec{P} instead of \vec{r} and v , thus we take a further step here. Notice that we can write

$$E = \sqrt{p^2 c^2 + m_0^2 c^4} = \sqrt{m^2 v^2 c^2 + m_0^2 c^4} = \sqrt{\frac{m_0^2 v^2 c^2}{\sqrt{1 - v^2/c^2}} + m_0^2 c^4} = \frac{m_0 c^2}{\sqrt{1 - v^2/c^2}} = mc^2.$$

Thus concluding we see here

$$H(\vec{P}, \vec{r}) = \sqrt{p^2 c^2 + m_0^2 c^4} + q \phi(\vec{r}) = \sqrt{(\vec{P} - q \vec{A}) \cdot (\vec{P} - q \vec{A}) c^2 + m_0^2 c^4} + q \phi(\vec{r}).$$

The Hamiltonian of the neutral particle (without charge, thus $\vec{P} = \vec{p}$ in this case) thus have the form

$$H(\vec{P}, \vec{r}) = \sqrt{(\vec{P} \cdot \vec{P}) c^2 + m_0^2 c^4} = \sqrt{p^2 c^2 + m_0^2 c^4},$$

as expected.

Thus going from neutral particle $q = 0$ to charged particle $q \neq 0$, a term $q \phi(\vec{r})$ should be added to the Hamiltonian, and \vec{P} becomes $\vec{p} - q \vec{A}$, giving rise to the minimal coupling.

Furthermore, from (4.9) we can compute the equation of motion from such Lagrangian. For instance, the x -component reads

$$\frac{d}{dt} \frac{\delta L}{\delta(dx/dt)} = \frac{dP_x}{dt} = \frac{d}{dt} (mv_x + qA_x) = \frac{dp_x}{dt} + q \frac{dA_x}{dt} = \frac{\delta L}{\delta x} = -q \partial_x \phi + q \partial_x (\vec{v} \cdot \vec{A}).$$

Note that in general \vec{A} is not just a function of \vec{r} (the position of the point charge) only, it can also depend on time t if the fields evolve. Thus we compute

$$\frac{dA_x}{dt} \Big|_{\vec{r}(t), t} = \frac{\partial A_x}{\partial t} + \frac{\partial A_x}{\partial x} \frac{dx}{dt} + \frac{\partial A_x}{\partial y} \frac{dy}{dt} + \frac{\partial A_x}{\partial z} \frac{dz}{dt} = \frac{\partial A_x}{\partial t} + \vec{v} \cdot (\nabla A_x).$$

Now combining we have

$$\frac{dp_x}{dt} + q \frac{\partial A_x}{\partial t} + q \vec{v} \cdot (\nabla A_x) = \frac{\delta L}{\delta x} = -q \partial_x \phi + q \partial_x (\vec{v} \cdot \vec{A}).$$

Rearranging we obtain

$$\frac{dp_x}{dt} = -q \left(\partial_x \phi + \frac{\partial A_x}{\partial t} \right) + q \left(\partial_x (\vec{v} \cdot \vec{A}) - \vec{v} \cdot (\nabla A_x) \right).$$

Similarly computation for y - and z - components gives

$$\begin{aligned} \frac{dp_y}{dt} &= -q \left(\partial_y \phi + \frac{\partial A_y}{\partial t} \right) + q \left(\partial_y (\vec{v} \cdot \vec{A}) - \vec{v} \cdot (\nabla A_y) \right), \\ \frac{dp_z}{dt} &= -q \left(\partial_z \phi + \frac{\partial A_z}{\partial t} \right) + q \left(\partial_z (\vec{v} \cdot \vec{A}) - \vec{v} \cdot (\nabla A_z) \right). \end{aligned}$$

Combining we have

$$\frac{d\vec{p}}{dt} = -q \left(\nabla\phi + \frac{\partial\vec{A}}{\partial t} \right) + q\vec{v} \times (\nabla \times \vec{A}) = q\vec{E} + q(\vec{v} \times \vec{B}),$$

which agrees with the classical limit.