MAT 560 Mathematical Physics I. Classical Field Theory

Leon A. Takhtajan

Department of Mathematics, Stony Brook University, Stony Brook, NY 11794-3651, USA

E-mail address: leontak@math.sunysb.edu

Contents

Part 1

Classical Mechanics

LECTURE 1

Equations of motion

We assume that the reader is familiar with the basic notions from the theory of smooth — C^∞ — manifolds, and recall here the standard notation. Unless it is stated explicitly otherwise, all maps are assumed to be smooth, and all functions are assumed to be smooth and real-valued. Local coordinates $\boldsymbol{q} = (q^1, \dots, q^n)$ on a smooth n-dimensional manifold M at point $q \in M$ are Cartesian coordinates on $\varphi(U) \subset \mathbb{R}^n$, where (U, φ) is a coordinate chart on M centered at $q \in U$. For $f : U \to \mathbb{R}^n$ we denote $(f \circ \varphi^{-1})(q^1, \dots, q^n)$ by $f(\boldsymbol{q})$, and we let

$$\frac{\partial f}{\partial \boldsymbol{q}} = \left(\frac{\partial f}{\partial q^1}, \dots, \frac{\partial f}{\partial q^n} \right)$$

stand for the gradient of a function f at point $\boldsymbol{q} \in \mathbb{R}^n$ with Cartesian coordinates $(q^1, \dots, q^n)$. We denote by

$$\mathcal{A}^\bullet(M) = \bigoplus_{k=0}^{n} \mathcal{A}^k(M)$$

the graded algebra of smooth differential forms on M with respect to the wedge product, and by d the de Rham differential — a graded derivation of $\mathcal{A}^\bullet(M)$ of degree 1, such that df is a differential of a function $f \in \mathcal{A}^0(M) = C^\infty(M)$. Let $\mathrm{Vect}(M)$ be the Lie algebra of smooth vector fields on M with the Lie bracket $[\ ,\ ]$, given by a commutator of vector fields. For $X \in \mathrm{Vect}(M)$ we denote by $\mathcal{L}_X$ and i_X, respectively, the Lie derivative along X and the inner product with X. The Lie derivative is a degree 0 derivation of $\mathcal{A}^\bullet(M)$ which commutes with d and satisfies $\mathcal{L}_X(f) = X(f)$ for $f \in \mathcal{A}^0(M)$, and the inner product is a degree -1 derivation of $\mathcal{A}^\bullet(M)$ satisfying $i_X(f) = 0$ and $i_X(df) = X(f)$ for $f \in \mathcal{A}^0(M)$. They satisfy Cartan formulas

$$\mathcal{L}_X = i_X \circ d + d \circ i_X = (d + i_X)^2, \tag{1.1}$$

$$i_{[X,Y]} = \mathcal{L}_X \circ i_Y - i_Y \circ \mathcal{L}_X. \tag{1.2}$$

For a smooth mapping of manifolds $f : M \to N$ we denote by $f_* : TM \to TN$ and $f^* : T^*N \to T^*M$, respectively, the induced mappings on tangent and cotangent bundles. Other notations, including those traditional for classical mechanics, will be introduced in the main text.

1.1. Generalized coordinates

Classical mechanics describes systems of finitely many interacting *particles*. Position of a system in space is specified by the positions of its particles and determines a point in some smooth, finite-dimensional manifold M, called a *configuration space* of the system. Coordinates on M are called *generalized coordinates* of a system, and the dimension $n = \dim M$ is called the number of *degrees of freedom*.

A *state* of the system at any instant of time is described by a point $q \in M$ and by a tangent vector $v \in T_qM$ at this point. The basic principle of classical mechanics is the *Newton-Laplace determinacy principle*, which asserts that a state of the system at a given instant of time completely determines its motion at all times t (in the future and in the past). The motion is described by a *classical trajectory* — a path $\gamma(t)$ in the configuration space M. In generalized coordinates $\gamma(t) = (q^1(t), \dots, q^n(t))$, and corresponding derivatives $\dot{q}^i = \dfrac{dq^i}{dt}$ are called *generalized velocities*. The Newton-Laplace principle is a fundamental experimental fact. It implies that *generalized accelerations* $\ddot{q}^i = \dfrac{d^2q^i}{dt^2}$ are uniquely determined by generalized coordinates q^i and generalized velocities $\dot{q}^i$, so that classical trajectories satisfy a system of second order ordinary differential equations, called *equations of motion*.

A *Lagrangian system* on a configuration space M is defined by a smooth, real-valued function L on $TM \times \mathbb{R}$ — the direct product of a tangent bundle TM of M and the time axis[1] — called the *Lagrangian function* (or simply, *Lagrangian*).

1.2. The principle of least action

The most general principle governing the motion of Lagrangian systems is the *principle of least action in the configuration space* (or *Hamilton's principle*), formulated as follows.

Let

$$P(M)_{q_0,t_0}^{q_1,t_1} = \{\gamma : [t_0, t_1] \to M;\ \gamma(t_0) = q_0,\ \gamma(t_1) = q_1\}$$

be the space of smooth parametrized paths in M connecting points q_0 and q_1. The path space $P(M) = P(M)_{q_0,t_0}^{q_1,t_1}$ is an infinite-dimensional real Fréchet manifold, and the tangent space $T_\gamma P(M)$ to $P(M)$ at $\gamma \in P(M)$ consists of all smooth vector fields along the path γ in M which vanish at the endpoints q_0 and q_1. A smooth path Γ in $P(M)$, passing through $\gamma \in P(M)$, is called a *variation with fixed ends* of the path $\gamma(t)$ in M. A variation Γ is a family $\gamma_\varepsilon(t) = \Gamma(t, \varepsilon)$ of paths in M given by a smooth map

$$\Gamma : [t_0, t_1] \times [-\varepsilon_0, \varepsilon_0] \to M$$

[1] It follows from the Newton-Laplace principle that L could depend only on generalized coordinates and velocities, and on time.

such that $\Gamma(t,0) = \gamma(t)$ for $t_0 \leq t \leq t_1$ and $\Gamma(t_0,\varepsilon) = q_0, \Gamma(t_1,\varepsilon) = q_1$ for $-\varepsilon_0 \leq \varepsilon \leq \varepsilon_0$. The tangent vector

$$\delta\gamma = \left.\frac{\partial\Gamma}{\partial\varepsilon}\right|_{\varepsilon=0} \in T_\gamma P(M)$$

corresponding to a variation $\gamma_\varepsilon(t)$ is traditionally called an *infinitesimal variation*. Explicitly,

$$\delta\gamma(t) = \Gamma_*(\tfrac{\partial}{\partial\varepsilon})(t,0) \in T_{\gamma(t)}M, \quad t_0 \leq t \leq t_1,$$

where $\frac{\partial}{\partial\varepsilon}$ is a tangent vector to the interval $[-\varepsilon_0,\varepsilon_0]$ at 0. Finally, a *tangential lift* of a path $\gamma : [t_0,t_1] \to M$ is the path $\gamma' : [t_0,t_1] \to TM$ defined by $\gamma'(t) = \gamma_*(\frac{\partial}{\partial t}) \in T_{\gamma(t)}M$, $t_0 \leq t \leq t_1$, where $\frac{\partial}{\partial t}$ is a tangent vector to $[t_0,t_1]$ at t. In other words, $\gamma'(t)$ is the velocity vector of a path $\gamma(t)$ at time t.

Definition. The *action functional* $S : P(M) \to \mathbb{R}$ of a Lagrangian system (M,L) is defined by

$$S(\gamma) = \int_{t_0}^{t_1} L(\gamma'(t),t)dt.$$

Principle of Least Action (Hamilton's principle). A path $\gamma \in PM$ describes the motion of a Lagrangian system (M,L) between the position $q_0 \in M$ at time t_0 and the position $q_1 \in M$ at time t_1 if and only if it is a critical point of the action functional S,

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

for all variations $\gamma_\varepsilon(t)$ of $\gamma(t)$ with fixed ends.

The critical points of the action functional are called *extremals* and the principle of the least action states that a Lagrangian system (M,L) moves along the extremals[2]. The extremals are characterized by equations of motion — a system of second order differential equations in local coordinates on TM. The equations of motion have the most elegant form for the following choice of local coordinates on TM.

Definition. Let (U,φ) be a coordinate chart on M with local coordinates $\boldsymbol{q} = (q^1,\dots,q^n)$. Coordinates

$$(\boldsymbol{q},\boldsymbol{v}) = (q^1,\dots,q^n,v^1,\dots,v^n)$$

on a chart TU on TM, where $\boldsymbol{v} = (v^1,\dots,v^n)$ are coordinates in the fiber corresponding to the basis $\dfrac{\partial}{\partial q^1},\dots,\dfrac{\partial}{\partial q^n}$ for T_qM, are called *standard coordinates*.

[2] The principle of least action does not state that an extremal connecting points q_0 and q_1 is a minimum of S, nor that such an extremal is unique. It also does not state that any two points can be connected by an extremal.

Standard coordinates are Cartesian coordinates on $\varphi_*(TU) \subset T\mathbb{R}^n \simeq \mathbb{R}^n \times \mathbb{R}^n$ and have the property that for $(q,v) \in TU$ and $f \in C^\infty(U)$,

$$v(f) = \sum_{i=1}^n v^i \frac{\partial f}{\partial q^i} = \boldsymbol{v}\frac{\partial f}{\partial \boldsymbol{q}}. \tag{1.3}$$

Let (U,φ) and (U',φ') be coordinate charts on M with the transition functions $F = (F^1,\dots,F^n) = \varphi' \circ \varphi^{-1} : \varphi(U \cap U') \to \varphi'(U \cap U')$, and let $(\boldsymbol{q},\boldsymbol{v})$ and $(\boldsymbol{q}',\boldsymbol{v}')$, respectively, be the standard coordinates on TU and TU'. We have $\boldsymbol{q}' = F(\boldsymbol{q})$, and it follows from (1.3) that

$$\boldsymbol{v}' = F_*(\boldsymbol{q})\boldsymbol{v}, \quad \text{where} \quad F_*(\boldsymbol{q}) = \left\{\frac{\partial F^i}{\partial q^j}(\boldsymbol{q})\right\}_{i,j=1}^n \tag{1.4}$$

is a matrix-valued function on $\varphi(U \cap U')$. In other words, "vertical" coordinates $\boldsymbol{v} = (v^1,\dots,v^n)$ in the fibers of $TM \to M$ transform like components of a tangent vector on M under the change of coordinates on M. In classical terminology, $\boldsymbol{v}$ is a contravariant vector.

The tangential lift $\gamma'(t)$ of a path $\gamma(t)$ in M in standard coordinates on TU is $(\boldsymbol{q}(t),\dot{\boldsymbol{q}}(t)) = (q^1(t),\dots,q^n(t),\dot q^1(t),\dots,\dot q^n(t))$, where the dot stands for the time derivative, so that

$$L(\gamma'(t),t) = L(\boldsymbol{q}(t),\dot{\boldsymbol{q}}(t),t).$$

Following a centuries long tradition[3], we will usually denote standard coordinates by

$$(\boldsymbol{q},\dot{\boldsymbol{q}}) = (q^1,\dots,q^n,\dot q^1,\dots,\dot q^n),$$

where the dot *does not* stand for the time derivative. Since we only consider paths in TM that are tangential lifts of paths in M, there will be no confusion[4].

Theorem 1.1. *The equations of motion of a Lagrangian system (M,L) in standard coordinates on TM are given by the Euler-Lagrange equations*

$$\frac{\partial L}{\partial \boldsymbol{q}}(\boldsymbol{q}(t),\dot{\boldsymbol{q}}(t),t) - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\boldsymbol{q}(t),\dot{\boldsymbol{q}}(t),t)\right) = 0.$$

Proof. Suppose first that an extremal $\gamma(t)$ lies in a coordinate chart U of M. Then a simple computation in standard coordinates, using integration by

[3] Used in all texts on classical mechanics and theoretical physics.

[4] We reserve the notation $(\boldsymbol{q}(t),\boldsymbol{v}(t))$ for general paths in TM.

parts, gives

$$\begin{aligned}
0 &= \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} S(\gamma_\varepsilon) \\
&= \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} \int_{t_0}^{t_1} L\left(\boldsymbol{q}(t,\varepsilon), \dot{\boldsymbol{q}}(t,\varepsilon), t\right) dt \\
&= \sum_{i=1}^{n} \int_{t_0}^{t_1} \left(\frac{\partial L}{\partial q^i}\delta q^i + \frac{\partial L}{\partial \dot{q}^i}\delta \dot{q}^i\right) dt \\
&= \sum_{i=1}^{n} \int_{t_0}^{t_1} \left(\frac{\partial L}{\partial q^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}\right) \delta q^i dt + \sum_{i=1}^{n} \frac{\partial L}{\partial \dot{q}^i}\delta q^i\bigg|_{t_0}^{t_1}.
\end{aligned}$$

The second sum in the last line vanishes due to the property $\delta q^i(t_0) = \delta q^i(t_1) = 0,\ i = 1, \dots, n$. The first sum is zero for arbitrary smooth functions δq^i on the interval $[t_0, t_1]$ which vanish at the endpoints. This implies that for each term in the sum the integrand is identically zero,

$$\frac{\partial L}{\partial q^i}(\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t), t) - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}(\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t), t)\right) = 0, \quad i = 1, \dots, n.$$

Since the restriction of an extremal of the action functional S to a coordinate chart on M is again an extremal, each extremal in standard coordinates on TM satisfies Euler-Lagrange equations. □

Remark. In calculus of variations, the directional derivative of a functional S with respect to a tangent vector $V \in T_\gamma P(M)$ — the *Gato derivative* — is defined by

$$\delta_V S = \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} S(\gamma_\varepsilon),$$

where γ_ε is a path in $P(M)$ with a tangent vector V at $\gamma_0 = \gamma$. The result of the above computation (when γ lies in a coordinate chart $U \subset M$) can be written as

$$\begin{aligned}
\delta_V S &= \int_{t_0}^{t_1} \sum_{i=1}^{n} \left(\frac{\partial L}{\partial q^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}\right)(\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t), t) v^i(t) dt \\
&= \int_{t_0}^{t_1} \left(\frac{\partial L}{\partial \boldsymbol{q}} - \frac{d}{dt}\frac{\partial L}{\partial \dot{\boldsymbol{q}}}\right)(\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t), t) \boldsymbol{v}(t) dt. \qquad (1.5)
\end{aligned}$$

Here $V(t) = \sum_{i=1}^{n} v^i(t)\frac{\partial}{\partial q^i}$ is a vector field along the path γ in M. Formula (1.5) is called the formula for the *first variation of the action with fixed ends*. The principle of least action is a statement that $\delta_V S(\gamma) = 0$ for all $V \in T_\gamma P(M)$.

Remark. It is also convenient to consider a space $\widehat{P(M)} = \{\gamma : [t_0, t_1] \to M\}$ of all smooth parametrized paths in M. The tangent space $T_\gamma \widehat{P(M)}$ to

$\widehat{P(M)}$ at $\gamma \in \widehat{P(M)}$ is the space of all smooth vector fields along the path γ in M (no condition at the endpoints). The computation in the proof of Theorem 1.1 yields the following formula for the *first variation of the action with free ends*:

$$\delta_V S = \int_{t_0}^{t_1} \left(\frac{\partial L}{\partial \boldsymbol{q}} - \frac{d}{dt} \frac{\partial L}{\partial \dot{\boldsymbol{q}}} \right) \boldsymbol{v}\, dt + \frac{\partial L}{\partial \dot{\boldsymbol{q}}} \boldsymbol{v} \Big|_{t_0}^{t_1}. \tag{1.6}$$

In expanded form, the Euler-Lagrange equations are given by the following system of second order ordinary differential equations:

$$\frac{\partial L}{\partial q^i}(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) \right)$$

$$= \sum_{j=1}^{n} \left(\frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j}(\boldsymbol{q}, \dot{\boldsymbol{q}}, t)\, \ddot{q}^j + \frac{\partial^2 L}{\partial \dot{q}^i \partial q^j}(\boldsymbol{q}, \dot{\boldsymbol{q}}, t)\, \dot{q}^j \right) + \frac{\partial^2 L}{\partial \dot{q}^i \partial t}(\boldsymbol{q}, \dot{\boldsymbol{q}}, t), \quad i = 1, \dots, n.$$

In order for this system to be solvable for the highest derivatives for all initial conditions in TU, the symmetric $n \times n$ matrix

$$H_L(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) = \left\{ \frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j}(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) \right\}_{i,j=1}^{n}$$

should be invertible on TU.

Definition. A Lagrangian system (M, L) is called *non-degenerate* if for every coordinate chart U on M the matrix $H_L(\boldsymbol{q}, \dot{\boldsymbol{q}}, t)$ is invertible on TU. Otherwise Lagrangian system is called *singular*.

Remark. Note that the $n \times n$ matrix H_L is a Hessian of the Lagrangian function L for vertical directions on TM. Under the change of standard coordinates $\boldsymbol{q}' = F(\boldsymbol{q})$ and $\dot{\boldsymbol{q}}' = F_*(\boldsymbol{q})\dot{\boldsymbol{q}}$ it has the transformation law

$$H_L(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) = F_*(\boldsymbol{q})^\tau H_L(\boldsymbol{q}', \dot{\boldsymbol{q}}', t) F_*(\boldsymbol{q}),$$

where $F_*(\boldsymbol{q})^\tau$ is the transposed matrix, so that the condition $\det H_L \neq 0$ does not depend on the choice of standard coordinates.

Inverting the matrix H_L, we can write Euler-Lagrange equations for a non-degenerate Lagrangian in the form

$$\ddot{q}^i = F^i(\boldsymbol{q}, \dot{\boldsymbol{q}}, t), \quad i = 1, \dots, n. \tag{1.7}$$

1.3. Newtonian spacetime

To describe a mechanical phenomena it is necessary to choose a *frame of reference*. The properties of the *spacetime* where the motion takes place depend on this choice. The spacetime is characterized by the following postulates[5].

[5]Strictly speaking, these postulates are valid only in the non-relativistic limit of special relativity, when the speed of light in the vacuum is assumed to be infinite.

Newtonian Space-Time. The space is a three-dimensional affine Euclidean space E^3. A choice of the *origin* $0 \in E^3$ — a *reference point* — establishes the isomorphism $E^3 \simeq \mathbb{R}^3$, where the vector space $\mathbb{R}^3$ carries the Euclidean inner product and has a fixed orientation. The time is one-dimensional — a time axis $\mathbb{R}$ — and the spacetime is a direct product $E^3 \times \mathbb{R}$. Points in the spacetime are called *events*. Two events $(\boldsymbol{r}, t)$ and $(\boldsymbol{r}', t')$ are called *simultaneous*, if $t = t'$. The distance can be defined only for simultaneous events and is the Euclidean distance $|\boldsymbol{r} - \boldsymbol{r}'|$.

An *inertial* reference frame is a coordinate system with respect to the origin $0 \in E^3$, initial time t_0, and an orthonormal basis in $\mathbb{R}^3$. In an inertial frame the space is *homogeneous* and *isotropic* and the time is *homogeneous*. The laws of motion are invariant with respect to the transformations

$$\boldsymbol{r} \mapsto g \cdot \boldsymbol{r} + \boldsymbol{r}_0, \quad t \mapsto t + t_0,$$

where $\boldsymbol{r}, \boldsymbol{r}_0 \in \mathbb{R}^3$ and $g \in \mathrm{O}(3)$ is an orthogonal linear transformation in $\mathbb{R}^3$. The time in classical mechanics is *absolute*.

The Galilean group G is a group of all affine transformations of $E^3 \times \mathbb{R}$, which preserve time intervals, and which for every $t \in \mathbb{R}$ are isometries in E^3. Every Galilean transformation is a composition of rotation, spacetime translation, and a special Galilean transformation

$$\boldsymbol{r} \mapsto \boldsymbol{r} + \boldsymbol{v}t, \quad t \mapsto t, \tag{1.8}$$

where $\boldsymbol{v} \in \mathbb{R}^3$. Any two inertial frames are related by a Galilean transformation.

The homogeneous Galilean group G_0 consists of rotations and special Galilean transformations (1.8). As a Lie group, G_0 is isomorphic to the Euclidean Lie group $E(3)$ — a semi-direct product $\mathbb{R}^3 \rtimes \mathrm{O}(3)$. Explicitly,

$$G_0 = \left\{ \begin{pmatrix} g & \boldsymbol{v} \\ 0 & 1 \end{pmatrix} : g \in \mathrm{O}(3),\ \boldsymbol{v} \in \mathbb{R}^3 \right\},$$

so that

$$\begin{pmatrix} \boldsymbol{r} \\ t \end{pmatrix} \mapsto \begin{pmatrix} g & \boldsymbol{v} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{r} \\ t \end{pmatrix} = \begin{pmatrix} g \cdot \boldsymbol{r} + \boldsymbol{v}t \\ t \end{pmatrix}.$$

Galileo's Relativity Principle. The laws of motion are invariant with respect to the Galilean group.

These postulates impose restrictions on Lagrangians of mechanical systems. In particular, Lagrangian L of a *closed system*[6] does not explicitly depend on time.

1.4. Examples of Lagrangian systems

Physical systems are described by special Lagrangians, in agreement with the experimental facts about the motion of material bodies.

[6] A system is called closed if its particles do not interact with the outside material bodies.

EXAMPLE 1.1 (Free particle). The configuration space for a free particle is $M = \mathbb{R}^3$, and it can be deduced from Galileo's relativity principle that the Lagrangian for a free particle is

$$L = \tfrac{1}{2}m\dot{\boldsymbol{r}}^2.$$

Here $m > 0$[7] is the mass of a particle and $\dot{\boldsymbol{r}}^2 = |\dot{\boldsymbol{r}}|^2$ is the length square of the velocity vector $\dot{\boldsymbol{r}} \in T_{\boldsymbol{r}}\mathbb{R}^3 \simeq \mathbb{R}^3$. Indeed, under the Galilean transformation (1.8)

$$L = \tfrac{1}{2}m\dot{\boldsymbol{r}}^2 \mapsto L' = L = \tfrac{1}{2}m(\dot{\boldsymbol{r}} + \boldsymbol{v})^2 = L + \frac{d}{dt}(m\boldsymbol{r}\boldsymbol{v} + \tfrac{1}{2}\boldsymbol{v}^2 t), \tag{1.9}$$

so that Lagrangians L and L' have the same equations of motion (cf. Problem 1.2). Specifically, Euler-Lagrange equations give *Newton's law of inertia*,

$$\ddot{\boldsymbol{r}} = 0.$$

EXAMPLE 1.2 (Interacting particles). A closed system of N interacting particles in $\mathbb{R}^3$ with masses $m_1, \dots, m_N$ is described by a configuration space

$$M = \mathbb{R}^{3N} = \underbrace{\mathbb{R}^3 \times \cdots \times \mathbb{R}^3}_{N}$$

with a position vector $\boldsymbol{r} = (\boldsymbol{r}_1, \dots, \boldsymbol{r}_N)$, where $\boldsymbol{r}_a \in \mathbb{R}^3$ is a position vector of the a-th particle, $a = 1, \dots, N$. It is found that the Lagrangian is given by

$$L = \sum_{a=1}^{N} \tfrac{1}{2}m_a\dot{\boldsymbol{r}}_a^2 - V(\boldsymbol{r}) = T - V,$$

where

$$T = \sum_{a=1}^{N} \tfrac{1}{2}m_a\dot{\boldsymbol{r}}_a^2$$

is called *kinetic energy* of a system and $V(\boldsymbol{r})$ is *potential energy*. The Euler-Lagrange equations give *Newton's equations*

$$m_a\ddot{\boldsymbol{r}}_a = \boldsymbol{F}_a,$$

where

$$\boldsymbol{F}_a = -\frac{\partial V}{\partial \boldsymbol{r}_a}$$

is a *force* on the a-th particle, $a = 1, \dots, N$. Forces of this form are called *conservative*. Thus the interaction of particles is through the action of potential forces, and is an *instantaneous action at a distance*[8].

---

[7] Otherwise the action functional is not bounded from below.

[8] This means a phenomenon in which a change in intrinsic properties of one system induces an instantaneous change in the intrinsic properties of a distant system without a process that carries this influence contiguously in space and time.

It follows from homogeneity of space that potential energy $V(\boldsymbol{r})$ of a closed system of N interacting particles with conservative forces depends only on relative positions of the particles, i.e., $V(\boldsymbol{r}_1+\boldsymbol{c},\dots,\boldsymbol{r}_N+\boldsymbol{c})=V(\boldsymbol{r}_1,\dots,\boldsymbol{r}_N)$ for all $\boldsymbol{c}\in\mathbb{R}^3$, which leads to the equation

$$\sum_{a=1}^{N}\boldsymbol{F}_a=0.$$

In particular, for a closed system of two particles $\boldsymbol{F}_1+\boldsymbol{F}_2=0$, which is the equality of action and reaction forces, also called *Newton's third law.*

The potential energy of a closed system with only pair-wise interaction between the particles has the form

$$V(\boldsymbol{r})=\sum_{1\leq a<b\leq N}V_{ab}(\boldsymbol{r}_a-\boldsymbol{r}_b).$$

It follows from the isotropy of space that $V(\boldsymbol{r})$ depends only on relative distances between the particles, so that the Lagrangian of a closed system of N particles with pair-wise interaction has the form

$$L=\sum_{a=1}^{N}\tfrac{1}{2}m_a\dot{\boldsymbol{r}}_a^2\;-\sum_{1\leq a<b\leq N}V_{ab}(|\boldsymbol{r}_a-\boldsymbol{r}_b|).$$

EXAMPLE 1.3 (Universal gravitation). According to *Newton's law of gravitation*, the potential energy of the gravitational force between two particles with masses m_a and m_b is

$$V(\boldsymbol{r}_a-\boldsymbol{r}_b)=-G\frac{m_am_b}{|\boldsymbol{r}_a-\boldsymbol{r}_b|},$$

where G is the gravitational constant. The configuration space of N particles with gravitational interaction is

$$M=\{(\boldsymbol{r}_1,\dots,\boldsymbol{r}_N)\in\mathbb{R}^{3N}:\boldsymbol{r}_a\neq\boldsymbol{r}_b\text{ for }a\neq b,\,a,b=1,\dots,N\}.$$

EXAMPLE 1.4 (Small oscillations). Consider a particle of mass m with n degrees of freedom moving in a potential field $V(\boldsymbol{q})$, and suppose that potential energy V has a minimum at $\boldsymbol{q}=0$. Expanding $V(\boldsymbol{q})$ in Taylor series around 0 and keeping only quadratic terms, one obtains a Lagrangian system which describes small oscillations from equilibrium. Explicitly,

$$L=\tfrac{1}{2}m\dot{\boldsymbol{q}}^2-V_0(\boldsymbol{q}),$$

where V_0 is a positive-definite quadratic form on $\mathbb{R}^n$ given by

$$V_0(\boldsymbol{q})=\tfrac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2V}{\partial q^i\partial q^j}(0)q^iq^j.$$

Since every quadratic form can be diagonalized by an orthogonal transformation, we can assume from the very beginning that coordinates $\boldsymbol{q} = (q^1, \dots, q^n)$ are chosen so that $V_0(\boldsymbol{q})$ is diagonal and

$$L = \tfrac{1}{2}m\big(\dot{\boldsymbol{q}}^2 - \sum_{i=1}^{n} \omega_i^2 (q^i)^2\big), \tag{1.10}$$

where $\omega_1, \dots, \omega_n > 0$. Such coordinates $\boldsymbol{q}$ are called *normal coordinates*. In normal coordinates Euler-Lagrange equations take the form

$$\ddot{q}^i + \omega_i^2 q^i = 0, \quad i = 1, \dots, n,$$

and describe n decoupled (i.e., non-interacting) *harmonic oscillators* with *frequencies* $\omega_1, \dots, \omega_n$.

Example 1.5 (Free particle on a Riemannian manifold). Let (M, ds^2) be a Riemannian manifold with the Riemannian metric ds^2. In local coordinates $x^1, \dots, x^n$ on M,

$$ds^2 = g_{\mu\nu}(x) dx^\mu dx^\nu,$$

where we are using summation over repeated indices. The Lagrangian of a free particle on M is

$$L(v) = \tfrac{1}{2}\langle v, v\rangle = \tfrac{1}{2}\|v\|^2,\ v \in TM,$$

where $\langle\ ,\ \rangle$ stands for the inner product in fibers of TM, given by the Riemannian metric. The corresponding functional

$$S(\gamma) = \tfrac{1}{2}\int_{t_0}^{t_1} \|\gamma'(t)\|^2 dt = \tfrac{1}{2}\int_{t_0}^{t_1} g_{\mu\nu}(x)\dot{x}^\mu \dot{x}^\nu dt$$

is called the action functional in Riemannian geometry. The Euler-Lagrange equations are

$$g_{\mu\nu}\ddot{x}^\mu + \frac{\partial g_{\mu\nu}}{\partial x^\lambda}\dot{x}^\mu \dot{x}^\lambda = \frac{1}{2}\frac{\partial g_{\mu\lambda}}{\partial x^\nu}\dot{x}^\mu \dot{x}^\lambda,$$

and after multiplying by the inverse metric tensor $g^{\sigma\nu}$ and summation over ν they take the form

$$\ddot{x}^\sigma + \Gamma^\sigma_{\mu\nu}\dot{x}^\mu \dot{x}^\nu = 0, \quad \sigma = 1, \dots, n,$$

where

$$\Gamma^\sigma_{\mu\nu} = \frac{1}{2} g^{\sigma\lambda}\left(\frac{\partial g_{\mu\lambda}}{\partial x^\nu} + \frac{\partial g_{\nu\lambda}}{\partial x^\mu} - \frac{\partial g_{\mu\nu}}{\partial x^\lambda}\right)$$

are Christoffel's symbols. The Euler-Lagrange equations of a free particle moving on a Riemannian manifold are geodesic equations.

Let ∇ be the Levi-Civita connection — the metric connection in the tangent bundle TM — and let ∇_ξ be a covariant derivative with respect to the vector field $\xi \in \mathrm{Vect}(M)$. Explicitly,

$$(\nabla_\xi\, \eta)^\mu = \left(\frac{\partial \eta^\mu}{\partial x^\nu} + \Gamma^\mu_{\nu\lambda}\eta^\lambda\right)\xi^\nu, \quad \text{where} \quad \xi = \xi^\mu(x)\frac{\partial}{\partial x^\mu},\ \eta = \eta^\mu(x)\frac{\partial}{\partial x^\mu}.$$

For a path $\gamma(t) = (x^\mu(t))$ denote by $\nabla_{\dot\gamma}$ a covariant derivative along γ,

$$(\nabla_{\dot\gamma}\eta)^\mu(t) = \frac{d\eta^\mu(t)}{dt} + \Gamma^\mu_{\nu\lambda}(\gamma(t))\dot x^\nu(t)\eta^\lambda(t), \quad \text{where} \quad \eta = \eta^\mu(t)\frac{\partial}{\partial x^\mu}$$

is a vector field along γ. Formula (1.5) can now be written in an invariant form

$$\delta S = -\int_{t_0}^{t_1} \langle \nabla_{\dot\gamma}\dot\gamma, \delta\gamma\rangle dt,$$

which is known as the formula for the first variation of the action in Riemannian geometry.

PROBLEM 1.1. Show that the action functional is given by the evaluation of the 1-form Ldt on $TM \times \mathbb{R}$ over the 1-chain $\tilde\gamma$ on $TM \times \mathbb{R}$,

$$S(\gamma) = \int_{\tilde\gamma} Ldt,$$

where $\tilde\gamma = \{(\gamma'(t), t); t_0 \le t \le t_1\}$ and $Ldt\left(w, c\frac{\partial}{\partial t}\right) = cL(q, v)$, $w \in T_{(q,v)}TM$, $c \in \mathbb{R}$.

PROBLEM 1.2. Let $f \in C^\infty(M)$. Show that Lagrangian systems (M, L) and $(M, L + df)$ (where df is a fibre-wise linear function on TM) have the same equations of motion. In general, the Lagrangian is defined up to an addition of a total time derivative of a function of coordinates and time.

PROBLEM 1.3. Give examples of Lagrangian systems such that an extremal connecting two given points (i) is not a local minimum; (ii) is not unique; (iii) does not exist.

PROBLEM 1.4. For γ an extremal of the action functional S, the *second variation* of S is defined by

$$\delta^2_{V_1V_2}S = \left.\frac{\partial^2}{\partial\varepsilon_1\partial\varepsilon_2}\right|_{\varepsilon_1=\varepsilon_2=0} S(\gamma_{\varepsilon_1,\varepsilon_2}),$$

where $\gamma_{\varepsilon_1,\varepsilon_2}$ is a smooth two-parameter family of paths in M such that the paths $\gamma_{\varepsilon_1,0}$ and γ_{0,ε_2} in $P(M)$ at the point $\gamma_{0,0} = \gamma \in P(M)$ have tangent vectors V_1 and V_2, respectively. For a Lagrangian system (M, L) find the second variation of S and verify that for given V_1 and V_2 it does not depend on the choice of $\gamma_{\varepsilon_1,\varepsilon_2}$.

PROBLEM 1.5. Prove that the second variation of the action functional in Riemannian geometry is given by

$$\delta^2 S = \int_{t_0}^{t_1} \langle \mathcal{J}(\delta_1\gamma), \delta_2\gamma\rangle dt.$$

Here $\delta_1\gamma, \delta_2\gamma \in T_\gamma PM$, $\mathcal{J} = -\nabla^2_{\dot\gamma} - R(\dot\gamma, \cdot\,)\dot\gamma$ is the Jacobi operator, and R is a curvature operator — a fibre-wise linear mapping $R : TM \otimes TM \to \mathrm{End}(TM)$ of vector bundles, defined by $R(\xi, \eta) = \nabla_\eta\nabla_\xi - \nabla_\xi\nabla_\eta + \nabla_{[\xi,\eta]} : TM \to TM$, where $\xi, \eta \in \mathrm{Vect}(M)$.

LECTURE 2

Integrals of motion and Noether's theorem

To describe the motion of a mechanical system one needs to solve the Euler-Lagrange equations — a system of second order ordinary differential equations for the generalized coordinates. This could be a very difficult problem. Therefore of particular interest are those functions of generalized coordinates and velocities, which remain constant during the motion.

Definition. A smooth function $I : TM \to \mathbb{R}$ is called an *integral of motion* (*first integral*, or *conservation law*) for a Lagrangian system (M, L) if

$$\frac{d}{dt} I(\gamma'(t)) = 0$$

for all extremals γ of the action functional.

2.1. Conservation of energy

Definition. The *energy* of a Lagrangian system (M, L) is a function E on $TM \times \mathbb{R}$, defined in standard coordinates on TM by

$$E(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) = \sum_{i=1}^{n} \dot{q}^i \frac{\partial L}{\partial \dot{q}^i}(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) - L(\boldsymbol{q}, \dot{\boldsymbol{q}}, t).$$

Lemma 2.1. *The energy* $E = \dot{\boldsymbol{q}} \dfrac{\partial L}{\partial \dot{\boldsymbol{q}}} - L$ *is a well-defined function on* $TM \times \mathbb{R}$.

Proof. Let (U, φ) and (U', φ') be coordinate charts on M with the transition functions $F = (F^1, \dots, F^n) = \varphi' \circ \varphi^{-1} : \varphi(U \cap U') \to \varphi'(U \cap U')$. Corresponding standard coordinates $(\boldsymbol{q}, \dot{\boldsymbol{q}})$ and $(\boldsymbol{q}', \dot{\boldsymbol{q}}')$ are related by $\boldsymbol{q}' = F(\boldsymbol{q})$ and $\dot{\boldsymbol{q}}' = F_*(\boldsymbol{q})\dot{\boldsymbol{q}}$ (see formula (1.3) in Lecture 1). We have

$$d\boldsymbol{q}' = F_*(\boldsymbol{q}) d\boldsymbol{q} \quad \text{and} \quad d\dot{\boldsymbol{q}}' = G(\boldsymbol{q}, \dot{\boldsymbol{q}}) d\boldsymbol{q} + F_*(\boldsymbol{q}) d\dot{\boldsymbol{q}},$$

where

$$G(\boldsymbol{q}, \dot{\boldsymbol{q}}) = \left\{ \sum_{k=1}^{n} \frac{\partial^2 F^i}{\partial q^j \partial q^k} \dot{q}^k \right\}_{i,j=1}^{n},$$

so that

$$\begin{aligned} dL &= \frac{\partial L}{\partial \boldsymbol{q}'} d\boldsymbol{q}' + \frac{\partial L}{\partial \dot{\boldsymbol{q}}'} d\dot{\boldsymbol{q}}' + \frac{\partial L}{\partial t} dt \\ &= \left(\frac{\partial L}{\partial \boldsymbol{q}'} F_*(\boldsymbol{q}) + \frac{\partial L}{\partial \dot{\boldsymbol{q}}'} G(\boldsymbol{q}, \dot{\boldsymbol{q}}) \right) d\boldsymbol{q} + \frac{\partial L}{\partial \dot{\boldsymbol{q}}'} F_*(\boldsymbol{q}) d\dot{\boldsymbol{q}} + \frac{\partial L}{\partial t} dt \\ &= \frac{\partial L}{\partial \boldsymbol{q}} d\boldsymbol{q} + \frac{\partial L}{\partial \dot{\boldsymbol{q}}} d\dot{\boldsymbol{q}} + \frac{\partial L}{\partial t} dt. \end{aligned}$$

Thus under a change of coordinates

$$\frac{\partial L}{\partial \dot{\boldsymbol{q}}'} F_*(\boldsymbol{q}) = \frac{\partial L}{\partial \dot{\boldsymbol{q}}} \quad \text{and} \quad \dot{\boldsymbol{q}}' \frac{\partial L}{\partial \dot{\boldsymbol{q}}'} = \dot{\boldsymbol{q}} \frac{\partial L}{\partial \dot{\boldsymbol{q}}},$$

so that E is a well-defined function on TM. □

Corollary 2.1. *Under a change of local coordinates on M, components of a vector $\frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) = \left(\frac{\partial L}{\partial \dot{q}^1}, \dots, \frac{\partial L}{\partial \dot{q}^n} \right)$ transform like components of a 1-form on M. In classical terminology, $\frac{\partial L}{\partial \dot{\boldsymbol{q}}}$ is a covariant vector.*

Let θ_L be a 1-form on TM, defined in standard coordinates associated with a coordinate chart $U \subset M$ by

$$\theta_L = \sum_{i=1}^{n} \frac{\partial L}{\partial \dot{q}^i} dq^i = \frac{\partial L}{\partial \dot{\boldsymbol{q}}} d\boldsymbol{q}. \tag{2.1}$$

It follows from Corollary 2.1 that θ_L is a well-defined 1-form on TM.

Proposition 2.1 (Conservation of energy). *The energy of a closed system is an integral of motion.*

Proof. For an extremal γ put $E(t) = E(\gamma'(t))$. We have, according to the Euler-Lagrange equations,

$$\begin{aligned} \frac{dE}{dt} &= \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\boldsymbol{q}}} \right) \dot{\boldsymbol{q}} + \frac{\partial L}{\partial \dot{\boldsymbol{q}}} \ddot{\boldsymbol{q}} - \frac{\partial L}{\partial \boldsymbol{q}} \dot{\boldsymbol{q}} - \frac{\partial L}{\partial \dot{\boldsymbol{q}}} \ddot{\boldsymbol{q}} - \frac{\partial L}{\partial t} \\ &= \left(\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\boldsymbol{q}}} \right) - \frac{\partial L}{\partial \boldsymbol{q}} \right) \dot{\boldsymbol{q}} - \frac{\partial L}{\partial t} = -\frac{\partial L}{\partial t}. \end{aligned}$$

Since for a closed system $\frac{\partial L}{\partial t} = 0$, the energy is conserved. □

Conservation of energy for a closed mechanical system is a fundamental law of physics, which follows from the homogeneity of time. For a general closed system of N interacting particles considered in Example 1.2,

$$E = \sum_{a=1}^{N} m_a \dot{\boldsymbol{r}}_a^2 - L = \sum_{a=1}^{N} \tfrac{1}{2} m_a \dot{\boldsymbol{r}}_a^2 + V(\boldsymbol{r}).$$

In other words, the total energy $E = T + V$ is a sum of the kinetic energy and the potential energy.

2.2. Noether theorem

DEFINITION. A Lagrangian $L : TM \to \mathbb{R}$ is invariant with respect to the diffeomorphism $g : M \to M$ if $L(g_*(v)) = L(v)$ for all $v \in TM$. The diffeomorphism g is called a *symmetry* of a closed Lagrangian system (M, L). A Lie group G is the *symmetry group* of (M, L) (group of *continuous symmetries*), if there is a left G-action on M such that for every $g \in G$ the mapping $M \ni x \mapsto g \cdot x \in M$ is a symmetry.

Continuous symmetries give rise to conservation laws.

THEOREM 2.2 (Noether). *Suppose that a Lagrangian $L : TM \to \mathbb{R}$ is invariant under a one-parameter group $\{g_s\}_{s\in\mathbb{R}}$ of diffeomorphisms of M. Then the Lagrangian system (M, L) admits an integral of motion I, given in standard coordinates on TM by*

$$I(\boldsymbol{q}, \dot{\boldsymbol{q}}) = \sum_{i=1}^{n} \frac{\partial L}{\partial \dot{q}^i}(\boldsymbol{q}, \dot{\boldsymbol{q}}) \left(\left. \frac{dg_s^i(\boldsymbol{q})}{ds} \right|_{s=0} \right) = \frac{\partial L}{\partial \dot{\boldsymbol{q}}} \boldsymbol{a},$$

where $X = \sum_{i=1}^{n} a^i(\boldsymbol{q}) \frac{\partial}{\partial q^i}$ is the vector field on M associated with the flow g_s. The integral of motion I is called the Noether integral.

PROOF. It follows from Corollary 2.1 that I is a well-defined function on TM. Now differentiating $L((g_s)_*(\gamma'(t))) = L(\gamma'(t))$ with respect to s at $s = 0$ and using the Euler-Lagrange equations, we get

$$0 = \frac{\partial L}{\partial \boldsymbol{q}} \boldsymbol{a} + \frac{\partial L}{\partial \dot{\boldsymbol{q}}} \dot{\boldsymbol{a}} = \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\boldsymbol{q}}} \right) \boldsymbol{a} + \frac{\partial L}{\partial \dot{\boldsymbol{q}}} \frac{d\boldsymbol{a}}{dt} = \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\boldsymbol{q}}} \boldsymbol{a} \right),$$

where $\boldsymbol{a}(t) = \big(a^1(\gamma(t)), \dots, a^n(\gamma(t))\big)$. □

REMARK. A vector field X on M is called an *infinitesimal symmetry*, if the corresponding "time s" local flow g_s of X (defined for each $s \in \mathbb{R}$ on some $U_s \subseteq M$ as a diffeomorphism $g_s : U_s \to U_{-s}$) is a symmetry: $L \circ (g_s)_* = L$ on U_s. Every vector field X on M lifts to a vector field X' on TM, defined by a local flow on TM, induced from the corresponding local flow on M. In standard coordinates on TM,

$$X = \sum_{i=1}^{n} a^i(\boldsymbol{q}) \frac{\partial}{\partial q^i},$$

and the corresponding local flow on M is given by

$$\frac{dq^i}{ds} = a^i(\boldsymbol{q}),$$

and induces the local flow on TM,

$$\frac{d\dot{q}^i}{ds} = \sum_{j=1}^{n} \dot{q}^j \frac{\partial a^i}{\partial q^j}(\boldsymbol{q}), \quad i = 1, \dots, n.$$

Thus

$$X' = \sum_{i=1}^{n} a^i(\boldsymbol{q}) \frac{\partial}{\partial q^i} + \sum_{i,j=1}^{n} \dot{q}^j \frac{\partial a^i}{\partial q^j}(\boldsymbol{q}) \frac{\partial}{\partial \dot{q}^i}, \tag{2.2}$$

and for every path γ in M,

$$dL(X')(\gamma'(t)) = \frac{\partial L}{\partial \boldsymbol{q}}\boldsymbol{a} + \frac{\partial L}{\partial \dot{\boldsymbol{q}}}\dot{\boldsymbol{a}}.$$

It is easy to verify that X is an infinitesimal symmetry if and only if $dL(X') = 0$ on TM, and $I(\boldsymbol{q}, \dot{\boldsymbol{q}}) = \dfrac{\partial L}{\partial \dot{\boldsymbol{q}}}\boldsymbol{a}$ is an integral of motion.

Remark. Using the 1-form θ_L, the Noether integral I in Theorem 2.2 can be written as

$$I = \theta_L(X'). \tag{2.3}$$

Remark. Noether's theorem generalizes to time-dependent Lagrangians $L : TM \times \mathbb{R} \to \mathbb{R}$. Namely, on the *extended configuration space* $M_1 = M \times \mathbb{R}$ define a time-independent Lagrangian L_1 by

$$L_1(\boldsymbol{q}, \tau, \dot{\boldsymbol{q}}, \dot{\tau}) = L\left(\boldsymbol{q}, \frac{\dot{\boldsymbol{q}}}{\dot{\tau}}, \tau\right)\dot{\tau},$$

where $(\boldsymbol{q}, \tau)$ are local coordinates on M_1 and $(\boldsymbol{q}, \tau, \dot{\boldsymbol{q}}, \dot{\tau})$ are standard coordinates on TM_1. The Noether integral I_1 for a closed system (M_1, L_1) defines an integral of motion I for a system (M, L) by the formula

$$I(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) = I_1(\boldsymbol{q}, t, \dot{\boldsymbol{q}}, 1).$$

When the Lagrangian L does not depend on time, L_1 is invariant with respect to the one-parameter group of translations $\tau \mapsto \tau + s$, and the Noether integral $I_1 = \dfrac{\partial L_1}{\partial \dot{\tau}}$ gives $I = -E$.

Noether's theorem can be generalized further as follows.

Proposition 2.2. *Suppose that for a given Lagrangian $L : TM \to \mathbb{R}$ there exist a vector field X on M and a function K on TM, such that for every path γ in M,*

$$dL(X')(\gamma'(t)) = \frac{d}{dt}K(\gamma'(t)).$$

Then

$$I = \sum_{i=1}^{n} a^i(\boldsymbol{q}) \frac{\partial L}{\partial \dot{q}^i}(\boldsymbol{q}, \dot{\boldsymbol{q}}) - K(\boldsymbol{q}, \dot{\boldsymbol{q}})$$

is an integral of motion for the Lagrangian system (M, L).

PROOF. Using Euler-Lagrange equations, we have along the extremal γ,

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\boldsymbol{q}}}\boldsymbol{a}\right) = \frac{\partial L}{\partial \boldsymbol{q}}\boldsymbol{a} + \frac{\partial L}{\partial \dot{\boldsymbol{q}}}\dot{\boldsymbol{a}} = \frac{dK}{dt}. \qquad \square$$

For a closed, non-degenerate Lagrangian system (M, L) this result can be generalized further by allowing coefficients $a^i(\boldsymbol{q})$ of the vector field X to depend also on $\dot{\boldsymbol{q}}$. Namely, rewrite Euler-Lagrange equations as in (1.7), and consider a vector field $\tilde{X}$ on TM, given in the standard coordinates by

$$\tilde{X} = \sum_{i=1}^{n} a^i(\boldsymbol{q}, \dot{\boldsymbol{q}})\frac{\partial}{\partial q^i} + \sum_{i,j=1}^{n}\left(\dot{q}^j\,\frac{\partial a^i}{\partial q^j}(\boldsymbol{q}, \dot{\boldsymbol{q}}) + F^j(\boldsymbol{q}, \dot{\boldsymbol{q}})\,\frac{\partial a^i}{\partial \dot{q}^j}(\boldsymbol{q}, \dot{\boldsymbol{q}})\right)\frac{\partial}{\partial \dot{q}^i}. \tag{2.4}$$

PROPOSITION 2.3. *Suppose that for a closed, non-degenerate Lagrangian L there exist a vector field $\tilde{X}$ on TM of the form (2.4), and a function K on TM, such that for every path γ in M,*

$$dL(\tilde{X})(\gamma'(t)) = \frac{d}{dt}K(\gamma'(t)). \tag{2.5}$$

Then

$$I = \sum_{i=1}^{n} a^i(\boldsymbol{q}, \dot{\boldsymbol{q}})\,\frac{\partial L}{\partial \dot{q}^i}(\boldsymbol{q}, \dot{\boldsymbol{q}}) - K(\boldsymbol{q}, \dot{\boldsymbol{q}})$$

is an integral of motion.

PROOF. Along the extremal $\gamma(t)$,

$$\frac{dI}{dt} = \frac{\partial L}{\partial \boldsymbol{q}}\boldsymbol{a} + \frac{\partial L}{\partial \dot{\boldsymbol{q}}}\dot{\boldsymbol{a}} - \frac{dK}{dt} = dL(\tilde{X})(\gamma'(t)) - \frac{d}{dt}K(\gamma'(t)) = 0. \qquad \square$$

2.3. Examples of conservation laws

EXAMPLE 2.1 (Conservation of momentum). Let $M = V$ be a vector space, and suppose that a Lagrangian L is invariant with respect to a one-parameter group $g_s(q) = q + sv$, $v \in V$. According to Noether's theorem,

$$I = \sum_{i=1}^{n} v^i\,\frac{\partial L}{\partial \dot{q}^i}$$

is an integral of motion. Now let (M, L) be closed Lagrangian system of N interacting particles, considered in Example 1.2. We have $M = V = \mathbb{R}^{3N}$, and the Lagrangian L is invariant under a simultaneous translation of coordinates $\boldsymbol{r}_a = (r_a^1, r_a^2, r_a^3)$ of all particles by the same vector $\boldsymbol{c} \in \mathbb{R}^3$. Thus $v = (\boldsymbol{c}, \dots, \boldsymbol{c}) \in \mathbb{R}^{3N}$, and for every $\boldsymbol{c} = (c^1, c^2, c^3) \in \mathbb{R}^3$,

$$I = \sum_{a=1}^{N}\left(c^1\frac{\partial L}{\partial \dot{r}_a^1} + c^2\frac{\partial L}{\partial \dot{r}_a^2} + c^3\frac{\partial L}{\partial \dot{r}_a^3}\right) = c^1P_1 + c^2P_2 + c^3P_3$$

is an integral of motion. The integrals of motion P_1, P_2, P_3 define the vector

$$\boldsymbol{P} = \sum_{a=1}^{N} \frac{\partial L}{\partial \dot{\boldsymbol{r}}_a} \in \mathbb{R}^3$$

(or rather a vector in a dual space to $\mathbb{R}^3$), called the *momentum* of a system. Explicitly,

$$\boldsymbol{P} = \sum_{a=1}^{N} m_a \dot{\boldsymbol{r}}_a,$$

so that the total momentum of a closed system is a sum of momenta of individual particles. Conservation of momentum is a fundamental physical law which reflects the homogeneity of space.

Traditionally, $p_i = \dfrac{\partial L}{\partial \dot{q}^i}$ are called *generalized momenta* corresponding to generalized coordinates q^i, and $F_i = \dfrac{\partial L}{\partial q^i}$ are called *generalized forces*. In these notation, the Euler-Lagrange equations have the same form

$$\dot{\boldsymbol{p}} = \boldsymbol{F}$$

as Newton's equations in Cartesian coordinates. Conservation of momentum implies Newton's third law.

Example 2.2 (Conservation of angular momentum). Let $M = V$ be a vector space with Euclidean inner product. Let $G = \mathrm{SO}(V)$ be the connected Lie group of automorphisms of V preserving the inner product, and let $\mathfrak{g} = \mathfrak{so}(V)$ be the Lie algebra of G. Suppose that a Lagrangian L is invariant with respect to the action of a one-parameter subgroup $g_s(q) = e^{sx} \cdot q$ of G on V, where $x \in \mathfrak{g}$ and e^x is the exponential map. According to Noether's theorem,

$$I = \sum_{i=1}^{n} (x \cdot q)^i \, \frac{\partial L}{\partial \dot{q}^i}$$

is an integral of motion. Now let (M, L) be a closed Lagrangian system of N interacting particles, considered in Example 1.2. We have $M = V = \mathbb{R}^{3N}$, and the Lagrangian L is invariant under a simultaneous rotation of coordinates $\boldsymbol{r}_a$ of all particles by the same orthogonal transformation in $\mathbb{R}^3$. Thus $x = (u, \dots, u) \in \underbrace{\mathfrak{so}(3) \oplus \cdots \oplus \mathfrak{so}(3)}_{N}$, and for every $u \in \mathfrak{so}(3)$,

$$I = \sum_{a=1}^{N} \left((u \cdot \boldsymbol{r}_a)^1 \, \frac{\partial L}{\partial \dot{r}_a^1} + (u \cdot \boldsymbol{r}_a)^2 \, \frac{\partial L}{\partial \dot{r}_a^2} + (u \cdot \boldsymbol{r}_a)^3 \, \frac{\partial L}{\partial \dot{r}_a^3} \right)$$

is an integral of motion. Let $u = u^1 X_1 + u^2 X_2 + u^3 X_3$, where $X_1 = \left(\begin{smallmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{smallmatrix}\right)$, $X_2 = \left(\begin{smallmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{smallmatrix}\right)$, $X_3 = \left(\begin{smallmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{smallmatrix}\right)$ is the basis in $\mathfrak{so}(3) \simeq \mathbb{R}^3$ corresponding to the rotations about the vectors e_1, e_2, e_3 of the standard orthonormal basis in $\mathbb{R}^3$. Since

$u \cdot \boldsymbol{r}_a = \boldsymbol{u} \times \boldsymbol{r}_a$, where $\boldsymbol{u} = (u^1, u^2, u^3)$, we have

$$I = u^1 M_1 + u^2 M_2 + u^3 M_3,$$

where $\boldsymbol{M} = (M_1, M_2, M_3) \in \mathbb{R}^3$ (or rather a vector in a dual space to $\mathfrak{so}(3)$) is given by

$$\boldsymbol{M} = \sum_{a=1}^{N} \boldsymbol{r}_a \times \frac{\partial L}{\partial \dot{\boldsymbol{r}}_a}.$$

The vector $\boldsymbol{M}$ is called the *angular momentum* of a system. Explicitly,

$$\boldsymbol{M} = \sum_{a=1}^{N} \boldsymbol{r}_a \times m_a \dot{\boldsymbol{r}}_a,$$

so that the total angular momentum of a closed system is a sum of angular momenta of individual particles. Conservation of angular momentum is a fundamental physical law which reflects the isotropy of space.

Example 2.3 (The center of mass). Let (M, L) be a closed Lagrangian system of N interacting particles, considered in Example 1.2. Under a simultaneous Galilean transformation (1.8) of all coordinates, $\boldsymbol{r}_a \mapsto \boldsymbol{r}_a + \boldsymbol{v}t$, and corresponding transformation of velocities $\dot{\boldsymbol{r}}_a \mapsto \dot{\boldsymbol{r}}_a + \boldsymbol{v}$, we have

$$L \mapsto L' = L + \frac{d}{dt} \sum_{a=1}^{N} m_a (\boldsymbol{r}_a \boldsymbol{v} + \tfrac{1}{2} \boldsymbol{v}^2 t).$$

Therefore for infinitesimal Galilean transformation — the time-dependent vector field

$$\tilde{X} = \sum_{a=1}^{N} \left(t\boldsymbol{v} \frac{\partial}{\partial \boldsymbol{r}_a} + \boldsymbol{v} \frac{\partial}{\partial \dot{\boldsymbol{r}}_a} \right)$$

equation (2.5) holds, where the functions K is given by

$$K = \sum_{a=1}^{N} m_a \boldsymbol{r}_a \boldsymbol{v}.$$

According to Proposition 2.3, the vector

$$\boldsymbol{I} = t\boldsymbol{P} - \sum_{a=1}^{N} m_a \boldsymbol{r}_a$$

is an integral of motion, $\dot{\boldsymbol{I}} = 0$ on the solutions of the Euler-Lagrange equations. This is equivalent to the statement that the *center of mass* of the system

$$\boldsymbol{R} = \frac{1}{M} \sum_{a=1}^{N} m_a \boldsymbol{r}_a, \quad \text{where} \quad M = \sum_{a=1}^{N} m_a$$

is the total mass, moves with the constant velocity $\boldsymbol{V} = \boldsymbol{P}/M$.

PROBLEM 2.1. Prove that a Lagrangian system (M, L) is non-degenerate if and only if the 2-form $d\theta_L$ on TM is non-degenerate.

PROBLEM 2.2 (Second tangent bundle). Let $\pi : TM \to M$ be the canonical projection and let $T_{\mathrm{V}}(TM)$ be a *vertical tangent bundle* of TM along the fibers of π — the kernel of the bundle mapping $\pi_* : T(TM) \to TM$. Prove that there is a natural bundle isomorphism $i : \pi^*(TM) \simeq T_{\mathrm{V}}(TM)$, where $\pi^*(TM) \to TM$ is the pullback of the tangent bundle TM of M under the map π.

PROBLEM 2.3 (Invariant definition of the 1-form θ_L). Show that $\theta_L(v) = dL((i \circ \pi_*)v)$, where $v \in T(TM)$.

PROBLEM 2.4. Prove that if a vector field X on M is an infinitesimal symmetry of the Lagrangian system (M, L), then $\mathcal{L}_{X'}(\theta_L) = 0$, where $\mathcal{L}_{X'}$ stands for the Lie derivative.

PROBLEM 2.5. Prove that a path $\gamma(t)$ in M is a trajectory for the Lagrangian system (M, L) if and only if

$$i_{\dot{\gamma}'(t)}(d\theta_L) + dE_L(\gamma'(t)) = 0,$$

where $\dot{\gamma}'(t)$ is the velocity vector of the path $\gamma'(t)$ in TM.

LECTURE 3

Integration of equations of motion

A complete general solution can be obtained for three very important examples: a motion on the real line, a system of two interacting particles, including the Kepler problem, and the rotation of a rigid body.

3.1. One-dimensional motion

The motion of systems with one degree of freedom is called one-dimensional. In terms of a Cartesian coordinate x on $M = \mathbb{R}$, the Lagrangian takes the form

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

The conservation of energy

$$E = \frac{1}{2} m\dot{x}^2 + V(x),$$

allows to solve the equation of motion in a closed form by separation of variables. We have

$$\frac{dx}{dt} = \sqrt{\frac{2}{m}(E - V(x))},$$

so that

$$t = \sqrt{\frac{m}{2}} \int \frac{dx}{\sqrt{E - V(x)}}.$$

The inverse function $x(t)$ is a general solution of Newton's equation

$$m\ddot{x} = -\frac{dV}{dx},$$

with two arbitrary constants, the energy E and the constant of integration.

Since kinetic energy is non-negative, for a given value of E the actual motion takes place in the region of $\mathbb{R}$ where $V(x) \leq E$. The points where $V(x) = E$ are called *turning points*. The motion which is confined between two turning points is called *finite*. The finite motion is periodic — the particle oscillates between the turning points x_1 and x_2 with the period

$$T(E) = \sqrt{2m} \int_{x_1}^{x_2} \frac{dx}{\sqrt{E - V(x)}}.$$

If the region $V(x) \leq E$ is unbounded, then the motion is called *infinite* and the particle eventually goes to infinity. The regions where $V(x) > E$ are forbidden.

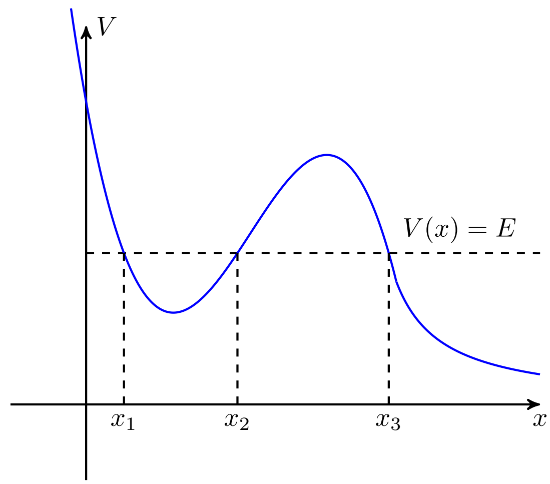

Figure 1

Thus on Fig. 1 the motion between points x_1 and x_2 is periodic, and in the region $x_3 \leq x$ the motion is infinite; all other regions there are forbidden.

On the phase plane with coordinates (x, y) Newton's equation reduces to the first order system

$$m\dot{x} = y, \quad \dot{y} = -\frac{dV}{dx}.$$

Trajectories correspond to the phase curves $(x(t), y(t))$, which lie on the level sets

$$\frac{y^2}{2m} + V(x) = E$$

of the energy function. The points $(x_0, 0)$, where x_0 is a critical point of the potential energy $V(x)$, correspond to the equilibrium solutions. The local minima correspond to the stable solutions and local maxima correspond to the unstable solutions. For the values of E which do not correspond to the equilibrium solutions the level sets are smooth curves. These curves are closed if the motion is finite.

The simplest non-trivial one-dimensional system, besides the free particle, is the harmonic oscillator with $V(x) = \frac{1}{2}kx^2$ $(k > 0)$, considered in Example 1.4. The general solution of the equation of motion is

$$x(t) = A\cos(\omega t + \alpha),$$

where A is the *amplitude*, $\omega = \sqrt{\dfrac{k}{m}}$ is the *frequency*, and α is the *phase* of a simple harmonic motion with the period $T = \dfrac{2\pi}{\omega}$. The energy is $E = \frac{1}{2}m\omega^2 A^2$ and the motion is finite with the same period T for $E > 0$.

3.2. Two-body problem

The motion of a system of two interacting particles — the *two-body problem* — can also be solved completely. Namely, in this case (see Example 1.2) $M = \mathbb{R}^6$ and

$$L = \frac{m_1 \dot{\boldsymbol{r}}_1^2}{2} + \frac{m_2 \dot{\boldsymbol{r}}_2^2}{2} - V(|\boldsymbol{r}_1 - \boldsymbol{r}_2|).$$

Introducing on $\mathbb{R}^6$ new coordinates

$$\boldsymbol{r} = \boldsymbol{r}_1 - \boldsymbol{r}_2 \quad \text{and} \quad \boldsymbol{R} = \frac{m_1 \boldsymbol{r}_1 + m_2 \boldsymbol{r}_2}{m_1 + m_2},$$

we get

$$L = \tfrac{1}{2} m \dot{\boldsymbol{R}}^2 + \tfrac{1}{2} \mu \dot{\boldsymbol{r}}^2 - V(|\boldsymbol{r}|),$$

where $m = m_1 + m_2$ is the *total mass* and $\mu = \dfrac{m_1 m_2}{m_1 + m_2}$ is the *reduced mass* of a two-body system. The Lagrangian L depends only on the velocity $\dot{\boldsymbol{R}}$ of the center of mass and not on its position $\boldsymbol{R}$. A generalized coordinate with this property is called *cyclic*. It follows from the Euler-Lagrange equations that generalized momentum corresponding to the cyclic coordinate is conserved. In our case it is a total momentum of the system,

$$\boldsymbol{P} = \frac{\partial L}{\partial \dot{\boldsymbol{R}}} = m \dot{\boldsymbol{R}},$$

so that the center of mass $\boldsymbol{R}$ moves uniformly. Thus in the reference frame $\boldsymbol{R} = 0$ the two-body problem reduces to the problem of a single particle of mass μ in the external central field $V(|\boldsymbol{r}|)$.

It follows from the conservation of angular momentum $\boldsymbol{M} = \mu \boldsymbol{r} \times \dot{\boldsymbol{r}}$ that during the motion position vector $\boldsymbol{r}$ lies in the plane P orthogonal to $\boldsymbol{M}$ in $\mathbb{R}^3$. Choosing the z-axis along $\boldsymbol{M}$, the plane P becomes the xy-plane and in polar coordinates

$$x = r \cos \varphi, \quad y = r \sin \varphi$$

the Lagrangian takes the form

$$L = \tfrac{1}{2} \mu (\dot{r}^2 + r^2 \dot{\varphi}^2) - V(r).$$

The coordinate φ is cyclic and its generalized momentum $\mu r^2 \dot{\varphi}$ coincides with $|\boldsymbol{M}|$ if $\dot{\varphi} > 0$ and with $-|\boldsymbol{M}|$ if $\dot{\varphi} < 0$. Denoting this quantity by M, we get the equation

$$\mu r^2 \dot{\varphi} = M, \tag{3.1}$$

which is equivalent to *Kepler's second law*[1]. Using (3.1) we get for the total energy

$$E = \tfrac{1}{2} \mu (\dot{r}^2 + r^2 \dot{\varphi}^2) + V(r) = \tfrac{1}{2} \mu \dot{r}^2 + V(r) + \frac{M^2}{2 \mu r^2}. \tag{3.2}$$

---

[1]It is the statement that *sectorial velocity* of a particle in a central field is constant.

Thus the radial motion reduces to a one-dimensional motion on the half-line $r > 0$ with the effective potential energy

$$V_{\text{eff}}(r) = V(r) + \frac{M^2}{2\mu r^2},$$

where the second term is called the *centrifugal energy.* As in the previous section, the solution is given by

$$t = \sqrt{\frac{\mu}{2}} \int \frac{dr}{\sqrt{E - V_{\text{eff}}(r)}}. \tag{3.3}$$

It follows from (3.1) that the angle φ is a monotonic function of t, given by another quadrature

$$\varphi = \frac{M}{\sqrt{2\mu}} \int \frac{dr}{r^2 \sqrt{E - V_{\text{eff}}(r)}}, \tag{3.4}$$

yielding an equation of the trajectory in polar coordinates.

The set $V_{\text{eff}}(r) \leq E$ is a union of annuli $0 \leq r_{min} \leq r \leq r_{max} \leq \infty$, and the motion is finite if $0 < r_{min} \leq r \leq r_{max} < \infty$. Though for a finite motion $r(t)$ oscillates between r_{min} and r_{max}, corresponding trajectories are not necessarily closed. The necessary and sufficient condition for a finite motion to have a closed trajectory is that the angle

$$\Delta\varphi = \frac{M}{\sqrt{2\mu}} \int_{r_{min}}^{r_{max}} \frac{dr}{r^2 \sqrt{E - V_{\text{eff}}(r)}}$$

is commensurable with 2π, i.e., $\Delta\varphi = 2\pi \frac{m}{n}$ for some $m, n \in \mathbb{Z}$. If the angle $\Delta\varphi$ is not commensurable with 2π, the orbit is everywhere dense in the annulus $r_{min} \leq r \leq r_{max}$. If

$$\lim_{r\to\infty} V_{\text{eff}}(r) = \lim_{r\to\infty} V(r) = V < \infty,$$

the motion is infinite for $E > V$ — the particle goes to ∞ with finite velocity $\sqrt{\frac{2}{\mu}(E - V)}$.

3.3. Kepler problem

A very important special case is when

$$V(r) = -\frac{\alpha}{r}.$$

It describes Newton's gravitational attraction ($\alpha > 0$) and Coulomb electrostatic interaction (either attractive or repulsive). First consider the case when $\alpha > 0$ — Kepler's problem. The effective potential energy is

$$V_{\text{eff}}(r) = -\frac{\alpha}{r} + \frac{M^2}{2\mu r^2}$$

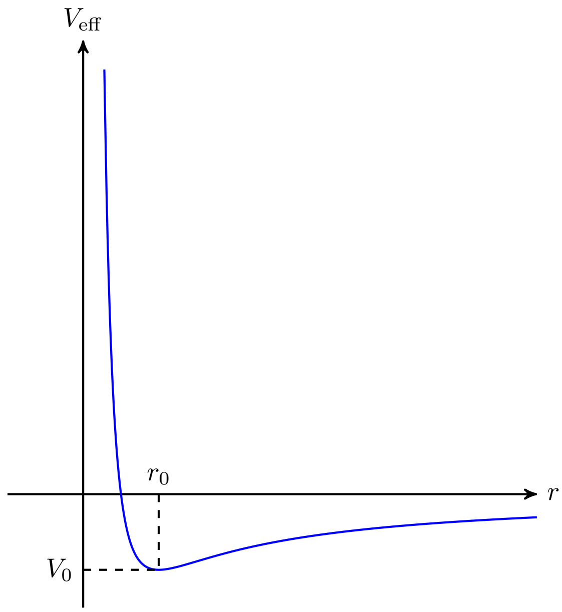

FIGURE 2

and has the global minimum

$$V_0 = -\frac{\alpha^2\mu}{2M^2}$$

at $r_0 = \dfrac{M^2}{\alpha\mu}$ (see Fig. 2). The motion is infinite for $E \geq 0$ and is finite for $V_0 \leq E < 0$. Since

$$2\mu(E - V_{\text{eff}}(r)) = 2\mu(E - V_0) - M^2\left(\frac{1}{r} - \frac{1}{r_0}\right)^2,$$

elementary integration in (3.4) gives

$$\varphi = \cos^{-1}\frac{\dfrac{M}{r} - \dfrac{M}{r_0}}{\sqrt{2\mu(E - V_0)}} + C,$$

which allows to determine the explicit form of trajectories.

Namely, choosing a constant of integration $C = 0$ and introducing the notation

$$p = r_0 \quad \text{and} \quad e = \sqrt{1 - \frac{E}{V_0}},$$

we get the equation of the orbit (trajectory)

$$\frac{p}{r} = 1 + e\cos\varphi. \tag{3.5}$$

This is the equation of a conic section with one focus at the origin. Quantity $2p$ is called the *latus rectum* of the orbit, and e is called the *eccentricity*. The choice $C = 0$ is such that the point with $\varphi = 0$ is the point nearest to the origin (called the *perihelion*). When $V_0 \leq E < 0$, the eccentricity $e < 1$ so that the orbit is the ellipse[2] with the major and minor semi-axes

$$a = \frac{p}{1-e^2} = \frac{\alpha}{2|E|}, \quad b = \frac{p}{\sqrt{1-e^2}} = \frac{|M|}{\sqrt{2\mu|E|}}. \tag{3.6}$$

Correspondingly, $r_{min} = \dfrac{p}{1+e}$, $r_{max} = \dfrac{p}{1-e}$, and the period T of elliptic orbit is given by

$$T = \pi\alpha\sqrt{\frac{\mu}{2|E|^3}}.$$

The last formula is *Kepler's third law*. When $E > 0$, the eccentricity $e > 1$ and the motion is infinite — the orbit is a hyperbola with the origin as internal focus. When $E = 0$, the eccentricity $e = 1$ — the particle starts from rest at ∞ and the orbit is a parabola.

For the repulsive case $\alpha < 0$ the effective potential energy $V_{\text{eff}}(r)$ is always positive and decreases monotonically from ∞ to 0. The motion is always infinite and the trajectories are hyperbolas (parabola if $E = 0$)

$$\frac{p}{r} = -1 + e\cos\varphi$$

with

$$p = \frac{M^2}{\alpha\mu} \quad \text{and} \quad e = \sqrt{1 + \frac{2EM^2}{\mu\alpha^2}}.$$

Kepler's problem is very special: for every $\alpha \in \mathbb{R}$ the Lagrangian system on $\mathbb{R}^3$ with

$$L = \tfrac{1}{2}\mu\dot{\boldsymbol{r}}^2 + \frac{\alpha}{r} \tag{3.7}$$

has three extra integrals of motion W_1, W_2, W_3 in addition to the components of the angular momentum $\boldsymbol{M}$. The corresponding vector $\boldsymbol{W} = (W_1, W_2, W_3)$, called the *Laplace-Runge-Lenz vector*, is given by

$$\boldsymbol{W} = \dot{\boldsymbol{r}} \times \boldsymbol{M} - \frac{\alpha\boldsymbol{r}}{r}. \tag{3.8}$$

Indeed, using equations of motion $\mu\ddot{\boldsymbol{r}} = -\dfrac{\alpha\boldsymbol{r}}{r^3}$ and conservation of the angular momentum $\boldsymbol{M} = \mu\boldsymbol{r} \times \dot{\boldsymbol{r}}$, we get

$$\begin{aligned}\dot{\boldsymbol{W}} &= \mu\ddot{\boldsymbol{r}} \times (\boldsymbol{r} \times \dot{\boldsymbol{r}}) - \frac{\alpha\dot{\boldsymbol{r}}}{r} + \frac{\alpha(\dot{\boldsymbol{r}} \cdot \boldsymbol{r})\boldsymbol{r}}{r^3} \\ &= (\mu\ddot{\boldsymbol{r}} \cdot \dot{\boldsymbol{r}})\boldsymbol{r} - (\mu\ddot{\boldsymbol{r}} \cdot \boldsymbol{r})\dot{\boldsymbol{r}} - \frac{\alpha\dot{\boldsymbol{r}}}{r} + \frac{\alpha(\dot{\boldsymbol{r}} \cdot \boldsymbol{r})\boldsymbol{r}}{r^3} \\ &= 0.\end{aligned}$$

[2]The statement that planets have elliptic orbits with a focus at the Sun is *Kepler's first law*.

Using $\mu(\dot{\boldsymbol{r}} \times \boldsymbol{M}) \cdot \boldsymbol{r} = \boldsymbol{M}^2$ and the identity $(\boldsymbol{a} \times \boldsymbol{b})^2 = \boldsymbol{a}^2\,\boldsymbol{b}^2 - (\boldsymbol{a} \cdot \boldsymbol{b})^2$, we get

$$\boldsymbol{W}^2 = \alpha^2 + \frac{2\boldsymbol{M}^2 E}{\mu} \tag{3.9}$$

where

$$E = \frac{\boldsymbol{p}^2}{2\mu} - \frac{\alpha}{r}$$

is the energy corresponding to the Lagrangian (3.7). The fact that all orbits are conic sections follows from this extra symmetry of the Kepler problem.

3.4. The motion of a rigid body

The configuration space of a rigid body in $\mathbb{R}^3$ with a fixed point is a Lie group $G = \mathrm{SO}(3)$ of orientation preserving orthogonal linear transformations in $\mathbb{R}^3$. Every left-invariant Riemannian metric $\langle\ ,\ \rangle$ on G defines a Lagrangian $L : TG \to \mathbb{R}$ by

$$L(v) = \tfrac{1}{2}\langle v, v\rangle, \quad v \in TG.$$

According to Example 1.5, equations of motion of a rigid body are geodesic equations on G with respect to the Riemannian metric $\langle\ ,\ \rangle$. Let $\mathfrak{g} = \mathfrak{so}(3)$ be the Lie algebra of G. A velocity vector $\dot{g} \in T_gG$ determines the *angular velocity of the body* $\Omega = (L_{g^{-1}})_*\dot{g} \in \mathfrak{g}$, where $L_g : G \to G$ are left translations on G. In terms of angular velocity, the Lagrangian takes the form

$$L = \tfrac{1}{2}\langle \Omega, \Omega\rangle_e, \tag{3.10}$$

where $\langle\ ,\ \rangle_e$ is an inner product on $\mathfrak{g} = T_eG$ given by the Riemannian metric $\langle\ ,\ \rangle$.

Let

$$B(x, y) = -\tfrac{1}{2}\operatorname{Tr} xy$$

be the Killing form on the Lie algebra $\mathfrak{g} = \mathfrak{so}(3)$ — the Lie algebra of 3×3 skew-symmetric matrices. It determines $\operatorname{ad}\mathfrak{g}$-invariant inner product on $\mathfrak{g}$,

$$B([x, z], y) + B(x, [y, z]) = 0$$

for all $x, y, z \in \mathfrak{g}$. Thus we have $\langle \Omega, \Omega\rangle_e = B(\boldsymbol{A}\cdot\Omega, \Omega)$ for some symmetric linear operator $\boldsymbol{A} : \mathfrak{g} \to \mathfrak{g}$, which is positive-definite with respect to the Killing form. Such linear operator $\boldsymbol{A}$ is called the *inertia tensor* of the body, and Lagrangian (3.10) takes the form

$$L = \tfrac{1}{2}B(\boldsymbol{A}\cdot\Omega, \Omega). \tag{3.11}$$

Now we are ready to derive equations of motion for Lagrangian (3.11). Similar to Sect. 1.2, for a path $g : [t_0, t_1] \to G$, consider the family

$$g(t, \varepsilon) = g(t)\exp\{\varepsilon u(t)\}, \quad \text{where} \quad u : [t_0, t_1] \to \mathfrak{g}, \quad u(t_0) = u(t_1) = 0,$$

and $\exp : \mathfrak{g} \to G$ is the exponential map. We have

$$\delta g(t) = \left.\frac{\partial g(t,\varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=0} = (L_g)_* u(t) \in T_{g(t)}G \quad \text{and} \quad u(t) = (L_{g(t)^{-1}})_* \delta g(t) \in \mathfrak{g}.$$

The corresponding angular velocity $\Omega(t,\varepsilon) = g^{-1}(t,\varepsilon)\dot{g}(t,\varepsilon) \in \mathfrak{g}$ takes the form

$$\Omega(t,\varepsilon) = \mathrm{Ad}_{h_\varepsilon(t)}\Omega(t) + \varepsilon\dot{u}(t), \quad \text{where} \quad h_\varepsilon(t) = \exp\{-\varepsilon u(t)\}, \quad \Omega(t) = \Omega(t,0),$$

and Ad_g stands for the adjoint action of G on $\mathfrak{g}$. Thus for the infinitesimal variation

$$\delta\Omega(t) = \left.\frac{\partial \Omega(t,\varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=0} \in \mathfrak{g}$$

we readily obtain

$$\delta\Omega = \dot{u} + [\Omega, u]. \tag{3.12}$$

As in Sect. 1.2 in Lecture 1, consider the action functional

$$S(g,\dot{g}) = \frac{1}{2}\int_{t_0}^{t_1} B(\boldsymbol{A}\cdot\Omega(t),\Omega(t))dt.$$

Using the symmetry of the operator $\boldsymbol{A}$ we obtain

$$\delta S = \frac{1}{2}\int_{t_0}^{t_1}\left(B(\boldsymbol{A}\cdot\delta\Omega(t),\Omega(t)) + B(\boldsymbol{A}\cdot\Omega(t),\delta\Omega(t))\right)dt = \int_{t_0}^{t_1} B(\boldsymbol{A}\cdot\Omega(t),\delta\Omega(t))dt,$$

and using (3.12), $\mathrm{ad}\,\mathfrak{g}$-invariance of the Killing form and integration by parts, we get

$$\begin{aligned}\delta S &= \int_{t_0}^{t_1} B(\boldsymbol{A}\cdot\Omega(t),\dot{u}(t) + [\Omega(t),u(t)])dt \\ &= \int_{t_0}^{t_1} B(-\boldsymbol{A}\cdot\dot{\Omega}(t) + [\boldsymbol{A}\cdot\Omega(t),\Omega(t)],u(t))dt.\end{aligned}$$

Since the Killing form is non-degenerate and $u(t)$ is an arbitrary smooth $\mathfrak{g}$-valued function with $u(t_1) = u(t_2) = 0$, from $\delta S = 0$ we obtain the following equations of motion

$$\boldsymbol{A}\cdot\dot{\Omega} = [\boldsymbol{A}\cdot\Omega,\Omega]. \tag{3.13}$$

Remark. Our derivation of equations of motion (3.13) is valid for any compact Lie group G and are called *Euler equations*. In general, for Lagrangian (3.10) we obtain the following equations of motion,

$$\boldsymbol{A}\cdot\dot{\Omega} = \mathrm{ad}^*_\Omega(\boldsymbol{A}\cdot\Omega),$$

where ad^*_Ω is the adjoint of the operator ad_Ω on $\mathfrak{g}$ with respect to the inner product $\langle\ ,\ \rangle_e$. These equations are called *Euler-Arnold* equations for the geodesics of a left-invariant Riemannian metric on a Lie group G, finite or infinite-dimensional.

Returning to the case $G = \mathrm{SO}(3)$, the *principal axes of inertia* of the body are orthonormal eigenvectors e_1, e_2, e_3 of $\boldsymbol{A}$; corresponding eigenvalues I_1, I_2, I_3 are called the *principal moments of inertia*. Choosing the principal axes of inertia as a basis in $\mathfrak{g}$ and setting $\Omega = \Omega_1 e_1 + \Omega_2 e_2 + \Omega_3 e_3$, we get the Lie algebra isomorphism $\mathfrak{g} \simeq \mathbb{R}^3$,

$$\mathfrak{g} \ni \Omega = \begin{pmatrix} 0 & -\Omega_3 & \Omega_2 \\ \Omega_3 & 0 & -\Omega_1 \\ -\Omega_2 & \Omega_1 & 0 \end{pmatrix} \mapsto (\Omega_1, \Omega_2, \Omega_3) \in \mathbb{R}^3,$$

where the Lie bracket in $\mathbb{R}^3$ is given by the cross-product (see Example 2.2). Indeed, for the matrices

$$a = \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} 0 & -b_3 & b_2 \\ b_3 & 0 & -b_1 \\ -b_2 & b_1 & 0 \end{pmatrix}$$

corresponding to the vectors $\boldsymbol{a} = (a_1, a_2, a_3)$ and $\boldsymbol{b} = (b_1, b_2, b_3)$ we have

$$[a, b] = c,$$

where c corresponds to the vector $\boldsymbol{c} = \boldsymbol{a} \times \boldsymbol{b}$. Moreover,

$$B(a, b) = \boldsymbol{a} \cdot \boldsymbol{b}.$$

It is easy to see that if $\boldsymbol{A} = \mathrm{diag}(I_1, I_2, I_3)$, then

$$\boldsymbol{A} \cdot \Omega = A\Omega + \Omega A,$$

where $A = \mathrm{diag}(l_1, l_2, l_3)$ and

$$l_1 = \frac{I_2 + I_3 - I_1}{2}, \; l_2 = \frac{I_1 + I_3 - I_2}{2}, \; l_3 = \frac{I_1 + I_2 - I_3}{2}.$$

Thus

$$[\boldsymbol{A} \cdot \Omega, \Omega] = A\Omega^2 - \Omega^2 A$$

and (3.13) become celebrated *Euler's equations for rotation of a free rigid body around a fixed point*,

$$\begin{aligned} I_1 \dot{\Omega}_1 &= (I_2 - I_3)\Omega_2\Omega_3, \\ I_2 \dot{\Omega}_2 &= (I_3 - I_1)\Omega_1\Omega_3, \\ I_3 \dot{\Omega}_3 &= (I_1 - I_2)\Omega_1\Omega_2 \end{aligned}$$

— the system of first order differential equations. Finally, the position $g(t)$ of a rigid body is determined from the first order linear matrix differential equation,

$$\dot{g} = g\Omega.$$

It is easy to see by direct computation that Euler's equations have two integrals of motion, total kinetic energy

$$T = I_1\Omega_1^2 + I_2\Omega_2^2 + I_3\Omega_3^2$$

and total angular momentum

$$M^2 = I_1^2\Omega_1^2 + I_2^2\Omega_2^2 + I_3^2\Omega_3^2.$$

Leaving aside the trivial case $I_1 = I_2 = I_3$, we conclude that the motion in $\mathbb{R}^3$ is constrained to the intersection of two quadrics which is a real form of elliptic curve.

Problem 3.1. Prove all statements in Sect. 3.2.

Problem 3.2. Show that if

$$\lim_{r\to 0} V_{\text{eff}}(r) = -\infty,$$

then there are orbits with $r_{min} = 0$ — "fall" of the particle to the center.

Problem 3.3. Prove that all finite trajectories in the central field are closed only when

$$V(r) = kr^2,\ \ k > 0, \quad \text{and} \quad V(r) = -\frac{\alpha}{r},\ \ \alpha > 0.$$

Problem 3.4 (Hamilton's Theorem). Prove that the velocity vector $\boldsymbol{v} = \dot{\boldsymbol{r}}(t)$ of the Kepler problem moves along a circle C in the plane P from Sect. 3.2, not in general centered at the origin. Any such "velocity circle" can occur, and a circle C, together with its orientation, determines the orbit $\boldsymbol{r} = \boldsymbol{r}(t)$ uniquely.

Problem 3.5. Derive Kepler's third law from Kepler's second law and equation (3.6).

Problem 3.6. Find parametric equations for orbits in the Kepler's problem.

Problem 3.7. For the Kepler problem, consider vector fields $\boldsymbol{Y} = (Y^1, Y^2, Y^3)$ on $\mathbb{R}^6$, defined by (2.4) with $a^{ij}(\boldsymbol{r}, \dot{\boldsymbol{r}}) = 2\dot{r}^i r^j - r^i\dot{r}^j - \delta^{ij}\boldsymbol{r}\cdot\dot{\boldsymbol{r}}$. Prove that they satisfy (2.5) with $\boldsymbol{K} = \dfrac{2\alpha\boldsymbol{r}}{r} = (K^1, K^2, K^3)$, and show that corresponding integrals of motions are components of the Laplace-Runge-Lenz vector.

Problem 3.8. Prove that the Laplace-Runge-Lenz vector $\boldsymbol{W}$ points in the direction of the major axis of the orbit and that $|\boldsymbol{W}| = \alpha e$, where e is the eccentricity of the orbit.

Problem 3.9. Using the conservation of the Laplace-Runge-Lenz vector, prove that trajectories in Kepler's problem with $E < 0$ are ellipses. (*Hint:* Evaluate $\boldsymbol{W}\cdot\boldsymbol{r}$ and use the previous problem.)

Problem 3.10. Derive Euler-Arnold equations.

Problem 3.11. In case $\mathfrak{g} = \mathfrak{so}(3)$ prove that for every symmetric $\boldsymbol{A} \in \operatorname{End}\mathfrak{g}$ there is a symmetric 3×3 matrix A such that

$$\boldsymbol{A}\cdot\Omega = A\Omega + \Omega A.$$

Problem 3.12. Solve Euler's equations.

LECTURE 4

Legendre transform and Hamilton's equations

4.1. Legendre transform

Let T^*M be the cotangent bundle of M. As in case of the tangent bundle, we have the following definition.

Definition. Let (U, φ) be a coordinate chart on M. Coordinates

$$(\boldsymbol{p}, \boldsymbol{q}) = (p_1, \dots, p_n, q^1, \dots, q^n)$$

on the chart $T^*U \simeq \mathbb{R}^n \times U$ on the cotangent bundle T^*M are called *standard coordinates*[1] if for $(p, q) \in T^*U$ and $f \in C^\infty(U)$

$$p_i(df) = \frac{\partial f}{\partial q^i}, \quad i = 1, \dots, n.$$

Equivalently, standard coordinates on T^*U are uniquely characterized by the condition that $\boldsymbol{p} = (p_1, \dots, p_n)$ are coordinates in the fiber corresponding to the basis $dq^1, \dots, dq^n$ for $T^*_q M$, dual to the basis $\dfrac{\partial}{\partial q^1}, \dots, \dfrac{\partial}{\partial q^n}$ for $T_q M$.

Definition. The 1-form θ on T^*M, defined in standard coordinates by

$$\theta = \sum_{i=1}^n p_i dq^i = \boldsymbol{p} d\boldsymbol{q},$$

is called *Liouville's canonical* 1*-form*.

Corollary 2.1 shows that θ is a well-defined 1-form on T^*M. It also admits invariant definition,

$$\theta(u) = p(\pi_*(u)), \quad \text{where} \quad u \in T_{(p,q)}T^*M,$$

and $\pi : T^*M \to M$ is the canonical projection.

Definition. A fibre-wise mapping $\tau_L : TM \to T^*M$ is called a *Legendre transform* associated with the Lagrangian L, if

$$\theta_L = \tau_L^*(\theta).$$

[1] Following tradition, the first n coordinates parametrize the fiber of T^*U and the last n coordinates parametrize the base.

In standard coordinates the Legendre transform is given by

$$\tau_L(\boldsymbol{q}, \dot{\boldsymbol{q}}) = (\boldsymbol{p}, \boldsymbol{q}), \quad \text{where} \quad \boldsymbol{p} = \frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\boldsymbol{q}, \dot{\boldsymbol{q}}).$$

The mapping τ_L is a local diffeomorphism if and only if the Lagrangian L is non-degenerate.

4.2. Hamiltonian function

DEFINITION. Suppose that the Legendre transform $\tau_L : TM \to T^*M$ is a diffeomorphism. The *Hamiltonian* function $H : T^*M \to \mathbb{R}$, associated with the Lagrangian $L : TM \to \mathbb{R}$, is defined by

$$H \circ \tau_L = E_L = \dot{\boldsymbol{q}} \frac{\partial L}{\partial \dot{\boldsymbol{q}}} - L.$$

In standard coordinates,

$$H(\boldsymbol{p}, \boldsymbol{q}) = \left. (\boldsymbol{p}\dot{\boldsymbol{q}} - L(\boldsymbol{q}, \dot{\boldsymbol{q}})) \right|_{\boldsymbol{p} = \frac{\partial L}{\partial \dot{\boldsymbol{q}}}},$$

where $\dot{\boldsymbol{q}}$ is a function of $\boldsymbol{p}$ and $\boldsymbol{q}$ defined by the equation $\boldsymbol{p} = \dfrac{\partial L}{\partial \dot{\boldsymbol{q}}}(\boldsymbol{q}, \dot{\boldsymbol{q}})$ through the implicit function theorem. The cotangent bundle T^*M is called the *phase space* of the Lagrangian system (M, L). It turns out that on the phase space the equations of motion take a very simple and symmetric form.

THEOREM 4.1. *Suppose that the Legendre transform $\tau_L : TM \to T^*M$ is a diffeomorphism. Then the Euler-Lagrange equations in standard coordinates on TM,*

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} = 0, \quad i = 1, \ldots, n,$$

are equivalent to the following system of first order differential equations in standard coordinates on T^*M:

$$\dot{p}_i = -\frac{\partial H}{\partial q^i}, \quad \dot{q}^i = \frac{\partial H}{\partial p_i}, \quad i = 1, \ldots, n.$$

PROOF. We have

$$\begin{aligned} dH &= \frac{\partial H}{\partial \boldsymbol{p}} d\boldsymbol{p} + \frac{\partial H}{\partial \boldsymbol{q}} d\boldsymbol{q} \\ &= \left. \left(\boldsymbol{p} d\dot{\boldsymbol{q}} + \dot{\boldsymbol{q}} d\boldsymbol{p} - \frac{\partial L}{\partial \boldsymbol{q}} d\boldsymbol{q} - \frac{\partial L}{\partial \dot{\boldsymbol{q}}} d\dot{\boldsymbol{q}} \right) \right|_{\boldsymbol{p} = \frac{\partial L}{\partial \dot{\boldsymbol{q}}}} \\ &= \left. \left(\dot{\boldsymbol{q}} d\boldsymbol{p} - \frac{\partial L}{\partial \boldsymbol{q}} d\boldsymbol{q} \right) \right|_{\boldsymbol{p} = \frac{\partial L}{\partial \dot{\boldsymbol{q}}}}. \end{aligned}$$

Thus under the Legendre transform,

$$\dot{\boldsymbol{q}} = \frac{\partial H}{\partial \boldsymbol{p}} \quad \text{and} \quad \dot{\boldsymbol{p}} = \frac{d}{dt}\frac{\partial L}{\partial \dot{\boldsymbol{q}}} = \frac{\partial L}{\partial \boldsymbol{q}} = -\frac{\partial H}{\partial \boldsymbol{q}}. \qquad \square$$

Corresponding first order differential equations on T^*M are called *Hamilton's equations* (*canonical equations*).

Corollary 4.2. *The Hamiltonian H is constant on the solutions of Hamilton's equations.*

Proof. For $H(t) = H(\boldsymbol{p}(t), \boldsymbol{q}(t))$ we have

$$\frac{dH}{dt} = \frac{\partial H}{\partial \boldsymbol{q}}\dot{\boldsymbol{q}} + \frac{\partial H}{\partial \boldsymbol{p}}\dot{\boldsymbol{p}} = \frac{\partial H}{\partial \boldsymbol{q}}\frac{\partial H}{\partial \boldsymbol{p}} - \frac{\partial H}{\partial \boldsymbol{p}}\frac{\partial H}{\partial \boldsymbol{q}} = 0. \qquad \square$$

For the Lagrangian

$$L = \frac{m\dot{\boldsymbol{r}}^2}{2} - V(\boldsymbol{r}) = T - V, \quad \boldsymbol{r} \in \mathbb{R}^3,$$

of a particle of mass m in a potential field $V(\boldsymbol{r})$ we have

$$\boldsymbol{p} = \frac{\partial L}{\partial \dot{\boldsymbol{r}}} = m\dot{\boldsymbol{r}}.$$

Thus the Legendre transform $\tau_L : T\mathbb{R}^3 \to T^*\mathbb{R}^3$ is a global diffeomorphism, linear on the fibers, and

$$H(\boldsymbol{p}, \boldsymbol{r}) = (\boldsymbol{p}\dot{\boldsymbol{r}} - L)|_{\dot{\boldsymbol{r}}=\frac{\boldsymbol{p}}{m}} = \frac{\boldsymbol{p}^2}{2m} + V(\boldsymbol{r}) = T + V.$$

Hamilton's equations

$$\begin{aligned}\dot{\boldsymbol{r}} &= \frac{\partial H}{\partial \boldsymbol{p}} = \frac{\boldsymbol{p}}{m},\\ \dot{\boldsymbol{p}} &= -\frac{\partial H}{\partial \boldsymbol{r}} = -\frac{\partial V}{\partial \boldsymbol{r}}\end{aligned}$$

are equivalent to Newton's equations with the force $\boldsymbol{F} = -\dfrac{\partial V}{\partial \boldsymbol{r}}$.

For the Lagrangian system describing small oscillators, considered in Example 1.4, we have $\boldsymbol{p} = m\dot{\boldsymbol{q}}$, and using normal coordinates we get

$$H(\boldsymbol{p}, \boldsymbol{q}) = (\boldsymbol{p}\dot{\boldsymbol{q}} - L(\boldsymbol{q}, \dot{\boldsymbol{q}}))|_{\dot{\boldsymbol{q}}=\frac{\boldsymbol{p}}{m}} = \frac{\boldsymbol{p}^2}{2m} + V_0(\boldsymbol{q}) = \frac{1}{2m}\big(\boldsymbol{p}^2 + m^2\sum_{i=1}^n \omega_i^2(q^i)^2\big).$$

Similarly, for the system of N interacting particles, considered in Example 1.2, we have $\boldsymbol{p} = (\boldsymbol{p}_1, \dots, \boldsymbol{p}_N)$, where

$$\boldsymbol{p}_a = \frac{\partial L}{\partial \dot{\boldsymbol{r}}_a} = m_a\dot{\boldsymbol{r}}_a, \quad a = 1, \dots, N.$$

The Legendre transform $\tau_L : T\mathbb{R}^{3N} \to T^*\mathbb{R}^{3N}$ is a global diffeomorphism, linear on the fibers, and

$$H(\boldsymbol{p}, \boldsymbol{r}) = (\boldsymbol{p}\dot{\boldsymbol{r}} - L)|_{\dot{\boldsymbol{r}}=\frac{\boldsymbol{p}}{m}} = \sum_{a=1}^N \frac{\boldsymbol{p}_a^2}{2m_a} + V(\boldsymbol{r}) = T + V.$$

In particular, for a closed system with pair-wise interaction,

$$H(\boldsymbol{p},\boldsymbol{r})=\sum_{a=1}^{N}\frac{\boldsymbol{p}_a^2}{2m_a}+\sum_{1\leq a<b\leq N}V_{ab}(\boldsymbol{r}_a-\boldsymbol{r}_b).$$

In general, consider the Lagrangian

$$L=\sum_{i,j=1}^{n}\tfrac{1}{2}a_{ij}(\boldsymbol{q})\dot{q}^i\dot{q}^j-V(\boldsymbol{q}),\ \ \boldsymbol{q}\in\mathbb{R}^n,$$

where $A(\boldsymbol{q})=\{a_{ij}(\boldsymbol{q})\}_{i,j=1}^n$ is a symmetric $n\times n$ matrix. We have

$$p_i=\frac{\partial L}{\partial\dot{q}^i}=\sum_{j=1}^{n}a_{ij}(\boldsymbol{q})\dot{q}^j,\quad i=1,\dots,n,$$

and the Legendre transform is a global diffeomorphism, linear on the fibers, if and only if the matrix $A(\boldsymbol{q})$ is non-degenerate for all $\boldsymbol{q}\in\mathbb{R}^n$. In this case,

$$H(\boldsymbol{p},\boldsymbol{q})=\left.(\boldsymbol{p}\dot{\boldsymbol{q}}-L(\boldsymbol{q},\dot{\boldsymbol{q}}))\right|_{\boldsymbol{p}=\frac{\partial L}{\partial\dot{\boldsymbol{q}}}}=\sum_{i,j=1}^{n}\tfrac{1}{2}a^{ij}(\boldsymbol{q})p_ip_j+V(\boldsymbol{q}),$$

where $\{a^{ij}(\boldsymbol{q})\}_{i,j=1}^n=A^{-1}(\boldsymbol{q})$ is the inverse matrix.

4.3. Hamilton's equations

With every function $H:T^*M\to\mathbb{R}$ on the phase space T^*M there are associated Hamilton's equations — a first-order system of ordinary differential equations, which in the standard coordinates on T^*U has the form

$$\dot{\boldsymbol{p}}=-\frac{\partial H}{\partial\boldsymbol{q}},\quad\dot{\boldsymbol{q}}=\frac{\partial H}{\partial\boldsymbol{p}}.\tag{4.1}$$

The corresponding vector field X_H on T^*U,

$$X_H=\sum_{i=1}^{n}\left(\frac{\partial H}{\partial p_i}\frac{\partial}{\partial q^i}-\frac{\partial H}{\partial q^i}\frac{\partial}{\partial p_i}\right)=\frac{\partial H}{\partial\boldsymbol{p}}\frac{\partial}{\partial\boldsymbol{q}}-\frac{\partial H}{\partial\boldsymbol{q}}\frac{\partial}{\partial\boldsymbol{p}},$$

gives rise to a well-defined vector field X_H on T^*M, called the *Hamiltonian vector field*. Suppose now that the vector field X_H on T^*M is complete, i.e., its integral curves exist for all times. The corresponding one-parameter group $\{g_t\}_{t\in\mathbb{R}}$ of diffeomorphisms of T^*M generated by X_H is called the *Hamiltonian phase flow*. It is defined by $g_t(p,q)=(p(t),q(t))$, where $p(t)$, $q(t)$ is a solution of Hamilton's equations satisfying $p(0)=p$, $q(0)=q$.

Liouville's canonical 1-form θ on T^*M defines a 2-form $\omega=d\theta$. In standard coordinates on T^*M it is given by

$$\omega=\sum_{i=1}^{n}dp_i\wedge dq^i=d\boldsymbol{p}\wedge d\boldsymbol{q},$$

and is a non-degenerate 2-form. The form ω is called *canonical symplectic form* on T^*M. The symplectic form ω defines an isomorphism

$$J : T^*(T^*M) \to T(T^*M)$$

between tangent and cotangent bundles to T^*M. For every $(p, q) \in T^*M$ the linear mapping $J^{-1} : T_{(p,q)}T^*M \to T^*_{(p,q)}T^*M$ is given by

$$\omega(u_1, u_2) = J^{-1}(u_2)(u_1), \quad u_1, u_2 \in T_{(p,q)}T^*M.$$

The mapping J induces the isomorphism

$$\mathcal{A}^1(T^*M) \simeq \mathrm{Vect}(T^*M)$$

between the infinite-dimensional vector spaces, which is linear over the ring $C^\infty(T^*M)$. Namely, if ϑ is a 1-form on T^*M, then the corresponding vector field $X = J(\vartheta)$ on T^*M satisfies

$$\omega(Y, X) = \vartheta(Y) \quad \text{for all} \quad Y \in \mathrm{Vect}(T^*M), \tag{4.2}$$

and, correspondingly,

$$\vartheta = J^{-1}(X) = -i_X\omega. \tag{4.3}$$

In particular, in standard coordinates,

$$J(d\boldsymbol{p}) = \frac{\partial}{\partial \boldsymbol{q}} \quad \text{and} \quad J(d\boldsymbol{q}) = -\frac{\partial}{\partial \boldsymbol{p}},$$

so that $X_H = J(dH)$. In this notation, for every $f \in C^\infty(T^*M)$,

$$df = -i_{X_f}\omega. \tag{4.4}$$

THEOREM 4.3. *The Hamiltonian phase flow on T^*M preserves the canonical symplectic form.*

PROOF. We need to prove that $(g_t)^*\omega = \omega$. Since g_t is a one-parameter group of diffeomorphisms, it is sufficient to show that

$$\left.\frac{d}{dt}(g_t)^*\omega\right|_{t=0} = \mathcal{L}_{X_H}\omega = 0,$$

where $\mathcal{L}_{X_H}$ is the Lie derivative along the Hamiltonian vector field X_H. Since for every vector field X,

$$\mathcal{L}_X(df) = d(X(f)),$$

we compute

$$\mathcal{L}_{X_H}(dp_i) = -d\left(\frac{\partial H}{\partial q^i}\right) \quad \text{and} \quad \mathcal{L}_{X_H}(dq^i) = d\left(\frac{\partial H}{\partial p_i}\right),$$

so that

$$\begin{aligned}\mathcal{L}_{X_H}\omega &= \sum_{i=1}^{n}\left(\mathcal{L}_{X_H}(dp_i)\wedge dq^i + dp_i\wedge \mathcal{L}_{X_H}(dq^i)\right)\\ &= \sum_{i=1}^{n}\left(-d\left(\frac{\partial H}{\partial q^i}\right)\wedge dq^i + dp_i\wedge d\left(\frac{\partial H}{\partial p_i}\right)\right) = -d(dH) = 0. \qquad \square\end{aligned}$$

The canonical symplectic form ω on T^*M defines the volume form

$$\frac{\omega^n}{n!} = \frac{1}{n!}\underbrace{\omega\wedge\cdots\wedge\omega}_{n}$$

on T^*M, called *Liouville's volume form.*

COROLLARY 4.4 (Liouville's theorem). *The Hamiltonian phase flow on T^*M preserves Liouville's volume form.*

The restriction of the symplectic form ω on T^*M to the configuration space M is 0. Generalizing this property, we get the following notion.

DEFINITION. A submanifold $\mathscr{L}$ of the phase space T^*M is called a *Lagrangian submanifold* if $\dim\mathscr{L} = \dim M$ and $\omega|_{\mathscr{L}} = 0$.

It follows from Theorem 4.3 that the image of a Lagrangian submanifold under the Hamiltonian phase flow is a Lagrangian submanifold.

PROBLEM 4.1. Suppose that for a Lagrangian system $(\mathbb{R}^n, L)$ the Legendre transform τ_L is a diffeomorphism and let H be the corresponding Hamiltonian. Prove that for fixed $\boldsymbol{q}$ and $\dot{\boldsymbol{q}}$ the function $\boldsymbol{p}\dot{\boldsymbol{q}} - H(\boldsymbol{p},\boldsymbol{q})$ has a single critical point at $\boldsymbol{p} = \dfrac{\partial L}{\partial \dot{\boldsymbol{q}}}$.

PROBLEM 4.2. Give an example of a non-degenerate Lagrangian system (M, L) such that the Legendre transform $\tau_L : TM \to T^*M$ is one-to-one but not onto.

PROBLEM 4.3. Verify that X_H is a well-defined vector field on T^*M.

PROBLEM 4.4. Show that if all level sets of the Hamiltonian H are compact submanifolds of T^*M, then the Hamiltonian vector field X_H is complete.

PROBLEM 4.5. Prove that $\mathcal{L}_{X_H}(\theta) = d(-H + \theta(X_H))$, where θ is Liouville's canonical 1-form.

LECTURE 5

Hamiltonian formalism

5.1. The action functional in the phase space

With every function H on the phase space T^*M there is an associated 1-form

$$\theta - H dt = \boldsymbol{p} d\boldsymbol{q} - H dt$$

on the *extended phase space* $T^*M \times \mathbb{R}$, called the *Poincaré-Cartan form.* Let $\gamma : [t_0, t_1] \to T^*M$ be a smooth parametrized path in T^*M such that $\pi(\gamma(t_0)) = q_0$ and $\pi(\gamma(t_1)) = q_1$, where $\pi : T^*M \to M$ is the canonical projection. By definition, the lift of a path γ to the extended phase space $T^*M \times \mathbb{R}$ is a path $\sigma : [t_0, t_1] \to T^*M \times \mathbb{R}$ given by $\sigma(t) = (\gamma(t), t)$, and a path σ in $T^*M \times \mathbb{R}$ is called an *admissible* path if it is a lift of a path γ in T^*M. The space of admissible paths in $T^*M \times \mathbb{R}$ is denoted by $\tilde{P}(T^*M)^{q_1,t_1}_{q_0,t_0}$. A variation of an admissible path σ is a smooth family of admissible paths σ_ε, where $\varepsilon \in [-\varepsilon_0, \varepsilon_0]$ and $\sigma_0 = \sigma$, and the corresponding infinitesimal variation is

$$\delta\sigma = \left.\frac{\partial \sigma_\varepsilon}{\partial \varepsilon}\right|_{\varepsilon=0} \in T_\sigma \tilde{P}(T^*M)^{q_1,t_1}_{q_0,t_0}$$

(cf. Section 1.2). The principle of least action in the phase space is the following statement.

THEOREM 5.1 (Poincaré). *The admissible path σ in $T^*M \times \mathbb{R}$ is an extremal for the action functional*

$$S(\sigma) = \int_\sigma (\boldsymbol{p} d\boldsymbol{q} - H dt) = \int_{t_0}^{t_1} (\boldsymbol{p}\dot{\boldsymbol{q}} - H) dt$$

if and only if it is a lift of a path $\gamma(t) = (\boldsymbol{p}(t), \boldsymbol{q}(t))$ in T^*M, where $\boldsymbol{p}(t)$ and $\boldsymbol{q}(t)$ satisfy canonical Hamilton's equations

$$\dot{\boldsymbol{p}} = -\frac{\partial H}{\partial \boldsymbol{q}}, \quad \dot{\boldsymbol{q}} = \frac{\partial H}{\partial \boldsymbol{p}}.$$

PROOF. As in the proof of Theorem 1.1, for an admissible family $\sigma_\varepsilon(t) = (\boldsymbol{p}(t, \varepsilon), \boldsymbol{q}(t, \varepsilon), t)$ we compute using integration by parts,

$$\begin{aligned}\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} S(\sigma_\varepsilon) &= \sum_{i=1}^n \int_{t_0}^{t_1} \left(\dot{q}^i \delta p_i - \dot{p}_i \delta q^i - \frac{\partial H}{\partial q^i} \delta q^i - \frac{\partial H}{\partial p_i} \delta p_i \right) dt \\ &\quad + \sum_{i=1}^n p_i \left.\delta q^i\right|_{t_0}^{t_1}.\end{aligned}$$

Since $\delta\boldsymbol{q}(t_0) = \delta\boldsymbol{q}(t_1) = 0$, the path σ is critical if and only if $\boldsymbol{p}(t)$ and $\boldsymbol{q}(t)$ satisfy canonical Hamilton's equations (4.1). $\square$

Remark. For a Lagrangian system (M, L), every path $\gamma(t) = (\boldsymbol{q}(t))$ in the configuration space M connecting points q_0 and q_1 defines an admissible path $\hat{\gamma}(t) = (\boldsymbol{p}(t), \boldsymbol{q}(t), t)$ in the phase space T^*M by setting $\boldsymbol{p} = \dfrac{\partial L}{\partial \dot{\boldsymbol{q}}}$. If the Legendre transform $\tau_L : TM \to T^*M$ is a diffeomorphism, then

$$S(\hat{\gamma}) = \int_{t_0}^{t_1} (\boldsymbol{p}\dot{\boldsymbol{q}} - H)dt = \int_{t_0}^{t_1} L(\gamma'(t), t)dt.$$

Thus the principle of the least action in a configuration space — Hamilton's principle — follows from the principle of the least action in a phase space. In fact, in this case the two principles are equivalent (see Problem 4.1).

From Corollary 4.2 we immediately get the following result.

Corollary 5.2. *Solutions of canonical Hamilton's equations lying on the hypersurface $H(\boldsymbol{p}, \boldsymbol{q}) = E$ are extremals of the functional $\int_\sigma \boldsymbol{p}d\boldsymbol{q}$ in the class of admissible paths σ lying on this hypersurface.*

Corollary 5.3 (Maupertuis' principle). *The trajectory $\gamma = (\boldsymbol{q}(\tau))$ of a closed Lagrangian system (M, L) connecting points q_0 and q_1 and having energy E is the extremal of the functional*

$$\int_\gamma \boldsymbol{p}d\boldsymbol{q} = \int_\gamma \frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\boldsymbol{q}(\tau), \dot{\boldsymbol{q}}(\tau))\dot{\boldsymbol{q}}(\tau)d\tau$$

on the space of all paths in the configuration space M connecting points q_0 and q_1 and parametrized such that $H(\frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\tau), \boldsymbol{q}(\tau)) = E$.

The functional

$$S_0(\gamma) = \int_\gamma \boldsymbol{p}d\boldsymbol{q}$$

is called the *abbreviated action*[1].

Proof. Every path $\gamma = \boldsymbol{q}(\tau)$, parametrized such that $H(\frac{\partial L}{\partial \dot{\boldsymbol{q}}}, \boldsymbol{q}) = E$, lifts to an admissible path $\sigma = (\frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\tau), \boldsymbol{q}(\tau), \tau)$, $a \leq \tau \leq b$, lying on the hypersurface $H(\boldsymbol{p}, \boldsymbol{q}) = E$. $\square$

5.2. The action as a function of coordinates

Consider a non-degenerate Lagrangian system (M, L) and denote by $\gamma(t; q_0, v_0)$ the solution of Euler-Lagrange equations

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{\boldsymbol{q}}} - \frac{\partial L}{\partial \boldsymbol{q}} = 0$$

[1]The accurate formulation of Maupertuis' principle is due to Euler and Lagrange.

with the initial conditions $\gamma(t_0) = q_0 \in M$ and $\dot{\gamma}(t_0) = v_0 \in T_{q_0}M$. Suppose that there exist a neighborhood $V_0 \subset T_{v_0}M$ of v_0 and $t_1 > t_0$ such that for all $v \in V_0$ the extremals $\gamma(t; q_0, v)$, which start at time t_0 at q_0, do not intersect in the extended configuration space $M \times \mathbb{R}$ for times $t_0 < t < t_1$. Such extremals are said to form a *central field* which includes the extremal $\gamma_0(t) = \gamma(t; q_0, v_0)$. The existence of the central field of extremals is equivalent to the condition that for every $t_0 < t < t_1$ there is a neighborhood $U_t \subset M$ of $\gamma_0(t) \in M$ such that the mapping

$$V_0 \ni v \mapsto q(t) = \gamma(t; q_0, v) \in U_t \tag{5.1}$$

is a diffeomorphism. Basic theorems in the theory of ordinary differential equations guarantee that for t_1 sufficiently close to t_0 every extremal $\gamma(t)$ for $t_0 < t < t_1$ can be included into the central field. In standard coordinates the mapping (5.1) is given by $\dot{\boldsymbol{q}} \mapsto \boldsymbol{q}(t) = \gamma(t; \boldsymbol{q}_0, \dot{\boldsymbol{q}})$.

For the central field of extremals $\gamma(t; \boldsymbol{q}_0, \dot{\boldsymbol{q}})$, $t_0 < t < t_1$, we define the *action as a function of coordinates and time* (or, *classical action*) by

$$S(\boldsymbol{q}, t; \boldsymbol{q}_0, t_0) = \int_{t_0}^{t} L(\gamma'(\tau))d\tau,$$

where $\gamma(\tau)$ is the extremal from the central field that connects $\boldsymbol{q}_0$ and $\boldsymbol{q}$. For given $\boldsymbol{q}_0$ and t_0, the classical action is defined for $t \in (t_0, t_1)$ and $\boldsymbol{q} \in \bigcup_{t_0<t<t_1} U_t$. For a fixed energy E,

$$S(\boldsymbol{q}, t; \boldsymbol{q}_0, t_0) = S_0(\boldsymbol{q}, t; \boldsymbol{q}_0, t_0) - E(t - t_0), \tag{5.2}$$

where S_0 is the abbreviated action from the previous section.

THEOREM 5.4. *The differential of the classical action $S(\boldsymbol{q}, t)$ with fixed initial point is given by*

$$dS = \boldsymbol{p}d\boldsymbol{q} - Hdt,$$

where $\boldsymbol{p} = \dfrac{\partial L}{\partial \dot{\boldsymbol{q}}}(\boldsymbol{q}, \dot{\boldsymbol{q}})$ and $H = \boldsymbol{p}\dot{\boldsymbol{q}} - L(\boldsymbol{q}, \dot{\boldsymbol{q}})$ are determined by the velocity $\dot{\boldsymbol{q}}$ of the extremal $\gamma(\tau)$ at time t.

PROOF. Let $\boldsymbol{q}_\varepsilon$ be a path in M passing through $\boldsymbol{q}$ at $\varepsilon = 0$ with the tangent vector $\boldsymbol{v} \in T_{\boldsymbol{q}}M \simeq \mathbb{R}^n$, and for ε small enough let $\gamma_\varepsilon(\tau)$ be the family of extremals from the central field satisfying $\gamma_\varepsilon(t_0) = \boldsymbol{q}_0$ and $\gamma_\varepsilon(t) = \boldsymbol{q}_\varepsilon$. For the infinitesimal variation $\delta\gamma$ we have $\delta\gamma(t_0) = 0$ and $\delta\gamma(t) = \boldsymbol{v}$, and for fixed t we get from the formula for variation with the free ends (1.6) that

$$dS(\boldsymbol{v}) = \frac{\partial L}{\partial \dot{\boldsymbol{q}}}\boldsymbol{v}.$$

This shows that $\dfrac{\partial S}{\partial \boldsymbol{q}} = \boldsymbol{p}$. Setting $\boldsymbol{q}(t) = \gamma(t)$, we obtain

$$\frac{d}{dt}S(\boldsymbol{q}(t), t) = \frac{\partial S}{\partial \boldsymbol{q}}\dot{\boldsymbol{q}} + \frac{\partial S}{\partial t} = L,$$

so that $\dfrac{\partial S}{\partial t} = L - \boldsymbol{p}\dot{\boldsymbol{q}} = -H.$ $\square$

Corollary 5.5. *The classical action satisfies the following nonlinear partial differential equation*

$$\frac{\partial S}{\partial t} + H\left(\frac{\partial S}{\partial \boldsymbol{q}}, \boldsymbol{q}\right) = 0. \tag{5.3}$$

This equation is called the *Hamilton-Jacobi equation.* Hamilton's equations (4.1) can be used for solving the Cauchy problem

$$S(\boldsymbol{q}, t)|_{t=0} = s(\boldsymbol{q}), \quad s \in C^\infty(M), \tag{5.4}$$

for Hamilton-Jacobi equation (5.3) by the method of characteristics.

We can also consider the action $S(\boldsymbol{q}, t; \boldsymbol{q}_0, t_0)$ as a function of both variables $\boldsymbol{q}$ and $\boldsymbol{q}_0$. The analog of Theorem 5.4 is the following statement.

Proposition 5.1. *The differential of the classical action as a function of initial and final points is given by*

$$dS = \boldsymbol{p}d\boldsymbol{q} - \boldsymbol{p}_0 d\boldsymbol{q}_0 - H(\boldsymbol{p}, \boldsymbol{q})dt + H(\boldsymbol{p}_0, \boldsymbol{q}_0)dt_0.$$

5.3. Classical observables and Poisson bracket

Smooth real-valued functions on the phase space T^*M are called *classical observables.* The vector space $C^\infty(T^*M)$ is an $\mathbb{R}$-algebra — an associative algebra over $\mathbb{R}$ with a unit given by the constant function 1, and with a multiplication given by the point-wise product of functions. The commutative algebra $C^\infty(T^*M)$ is called the *algebra of classical observables.* Assuming that the Hamiltonian phase flow g_t exists for all times, the time evolution of every observable $f \in C^\infty(T^*M)$ is given by

$$f_t(p, q) = f(g_t(p, q)) = f(p(t), q(t)), \quad (p, q) \in T^*M.$$

Equivalently, using the Hamiltonian vector field

$$X_H = \frac{\partial H}{\partial \boldsymbol{p}}\frac{\partial}{\partial \boldsymbol{q}} - \frac{\partial H}{\partial \boldsymbol{q}}\frac{\partial}{\partial \boldsymbol{p}},$$

the time evolution is described by the differential equation

$$\begin{aligned}
\frac{df_t}{dt} &= \left.\frac{df_{s+t}}{ds}\right|_{s=0} = \left.\frac{d(f_t \circ g_s)}{ds}\right|_{s=0} = X_H(f_t) \\
&= \sum_{i=1}^n \left(\frac{\partial H}{\partial p_i}\frac{\partial f_t}{\partial q^i} - \frac{\partial H}{\partial q^i}\frac{\partial f_t}{\partial p_i}\right) = \frac{\partial H}{\partial \boldsymbol{p}}\frac{\partial f_t}{\partial \boldsymbol{q}} - \frac{\partial H}{\partial \boldsymbol{q}}\frac{\partial f_t}{\partial \boldsymbol{p}},
\end{aligned}$$

called Hamilton's equation for classical observables. Setting

$$\{f, g\} = X_f(g) = \frac{\partial f}{\partial \boldsymbol{p}}\frac{\partial g}{\partial \boldsymbol{q}} - \frac{\partial f}{\partial \boldsymbol{q}}\frac{\partial g}{\partial \boldsymbol{p}}, \quad f, g \in C^\infty(T^*M), \tag{5.5}$$

we can rewrite Hamilton's equation in the concise form

$$\frac{df}{dt} = \{H, f\}, \tag{5.6}$$

where it is understood that (5.6) is a differential equation for a family of functions f_t on T^*M with the initial condition $f_t(p,q)|_{t=0} = f(p,q)$. The properties of the bilinear mapping

$$\{\ ,\ \} : C^\infty(T^*M) \times C^\infty(T^*M) \to C^\infty(T^*M)$$

are summarized below.

THEOREM 5.6. *The mapping* $\{\ ,\ \}$ *satisfies the following properties.*

(i) *(Relation with the symplectic form)*

$$\{f, g\} = \omega(J(df), J(dg)) = \omega(X_f, X_g).$$

(ii) *(Skew-symmetry)*

$$\{f, g\} = -\{g, f\}.$$

(iii) *(Leibniz rule)*

$$\{fg, h\} = f\{g, h\} + g\{f, h\}.$$

(iv) *(Jacobi identity)*

$$\{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{f, g\}\} = 0$$

for all $f, g, h \in C^\infty(T^*M)$.

PROOF. Property (i) immediately follows from the definitions of ω and J in Section 4.3. Namely, it follows from (4.2) that

$$\omega(X_f, X_g) = \omega(X_f, J(dg)) = dg(X_f) = X_f(g) = \{f, g\}.$$

Properties (ii)-(iii) are obvious. The Jacobi identity could be verified by a direct computation using (5.5), or by the following elegant argument. Observe that $\{f, g\}$ is a bilinear form in the first partial derivatives of f and g, and every term in the left-hand side of the Jacobi identity is a linear homogenous function of second partial derivatives of f, g, and h. Now the only terms in the Jacobi identity which could actually contain second partial derivatives of a function h are the following:

$$\{f, \{g, h\}\} + \{g, \{h, f\}\} = (X_f X_g - X_g X_f)(h).$$

However, this expression does not contain second partial derivatives of h since it is a commutator of two differential operators of the first order which is again a differential operator of the first order! $\square$

The observable $\{f, g\}$ is called the *canonical Poisson bracket* of the observables f and g. The Poisson bracket map $\{\ ,\ \} : C^\infty(T^*M) \times C^\infty(T^*M) \to C^\infty(T^*M)$ turns the algebra of classical observables $C^\infty(T^*M)$ into a Lie algebra with a Lie bracket given by the Poisson bracket. It has an important property that the Lie bracket is a bi-derivation with respect to the multiplication in $C^\infty(T^*M)$. The algebra of classical observables $C^\infty(T^*M)$ is an example of the *Poisson algebra* — a commutative algebra over $\mathbb{R}$ carrying a structure of a Lie algebra with the property that the Lie bracket is a derivation with respect to the algebra product.

In Lagrangian mechanics, a function I on TM is an integral of motion for the Lagrangian system (M, L) if it is constant along the trajectories. In Hamiltonian mechanics, an observable I — a function on the phase space T^*M — is called an integral of motion (first integral) for Hamilton's equations (4.1) if it is constant along the Hamiltonian phase flow. According to (5.6), this is equivalent to the condition

$$\{H, I\} = 0.$$

It is said that the observables H and I are *in involution* (*Poisson commute*).

5.4. Canonical transformations and generating functions

Definition. A diffeomorphism g of the phase space T^*M is called a *canonical transformation*, if it preserves the canonical symplectic form ω on T^*M, i.e., $g^*(\omega) = \omega$. By Theorem 4.3, the Hamiltonian phase flow g_t is a one-parameter group of canonical transformations.

Proposition 5.2. *Canonical transformations preserve Hamilton's equations.*

Proof. From $g^*(\omega) = \omega$ it follows that the mapping $J : T^*(T^*M) \to T(T^*M)$ satisfies

$$g_* \circ J \circ g^* = J. \tag{5.7}$$

Indeed, for all $X, Y \in \mathrm{Vect}(M)$ we have[2]

$$\omega(X, Y) = g^*(\omega)(X, Y) = \omega(g_*(X), g_*(Y)) \circ g,$$

so that for every 1-form ϑ on M,

$$\omega(X, J(g^*(\vartheta))) = g^*(\vartheta)(X) = \vartheta(g_*(X)) \circ g = \omega(g_*(X), J(\vartheta)) \circ g,$$

which gives $J(g^*(\vartheta)) = g_*^{-1}(J(\vartheta))$. Using (5.7), we get

$$g_*(X_H) = g_*(J(dH)) = J((g^*)^{-1}(dH)) = X_K,$$

where $K = H \circ g^{-1}$. Thus the canonical transformation g maps trajectories of the Hamiltonian vector field X_H into the trajectories of the Hamiltonian vector field X_K. □

[2] Since g is a diffeomorphism, g_*X is a well-defined vector field on M.

Remark. In classical terms, Proposition 5.2 means that canonical Hamilton's equations

$$\dot{\boldsymbol{p}} = -\frac{\partial H}{\partial \boldsymbol{q}}(\boldsymbol{p},\boldsymbol{q}), \quad \dot{\boldsymbol{q}} = \frac{\partial H}{\partial \boldsymbol{p}}(\boldsymbol{p},\boldsymbol{q})$$

in new coordinates $(\boldsymbol{P},\boldsymbol{Q}) = g(\boldsymbol{p},\boldsymbol{q})$ continue to have the canonical form

$$\dot{\boldsymbol{P}} = -\frac{\partial K}{\partial \boldsymbol{Q}}(\boldsymbol{P},\boldsymbol{Q}), \quad \dot{\boldsymbol{Q}} = \frac{\partial K}{\partial \boldsymbol{P}}(\boldsymbol{P},\boldsymbol{Q})$$

with the old Hamiltonian function $K(\boldsymbol{P},\boldsymbol{Q}) = H(\boldsymbol{p},\boldsymbol{q})$.

Consider now the classical case $M = \mathbb{R}^n$. For a canonical transformation $(\boldsymbol{P},\boldsymbol{Q}) = g(\boldsymbol{p},\boldsymbol{q})$ set $\boldsymbol{P} = \boldsymbol{P}(\boldsymbol{p},\boldsymbol{q})$ and $\boldsymbol{Q} = \boldsymbol{Q}(\boldsymbol{p},\boldsymbol{q})$. Since $d\boldsymbol{P}\wedge d\boldsymbol{Q} = d\boldsymbol{p}\wedge d\boldsymbol{q}$ on $T^*M \simeq \mathbb{R}^{2n}$, the 1-form $\boldsymbol{p}d\boldsymbol{q} - \boldsymbol{P}d\boldsymbol{Q}$ — the difference between the canonical Liouville 1-form and its pullback by the mapping g — is closed. From the Poincaré lemma it follows that there exists a function $F(\boldsymbol{p},\boldsymbol{q})$ on $\mathbb{R}^{2n}$ such that

$$\boldsymbol{p}d\boldsymbol{q} - \boldsymbol{P}d\boldsymbol{Q} = dF(\boldsymbol{p},\boldsymbol{q}). \tag{5.8}$$

Now assume that at some point $(\boldsymbol{p}_0,\boldsymbol{q}_0)$ the $n\times n$ matrix $\dfrac{\partial \boldsymbol{P}}{\partial \boldsymbol{p}} = \left\{\dfrac{\partial P_i}{\partial p_j}\right\}_{i,j=1}^n$ is non-degenerate. By the inverse function theorem, there exists a neighborhood U of $(\boldsymbol{p}_0,\boldsymbol{q}_0)$ in $\mathbb{R}^{2n}$ for which the functions $\boldsymbol{P},\boldsymbol{q}$ are coordinate functions. The function

$$S(\boldsymbol{P},\boldsymbol{q}) = F(\boldsymbol{p},\boldsymbol{q}) + \boldsymbol{P}\boldsymbol{Q}$$

is called a *generating function* of the canonical transformation g in U. It follows from (5.8) that

$$dS = \boldsymbol{p}d\boldsymbol{q} + \boldsymbol{Q}d\boldsymbol{P},$$

whence in new coordinates $\boldsymbol{P},\boldsymbol{q}$ on U,

$$\boldsymbol{p} = \frac{\partial S}{\partial \boldsymbol{q}}(\boldsymbol{P},\boldsymbol{q}) \quad \text{and} \quad \boldsymbol{Q} = \frac{\partial S}{\partial \boldsymbol{P}}(\boldsymbol{P},\boldsymbol{q}).$$

The converse statement below easily follows from the implicit function theorem.

Proposition 5.3. *Let $S(\boldsymbol{P},\boldsymbol{q})$ be a function in some neighborhood U of a point $(\boldsymbol{P}_0,\boldsymbol{q}_0)\in\mathbb{R}^{2n}$ such that the $n\times n$ matrix*

$$\frac{\partial^2 S}{\partial \boldsymbol{P}\partial \boldsymbol{q}}(\boldsymbol{P}_0,\boldsymbol{q}_0) = \left\{\frac{\partial^2 S}{\partial P_i \partial q^j}(\boldsymbol{P}_0,\boldsymbol{q}_0)\right\}_{i,j=1}^n$$

is non-degenerate. Then S is a generating function of a local (i.e., defined in some neighborhood of $(\boldsymbol{P}_0,\boldsymbol{q}_0)$ in $\mathbb{R}^{2n}$) canonical transformation.

Suppose there is a canonical transformation $(\boldsymbol{P},\boldsymbol{Q}) = g(\boldsymbol{p},\boldsymbol{q})$ such that $H(\boldsymbol{p},\boldsymbol{q}) = K(\boldsymbol{P})$ for some function K. Then in the new coordinates Hamilton's equations take the form

$$\dot{\boldsymbol{P}} = 0, \quad \dot{\boldsymbol{Q}} = \frac{\partial K}{\partial \boldsymbol{P}}, \tag{5.9}$$

and are trivially integrated:

$$\boldsymbol{P}(t)=\boldsymbol{P}(0),\quad \boldsymbol{Q}(t)=\boldsymbol{Q}(0)+t\frac{\partial K}{\partial \boldsymbol{P}}(\boldsymbol{P}(0)).$$

Assuming that the matrix $\dfrac{\partial \boldsymbol{P}}{\partial \boldsymbol{p}}$ is non-degenerate, the generating function $S(\boldsymbol{P},\boldsymbol{q})$ satisfies the differential equation

$$H\Big(\frac{\partial S}{\partial \boldsymbol{q}}(\boldsymbol{P},\boldsymbol{q}),\boldsymbol{q}\Big)=K(\boldsymbol{P}), \tag{5.10}$$

where after the differentiation one should substitute $\boldsymbol{q}=\boldsymbol{q}(\boldsymbol{P},\boldsymbol{Q})$, defined by the canonical transformation g^{-1}. The differential equation (5.10) for fixed $\boldsymbol{P}$, as it follows from (5.2), coincides with the Hamilton-Jacobi equation for the abbreviated action $S_0=S-Et$ where $E=K(\boldsymbol{P})$,

$$H\Big(\frac{\partial S_0}{\partial \boldsymbol{q}}(\boldsymbol{P},\boldsymbol{q}),\boldsymbol{q}\Big)=E.$$

Theorem 5.7 (Jacobi). *Suppose that there is a function $S(\boldsymbol{P},\boldsymbol{q})$ which depends on n parameters $\boldsymbol{P}=(P_1,\dots,P_n)$, satisfies the Hamilton-Jacobi equation (5.10) for some function $K(\boldsymbol{P})$, and has the property that the $n\times n$ matrix $\dfrac{\partial^2 S}{\partial \boldsymbol{P}\partial \boldsymbol{q}}$ is non-degenerate. Then Hamilton's equations*

$$\dot{\boldsymbol{p}}=-\frac{\partial H}{\partial \boldsymbol{q}},\quad \dot{\boldsymbol{q}}=\frac{\partial H}{\partial \boldsymbol{p}}$$

can be solved explicitly, and the functions $\boldsymbol{P}(\boldsymbol{p},\boldsymbol{q})=(P_1(\boldsymbol{p},\boldsymbol{q}),\dots,P_n(\boldsymbol{p},\boldsymbol{q}))$, defined by the equations $\boldsymbol{p}=\dfrac{\partial S}{\partial \boldsymbol{q}}(\boldsymbol{P},\boldsymbol{q})$, are integrals of motion in involution.

Proof. Set $\boldsymbol{p}=\dfrac{\partial S}{\partial \boldsymbol{q}}(\boldsymbol{P},\boldsymbol{q})$ and $\boldsymbol{Q}=\dfrac{\partial S}{\partial \boldsymbol{P}}(\boldsymbol{P},\boldsymbol{q})$. By the inverse function theorem, $g(\boldsymbol{p},\boldsymbol{q})=(\boldsymbol{P},\boldsymbol{Q})$ is a local canonical transformation with the generating function S. It follows from (5.10) that $H(\boldsymbol{p}(\boldsymbol{P},\boldsymbol{Q}),\boldsymbol{q}(\boldsymbol{P},\boldsymbol{Q}))=K(\boldsymbol{P})$, so that Hamilton's equations take the form (5.9). Since $\omega=d\boldsymbol{P}\wedge d\boldsymbol{Q}$, integrals of motion $P_1(\boldsymbol{p},\boldsymbol{q}),\dots,P_n(\boldsymbol{p},\boldsymbol{q})$ are in involution. □

The solution of the Hamilton-Jacobi equation satisfying conditions in Theorem 5.7 is called the *complete integral*. At first glance it seems that solving the Hamilton-Jacobi equation, which is a nonlinear partial differential equation, is a more difficult problem then solving Hamilton's equations, which is a system of ordinary differential equations. It is quite remarkable that for many problems of classical mechanics one can find the complete integral of the Hamilton-Jacobi equation by the method of separation of variables. By Theorem 5.7, this solves the corresponding Hamilton's equations.

PROBLEM 5.1. Let $\pi : T^*M \to M$ be the canonical projection, and let $\mathscr{L}$ be a Lagrangian submanifold. Show that if the mapping $\pi|_{\mathscr{L}} : \mathscr{L} \to M$ is a diffeomorphism, then $\mathscr{L}$ is a graph of a smooth function on M. Give examples when for some $t > 0$ the corresponding projection of $g_t(\mathscr{L})$ onto M is no longer a diffeomorphism.

PROBLEM 5.2. Find the generating function for the identity transformation $\boldsymbol{P} = \boldsymbol{p}, \boldsymbol{Q} = \boldsymbol{q}$.

PROBLEM 5.3. Prove Proposition 5.3.

PROBLEM 5.4. Suppose that the canonical transformation $g(\boldsymbol{p}, \boldsymbol{q}) = (\boldsymbol{P}, \boldsymbol{Q})$ is such that locally $(\boldsymbol{Q}, \boldsymbol{q})$ can be considered as new coordinates (canonical transformations with this property are called *free*). Prove that $S_1(\boldsymbol{Q}, \boldsymbol{q}) = F(\boldsymbol{p}, \boldsymbol{q})$, also called a generating function, satisfies

$$\boldsymbol{p} = \frac{\partial S_1}{\partial \boldsymbol{q}} \quad \text{and} \quad \boldsymbol{P} = -\frac{\partial S_1}{\partial \boldsymbol{Q}}.$$

PROBLEM 5.5. Find the complete integral for the case of a particle in $\mathbb{R}^3$ moving in a central field.

LECTURE 6

Symplectic and Poisson manifolds

6.1. Symplectic manifolds

The notion of a symplectic manifold is a generalization of the example of a cotangent bundle T^*M.

Definition. A non-degenerate, closed 2-form ω on a manifold $\mathscr{M}$ is called a *symplectic form*, and the pair $(\mathscr{M},\omega)$ is called a *symplectic manifold*.

Since a symplectic form ω is non-degenerate, a symplectic manifold $\mathscr{M}$ is necessarily even-dimensional, $\dim\mathscr{M}=2n$. The nowhere vanishing $2n$-form ω^n defines a canonical orientation on $\mathscr{M}$, and as in the case $\mathscr{M}=T^*M$, $\dfrac{\omega^n}{n!}$ is called Liouville's volume form. We also have the general notion of a Lagrangian submanifold.

Definition. A submanifold $\mathscr{L}$ of a symplectic manifold $(\mathscr{M},\omega)$ is called a *Lagrangian submanifold*, if $\dim\mathscr{L}=\frac{1}{2}\dim\mathscr{M}$ and the restriction of the symplectic form ω to $\mathscr{L}$ is 0.

Besides cotangent bundles, another important class of symplectic manifolds is given by Kähler manifolds. The simplest compact Kähler manifold is $\mathbb{C}P^1\simeq S^2$ with the symplectic form given by the area 2-form of the Hermitian metric of Gaussian curvature 1 — the round metric on the 2-sphere. In terms of the local coordinate z associated with the stereographic projection $\mathbb{C}P^1\simeq\mathbb{C}\cup\{\infty\}$,

$$\omega=2i\,\frac{dz\wedge d\bar{z}}{(1+|z|^2)^2}.$$

Similarly, the natural symplectic form on the complex projective space $\mathbb{C}P^n$ is the symplectic form of the Fubini-Study metric. By pull-back, it defines symplectic forms on complex projective varieties.

The simplest non-compact Kähler manifold is the n-dimensional complex vector space $\mathbb{C}^n$ with the standard Hermitian metric. In complex coordinates $\boldsymbol{z}=(z^1,\dots,z^n)$ on $\mathbb{C}^n$ it is given by

$$h=d\boldsymbol{z}\otimes d\bar{\boldsymbol{z}}=\sum_{\alpha=1}^{n}dz^\alpha\otimes d\bar{z}^\alpha.$$

In terms of real coordinates $(\boldsymbol{x},\boldsymbol{y})=(x^1,\dots,x^n,y^1,\dots,y^n)$ on $\mathbb{R}^{2n}\simeq\mathbb{C}^n$, where $\boldsymbol{z}=\boldsymbol{x}+i\boldsymbol{y}$, the corresponding symplectic form $\omega=-\operatorname{Im}h$ has the canonical form

$$\omega = \frac{i}{2} d\boldsymbol{z} \wedge d\bar{\boldsymbol{z}} = \sum_{\alpha=1}^{n} dx^{\alpha} \wedge dy^{\alpha} = d\boldsymbol{x} \wedge d\boldsymbol{y}.$$

This example naturally leads to the following definition.

Definition. A *symplectic vector space* is a pair (V, ω), where V is a vector space over $\mathbb{R}$ and ω is a non-degenerate, skew-symmetric bilinear form on V.

It follows from basic linear algebra that every symplectic vector space V has a *symplectic basis* — a basis $e^1, \dots, e^n, f_1, \dots, f_n$ of V, where $2n = \dim V$, such that

$$\omega(e^i, e^j) = \omega(f_i, f_j) = 0 \quad \text{and} \quad \omega(e^i, f_j) = \delta^i_j, \quad i, j = 1, \dots, n.$$

In coordinates $(\boldsymbol{p}, \boldsymbol{q}) = (p_1, \dots, p_n, q^1, \dots, q^n)$ corresponding to this basis, $V \simeq \mathbb{R}^{2n}$ and

$$\omega = d\boldsymbol{p} \wedge d\boldsymbol{q} = \sum_{i=1}^{n} dp_i \wedge dq^i.$$

Thus every symplectic vector space is isomorphic to a direct product of the phase planes $\mathbb{R}^2$ with the canonical symplectic form $dp \wedge dq$. Introducing complex coordinates $\boldsymbol{z} = \boldsymbol{p} + i\boldsymbol{q}$, we get the isomorphism $V \simeq \mathbb{C}^n$, so that every symplectic vector space admits a Kähler structure.

It is a basic fact of symplectic geometry that every symplectic manifold is locally isomorphic to a symplectic vector space.

Theorem 6.1 (Darboux' theorem). *Let $(\mathscr{M}, \omega)$ be a $2n$-dimensional symplectic manifold. For every point $x \in \mathscr{M}$ there is a neighborhood U of x with local coordinates $(\boldsymbol{p}, \boldsymbol{q}) = (p_1, \dots, p_n, q^1, \dots, q^n)$ such that on U*

$$\omega = d\boldsymbol{p} \wedge d\boldsymbol{q} = \sum_{i=1}^{n} dp_i \wedge dq^i.$$

Coordinates $\boldsymbol{p}, \boldsymbol{q}$ are called *canonical coordinates* (*Darboux coordinates*). The proof proceeds by induction on n with the two main steps stated as Problems 6.1 and 6.2.

A non-degenerate 2-form ω for every $x \in \mathscr{M}$ defines an isomorphism $J : T^*_x\mathscr{M} \to T_x\mathscr{M}$ by

$$\omega(u_1, u_2) = J^{-1}(u_2)(u_1), \ u_1, u_2 \in T_x\mathscr{M}.$$

Explicitly, for every $X \in \mathrm{Vect}(\mathscr{M})$ and $\vartheta \in \mathcal{A}^1(\mathscr{M})$ we have

$$\omega(X, J(\vartheta)) = \vartheta(X) \quad \text{and} \quad J^{-1}(X) = -i_X(\omega).$$

Defining the *Hamiltonian vector field* associated with the function f by the formula $X_f = J(df)$ we have

$$df = -i_{X_f}(\omega), \tag{6.1}$$

cf. formulas (4.2)–(4.4). This proves the following result.

LEMMA 6.1. *A vector field X on $\mathscr{M}$ is a Hamiltonian vector field if and only if the 1-form $i_X(\omega)$ is exact.*

In local coordinates $\boldsymbol{x} = (x^1, \dots, x^{2n})$ for the coordinate chart (U, φ) on $\mathscr{M}$, the 2-form ω is given by

$$\omega = \tfrac{1}{2} \sum_{i,j=1}^{2n} \omega_{ij}(\boldsymbol{x})\, dx^i \wedge dx^j,$$

where $\{\omega_{ij}(\boldsymbol{x})\}_{i,j=1}^{2n}$ is a non-degenerate, skew-symmetric matrix-valued function on $\varphi(U)$. Denoting the inverse matrix by $\{\omega^{ij}(\boldsymbol{x})\}_{i,j=1}^{2n}$, we have

$$J(dx^i) = -\sum_{j=1}^{2n} \omega^{ij}(\boldsymbol{x}) \frac{\partial}{\partial x^j}, \quad i = 1, \dots, 2n.$$

DEFINITION. A *Hamiltonian system* is a pair consisting of a symplectic manifold $(\mathscr{M}, \omega)$, called a *phase space*, and a smooth real-valued function H on $\mathscr{M}$, called *Hamiltonian*. The motion of points on the phase space is described by the vector field

$$X_H = J(dH),$$

called a *Hamiltonian vector field*.

The trajectories of a Hamiltonian system $((\mathscr{M}, \omega), H)$ are the integral curves of a Hamiltonian vector field X_H on $\mathscr{M}$. In canonical coordinates $(\boldsymbol{p}, \boldsymbol{q})$ they are described by the canonical Hamilton's equations (4.1),

$$\dot{\boldsymbol{p}} = -\frac{\partial H}{\partial \boldsymbol{q}}, \quad \dot{\boldsymbol{q}} = \frac{\partial H}{\partial \boldsymbol{p}}.$$

Suppose now that the Hamiltonian vector field X_H on $\mathscr{M}$ is complete. The *Hamiltonian phase flow* on $\mathscr{M}$ associated with a Hamiltonian H is a one-parameter group $\{g_t\}_{t\in\mathbb{R}}$ of diffeomorphisms of $\mathscr{M}$ generated by X_H. The following statement generalizes Theorem 4.3.

THEOREM 6.2. *The Hamiltonian phase flow preserves the symplectic form.*

PROOF. It is sufficient to show that $\mathcal{L}_{X_H}\omega = 0$. Using Cartan's formula (1.1) and $d\omega = 0$, we get for every $X \in \mathrm{Vect}(\mathscr{M})$,

$$\mathcal{L}_X \omega = (d \circ i_X)(\omega),$$

and it follows from Lemma 6.1 that

$$\mathcal{L}_{X_H}\omega = -d(dH) = 0. \quad \square$$

DEFINITION. A vector field X on a symplectic manifold $(\mathscr{M}, \omega)$ is called a *symplectic vector field* if the 1-form $i_X(\omega)$ is closed, which is equivalent to the condition $\mathscr{L}_X\omega = 0$.

The commutative algebra $C^\infty(\mathscr{M})$, with a multiplication given by the pointwise product of functions, is called the *algebra of classical observables.* Assuming that the Hamiltonian phase flow g_t exists for all times, the time evolution of every observable $f \in C^\infty(\mathscr{M})$ is given by

$$f_t(x) = f(g_t(x)), \quad x \in \mathscr{M},$$

and is described by the differential equation

$$\frac{df_t}{dt} = X_H(f_t)$$

— Hamilton's equation for classical observables. Hamilton's equations for observables on $\mathscr{M}$ have the same form as Hamilton's equations on $\mathscr{M} = T^*M$, considered in Section 2.3. Since

$$X_H(f) = df(X_H) = \omega(X_H, J(df)) = \omega(X_H, X_f),$$

we have the following definition.

DEFINITION. A Poisson bracket on the algebra $C^\infty(\mathscr{M})$ of classical observables on a symplectic manifold $(\mathscr{M}, \omega)$ is a bilinear mapping $\{\ ,\ \} : C^\infty(\mathscr{M}) \times C^\infty(\mathscr{M}) \to C^\infty(\mathscr{M})$, defined by

$$\{f, g\} = \omega(X_f, X_g), \quad f, g \in C^\infty(\mathscr{M}).$$

Now Hamilton's equation takes the concise form

$$\frac{df}{dt} = \{H, f\}, \tag{6.2}$$

understood as a differential equation for a family of functions f_t on $\mathscr{M}$ with the initial condition $f_t|_{t=0} = f$. In local coordinates $\boldsymbol{x} = (x^1, \dots, x^{2n})$ on $\mathscr{M}$,

$$\{f, g\}(\boldsymbol{x}) = -\sum_{i,j=1}^{2n} \omega^{ij}(\boldsymbol{x}) \frac{\partial f(\boldsymbol{x})}{\partial x^i} \frac{\partial g(\boldsymbol{x})}{\partial x^j}.$$

THEOREM 6.3. *The Poisson bracket $\{\ ,\ \}$ on a symplectic manifold $(\mathscr{M}, \omega)$ is skew-symmetric and satisfies Leibniz rule and the Jacobi identity.*

PROOF. The first two properties are obvious. It follows from the definition of a Poisson bracket and the formula

$$[X_f, X_g](h) = (X_g X_f - X_f X_g)(h) = \{g, \{f, h\}\} - \{f, \{g, h\}\}$$

that the Jacobi identity is equivalent to the property

$$[X_f, X_g] = X_{\{f,g\}}. \tag{6.3}$$

Let X and Y be symplectic vector fields. Using Cartan's formulas (1.1)–(1.2) and (4.4), we get

$$\begin{aligned} i_{[X,Y]}(\omega) &= \mathcal{L}_X(i_Y(\omega)) - i_Y(\mathcal{L}_X(\omega)) \\ &= d(i_X \circ i_Y(\omega)) + i_X d(i_Y(\omega)) \\ &= d(\omega(Y,X)) = i_Z(\omega), \end{aligned}$$

where Z is a Hamiltonian vector field corresponding to $\omega(X,Y) \in C^\infty(\mathscr{M})$. Since the 2-form ω is non-degenerate, this implies $[X,Y] = Z$, so that setting $X = X_f, Y = X_g$ and using $\{f,g\} = \omega(X_f, X_g)$, we get (6.3). ☐

From (6.3) we immediately get the following result.

COROLLARY 6.4. *The subspace* $\mathrm{Ham}(\mathscr{M})$ *of Hamiltonian vector fields on* $\mathscr{M}$ *is a Lie subalgebra of* $\mathrm{Vect}(\mathscr{M})$*. The mapping* $C^\infty(\mathscr{M}) \to \mathrm{Ham}(\mathscr{M})$*, given by* $f \mapsto X_f$*, is a Lie algebra homomorphism with the kernel consisting of locally constant functions on* $\mathscr{M}$*.*

As in the case $\mathscr{M} = T^*M$ (see Section 5.3), an observable I — a function on the phase space $\mathscr{M}$ — is called an integral of motion (first integral) for the Hamiltonian system $((\mathscr{M},\omega), H)$ if it is constant along the Hamiltonian phase flow. According to (6.2), this is equivalent to the condition

$$\{H, I\} = 0. \tag{6.4}$$

It is said that the observables H and I are *in involution* (*Poisson commute*). From the Jacobi identity for the Poisson bracket we get the following result.

COROLLARY 6.5 (Poisson's theorem). *The Poisson bracket of two integrals of motion is an integral of motion.*

PROOF. If $\{H, I_1\} = \{H, I_2\} = 0$, then

$$\{H, \{I_1, I_2\}\} = \{\{H, I_1\}, I_2\} - \{\{H, I_2\}, I_1\} = 0. \qquad \square$$

It follows from Poisson's theorem that integrals of motion form a Lie algebra and, by (6.3), corresponding Hamiltonian vector fields form a Lie subalgebra in $\mathrm{Vect}(\mathscr{M})$. Since $\{I, H\} = dH(X_I) = 0$, the vector fields X_I are tangent to submanifolds $\mathscr{M}_E = \{x \in \mathscr{M} : H(x) = E\}$ — the level sets of the Hamiltonian H. This defines a Lie algebra of integrals of motion for the Hamiltonian system $((\mathscr{M},\omega), H)$ at the level set $\mathscr{M}_E$.

6.2. Poisson manifolds

The notion of a Poisson manifold generalizes the notion of a symplectic manifold.

DEFINITION. A *Poisson manifold* is a manifold $\mathscr{M}$ equipped with a *Poisson structure* — a skew-symmetric bilinear mapping

$$\{\ ,\ \} : C^\infty(\mathscr{M}) \times C^\infty(\mathscr{M}) \to C^\infty(\mathscr{M})$$

which satisfies the Leibniz rule and Jacobi identity.

Equivalently, $\mathscr{M}$ is a Poisson manifold if the algebra $\mathcal{A} = C^\infty(\mathscr{M})$ of classical observables is a Poisson algebra — a Lie algebra such that the Lie bracket is a bi-derivation with respect to the multiplication in $\mathcal{A}$ (a point-wise product of functions). It follows from the derivation property that in local coordinates $\boldsymbol{x} = (x^1, \dots, x^N)$ on $\mathscr{M}$, the Poisson bracket has the form

$$\{f, g\}(\boldsymbol{x}) = \sum_{i,j=1}^{N} \eta^{ij}(\boldsymbol{x}) \frac{\partial f(\boldsymbol{x})}{\partial x^i} \frac{\partial g(\boldsymbol{x})}{\partial x^j}.$$

The 2-tensor $\eta^{ij}(\boldsymbol{x})$, called a *Poisson tensor*, defines a global section η of the vector bundle $T\mathscr{M} \wedge T\mathscr{M}$ over $\mathscr{M}$.

The evolution of classical observables on a Poisson manifold is given by Hamilton's equations, which have the same form as (6.2),

$$\frac{df}{dt} = X_H(f) = \{H, f\}.$$

The phase flow g_t for a complete Hamiltonian vector field $X_H = \{H, \cdot\}$ defines the *evolution operator* $U_t : \mathcal{A} \to \mathcal{A}$ by

$$U_t(f)(x) = f(g_t(x)), \ f \in \mathcal{A}.$$

THEOREM 6.6. *Suppose that every Hamiltonian vector field on a Poisson manifold* $(\mathscr{M}, \{\ ,\ \})$ *is complete. Then for every* $H \in \mathcal{A}$*, the corresponding evolution operator* U_t *is an automorphism of the Poisson algebra* $\mathcal{A}$*, i.e.,*

$$U_t(\{f, g\}) = \{U_t(f), U_t(g)\} \quad \textit{for all} \quad f, g \in \mathcal{A}. \tag{6.5}$$

Conversely, if a skew-symmetric bilinear mapping $\{\ ,\ \} : C^\infty(\mathscr{M}) \times C^\infty(\mathscr{M}) \to C^\infty(\mathscr{M})$ *is such that* $X_H = \{H, \cdot\}$ *are complete vector fields for all* $H \in \mathcal{A}$*, and corresponding evolution operators* U_t *satisfy* (6.5)*, then* $(\mathscr{M}, \{\ ,\ \})$ *is a Poisson manifold.*

PROOF. Let $f_t = U_t(f)$, $g_t = U_t(g)$, and[1] $h_t = U_t(\{f, g\})$. By definition,

$$\frac{d}{dt}\{f_t, g_t\} = \{\{H, f_t\}, g_t\} + \{f_t, \{H, g_t\}\} \quad \text{and} \quad \frac{dh_t}{dt} = \{H, h_t\}.$$

If $(\mathscr{M}, \{\ ,\ \})$ is a Poisson manifold, then it follows from the Jacobi identity that

$$\{\{H, f_t\}, g_t\} + \{f_t, \{H, g_t\}\} = \{H, \{f_t, g_t\}\},$$

so that h_t and $\{f_t, g_t\}$ satisfy the same differential equation (6.2). Since these functions coincide at $t = 0$, (6.5) follows from the uniqueness theorem for the ordinary differential equations.

Conversely, we get the Jacobi identity for the functions f, g, and H by differentiating (6.5) with respect to t at $t = 0$. ∎

[1] Here g_t is not the phase flow!

COROLLARY 6.7. *A global section η of $T\mathscr{M}\wedge T\mathscr{M}$ is a Poisson tensor if and only if*

$$\mathcal{L}_{X_f}\eta = 0 \quad \textit{for all} \quad f \in \mathcal{A}.$$

DEFINITION. The *center* of a Poisson algebra $\mathcal{A}$ is

$$\mathcal{Z}(\mathcal{A}) = \{f \in \mathcal{A} : \{f, g\} = 0 \quad \text{for all} \quad g \in \mathcal{A}\}.$$

A Poisson manifold $(\mathscr{M}, \{\ ,\ \})$ is called *non-degenerate* if the center of a Poisson algebra of classical observables $\mathcal{A} = C^\infty(\mathscr{M})$ consists only of locally constant functions ($\mathcal{Z}(\mathcal{A}) = \mathbb{R}$ for connected $\mathscr{M}$).

Equivalently, a Poisson manifold $(\mathscr{M}, \{\ ,\ \})$ is non-degenerate if the Poisson tensor η is non-degenerate everywhere on $\mathscr{M}$, so that $\mathscr{M}$ is necessarily an even-dimensional manifold. A non-degenerate Poisson tensor for every $x \in \mathscr{M}$ defines an isomorphism $J : T_x^*\mathscr{M} \to T_x\mathscr{M}$ by

$$\eta(u_1, u_2) = u_2(J(u_1)),\ u_1, u_2 \in T_x^*\mathscr{M}.$$

In local coordinates $\boldsymbol{x} = (x^1, \dots, x^N)$ for the coordinate chart (U, φ) on $\mathscr{M}$, we have

$$J(dx^i) = \sum_{j=1}^N \eta^{ij}(\boldsymbol{x})\frac{\partial}{\partial x^j}, \quad i = 1, \dots, N.$$

A map $\varphi : \mathscr{M}_1 \to \mathscr{M}_2$ of Poisson manifolds $(\mathscr{M}_1, \{\ ,\ \}_1)$ and $(\mathscr{M}_2, \{\ ,\ \}_2)$ is called a *Poisson mapping*, if

$$\{f \circ \varphi, g \circ \varphi\}_1 = \{f, g\}_2 \circ \varphi \quad \text{for all} \quad f, g \in C^\infty(\mathscr{M}_2).$$

A symplectic manifold carries a natural Poisson structure. Its non-degeneracy follows from the non-degeneracy of a symplectic form. Converse statement also holds.

THEOREM 6.8. *A non-degenerate Poisson manifold is a symplectic manifold.*

PROOF. Let $(\mathscr{M}, \{\ ,\ \})$ be a non-degenerate Poisson manifold. Define the 2-form ω on $\mathscr{M}$ by

$$\omega(X, Y) = J^{-1}(Y)(X), \quad X, Y \in \mathrm{Vect}(\mathscr{M}),$$

where the isomorphism $J : T^*\mathscr{M} \to T\mathscr{M}$ is defined by the Poisson tensor η. In local coordinates $\boldsymbol{x} = (x^1, \dots, x^N)$ on $\mathscr{M}$,

$$\omega = -\sum_{1\le i<j\le N} \eta_{ij}(\boldsymbol{x})\, dx^i \wedge dx^j,$$

where $\{\eta_{ij}(\boldsymbol{x})\}_{i,j=1}^N$ is the inverse matrix to $\{\eta^{ij}(\boldsymbol{x})\}_{i,j=1}^N$. The 2-form ω is skew-symmetric and non-degenerate. For every $f \in \mathcal{A}$ let $X_f = \{f, \cdot\}$ be the corresponding vector field on $\mathscr{M}$. The Jacobi identity for the Poisson bracket $\{\ ,\ \}$ is equivalent to $\mathcal{L}_{X_f}\eta = 0$ for every $f \in \mathcal{A}$, so that

$$\mathcal{L}_{X_f}\omega = 0.$$

Since $X_f = Jdf$, we have $\omega(X, Jdf) = df(X)$ for every $X \in \mathrm{Vect}(\mathscr{M})$, so that

$$\omega(X_f, X_g) = \{f, g\}.$$

By Cartan's formula for the exterior differential,

$$\begin{aligned} d\omega(X,Y,Z) = \tfrac{1}{3}\,(&\mathcal{L}_X\omega(Y,Z) + \mathcal{L}_Y\omega(Z,X) + \mathcal{L}_Z\omega(X,Y) \\ &-\omega([X,Y],Z) - \omega([Z,X],Y) - \omega([Y,Z],X)), \end{aligned}$$

where $X, Y, Z \in \mathrm{Vect}(\mathscr{M})$. Now setting $X = X_f, Y = X_g, Z = X_h$, we get

$$\begin{aligned} d\omega(X_f, X_g, X_h) &= \tfrac{1}{3}\left(\omega(X_h,[X_f,X_g]) + \omega(X_f,[X_g,X_h]) + \omega(X_g,[X_h,X_f])\right) \\ &= \tfrac{1}{3}\left(\omega(X_h, X_{\{f,g\}}) + \omega(X_f, X_{\{g,h\}}) + \omega(X_g, X_{\{h,f\}})\right) \\ &= \tfrac{1}{3}\left(\{h,\{f,g\}\} + \{f,\{g,h\}\} + \{g,\{h,f\}\}\right) \\ &= 0. \end{aligned}$$

The exact 1-forms df, $f \in \mathcal{A}$, generate the vector space of 1-forms $\mathcal{A}^1(\mathscr{M})$ as a module over $\mathcal{A}$, so that Hamiltonian vector fields $X_f = Jdf$ generate the vector space $\mathrm{Vect}(\mathscr{M})$ as a module over $\mathcal{A}$. Thus $d\omega = 0$ and $(\mathscr{M}, \omega)$ is a symplectic manifold associated with the Poisson manifold $(\mathscr{M}, \{\ ,\ \})$. □

Remark. One can also prove this theorem by a straightforward computation in local coordinates $\boldsymbol{x} = (x^1, \dots, x^N)$ on $\mathscr{M}$. Just observe that the condition

$$\frac{\partial \eta_{ij}(\boldsymbol{x})}{\partial x^l} + \frac{\partial \eta_{jl}(\boldsymbol{x})}{\partial x^i} + \frac{\partial \eta_{li}(\boldsymbol{x})}{\partial x^j} = 0, \quad i, j, l = 1, \dots, N,$$

which is a coordinate form of $d\omega = 0$, follows from the condition

$$\sum_{j=1}^{N}\left(\eta^{ij}(\boldsymbol{x})\frac{\partial \eta^{kl}(\boldsymbol{x})}{\partial x^j} + \eta^{lj}(\boldsymbol{x})\frac{\partial \eta^{ik}(\boldsymbol{x})}{\partial x^j} + \eta^{kj}(\boldsymbol{x})\frac{\partial \eta^{li}(\boldsymbol{x})}{\partial x^j}\right) = 0,$$

which is a coordinate form of the Jacobi identity, by multiplying it three times by the inverse matrix $\eta_{ij}(\boldsymbol{x})$ using

$$\sum_{p=1}^{N}\left(\eta^{ip}(\boldsymbol{x})\frac{\partial \eta_{pk}(\boldsymbol{x})}{\partial x^j} + \frac{\partial \eta^{ip}(\boldsymbol{x})}{\partial x^j}\eta_{pk}(\boldsymbol{x})\right) = 0.$$

Remark. Let $\mathscr{M} = T^*\mathbb{R}^n$ with the Poisson bracket $\{\ ,\ \}$ given by the canonical symplectic form $\omega = d\boldsymbol{p} \wedge d\boldsymbol{q}$, where $(\boldsymbol{p}, \boldsymbol{q}) = (p_1, \dots, p_n, q^1, \dots, q^n)$ are coordinate functions on $T^*\mathbb{R}^n$. The non-degeneracy of the Poisson manifold $(T^*\mathbb{R}^n, \{\ ,\ \})$ can be formulated as the property that the only observable $f \in C^\infty(T^*\mathbb{R}^n)$ satisfying

$$\{f, p_1\} = \cdots = \{f, p_n\} = 0, \quad \{f, q^1\} = \cdots = \{f, q^n\} = 0$$

is $f(\boldsymbol{p}, \boldsymbol{q}) = \mathrm{const}$.

6.3. Noether theorem with symmetries

Let G be a finite-dimensional Lie group that acts on a connected symplectic manifold $(\mathscr{M}, \omega)$ by symplectomorphisms. The Lie algebra $\mathfrak{g}$ of G acts on $\mathscr{M}$ by vector fields

$$X_\xi(f)(x) = \frac{d}{ds}\bigg|_{s=0} f(e^{-s\xi} \cdot x),$$

and the linear mapping $\mathfrak{g} \ni \xi \mapsto X_\xi \in \mathrm{Vect}(\mathscr{M})$ is a homomorphism of Lie algebras,

$$[X_\xi, X_\eta] = X_{[\xi,\eta]}, \quad \xi, \eta \in \mathfrak{g}.$$

The G-action on M is called a *Hamiltonian action*, if X_ξ are Hamiltonian vector fields, i.e., for every $\xi \in \mathfrak{g}$ there is $\Phi_\xi \in C^\infty(\mathscr{M})$, defined up to an additive constant, such that $X_\xi = X_{\Phi_\xi} = J(d\Phi_\xi)$. Using (6.3), we see that for the Hamiltonian action

$$X_{\{\Phi_\xi, \Phi_\eta\}} = X_{\Phi_{[\xi,\eta]}},$$

so that

$$\{\Phi_\xi, \Phi_\eta\} = \Phi_{[\xi,\eta]} + c(\xi, \eta)$$

for some constants $c(\xi, \eta)$. The Hamiltonian action is called a *Poisson action* if there is a choice of functions Φ_ξ such that the linear mapping $\Phi : \mathfrak{g} \to C^\infty(\mathscr{M})$ is a homomorphism of Lie algebras,

$$\{\Phi_\xi, \Phi_\eta\} = \Phi_{[\xi,\eta]}, \quad \xi, \eta \in \mathfrak{g}. \tag{6.6}$$

Definition. A Lie group G is a *symmetry group* of the Hamiltonian system $((\mathscr{M}, \omega), H)$ if there is a Hamiltonian action of G on $\mathscr{M}$ such that

$$H(g \cdot x) = H(x), \quad g \in G, \ x \in \mathscr{M}.$$

Theorem 6.9 (Noether theorem with symmetries). *If G is a symmetry group of the Hamiltonian system $((\mathscr{M}, \omega), H)$, then the functions Φ_ξ, $\xi \in \mathfrak{g}$, are the integrals of motion. If the action of G is Poisson, the integrals of motion satisfy (6.6).*

Proof. By definition of the Hamiltonian action, for every $\xi \in \mathfrak{g}$,

$$0 = X_\xi(H) = X_{\Phi_\xi}(H) = \{\Phi_\xi, H\}. \qquad \square$$

Corollary 6.10. *Let (M, L) be a Lagrangian system such that the Legendre transform $\tau_L : TM \to T^*M$ is a diffeomorphism. Then if a Lie group G is a symmetry of (M, L), then G is a symmetry group of the corresponding Hamiltonian system $((T^*M, \omega), H = E_L \circ \tau_L^{-1})$, and the corresponding G-action on T^*M is Poisson. In particular, $\Phi_\xi = -I_\xi \circ \tau_L^{-1}$, where I_ξ are Noether integrals of motion for the one-parameter subgroups of G generated by $\xi \in \mathfrak{g}$.*

PROOF. Let X be the vector field associated with the one-parameter subgroup $\{e^{s\xi}\}_{s\in\mathbb{R}}$ of diffeomorphisms of M, used in Theorem 2.2, and let X' be its lift to TM. We have[2]

$$X_\xi = -(\tau_L)_*(X'), \tag{6.7}$$

and it follows from (2.3) that $\Phi_\xi = i_{X_\xi}(\theta) = \theta(X_\xi)$, where θ is the canonical Liouville 1-form on T^*M. From Cartan's formula (1.1) and formula $\mathcal{L}_{X'}(\theta_L) = 0$ (see Problem 2.4) we get

$$d\Phi_\xi = d(i_{X_\xi}(\theta)) = -i_{X_\xi}(d\theta) + \mathcal{L}_{X_\xi}(\theta) = -i_{X_\xi}(\omega).$$

It follows from (6.1) that $X_\xi = J(d\Phi_\xi)$ and the G-action is Hamiltonian. Using again the formula $\mathcal{L}_{X'}(\theta_L) = 0$ and Cartan's formula (1.2), we obtain

$$\begin{aligned}\Phi_{[\xi,\eta]} &= i_{[X_\xi,X_\eta]}(\theta) = \mathcal{L}_{X_\xi}(i_{X_\eta}(\theta)) + i_{X_\eta}(\mathcal{L}_{X_\xi}(\theta)) \\ &= X_\xi(\Phi_\eta) = \{\Phi_\xi, \Phi_\eta\}.\end{aligned}$$

□

EXAMPLE 6.1. The Lagrangian

$$L = \tfrac{1}{2}m\dot{\boldsymbol{r}}^2 - V(r)$$

for a particle in $\mathbb{R}^3$ moving in a central field (see Section 3.2) is invariant with respect to the action of the group SO(3) of orthogonal transformations of the Euclidean space $\mathbb{R}^3$. Let u_1, u_2, u_3 be a basis for the Lie algebra $\mathfrak{so}(3)$ corresponding to the rotations with the axes given by the vectors of the standard basis e_1, e_2, e_3 for $\mathbb{R}^3$ (see Example 2.2 in Section 2.2). These generators satisfy the commutation relations

$$[u_i, u_j] = \varepsilon_{ijk}u_k,$$

where $i, j, k = 1, 2, 3$, and ε_{ijk} is a totally anti-symmetric tensor, $\varepsilon_{123} = 1$. Corresponding Noether integrals of motion are given by $\Phi_{u_i} = -M_i$, where

$$\begin{aligned}M_1 &= (\boldsymbol{r}\times\boldsymbol{p})_1 = r_2p_3 - r_3p_2, \\ M_2 &= (\boldsymbol{r}\times\boldsymbol{p})_2 = r_3p_1 - r_1p_3, \\ M_3 &= (\boldsymbol{r}\times\boldsymbol{p})_3 = r_1p_2 - r_2p_1\end{aligned}$$

are components of the angular momentum vector $\boldsymbol{M} = \boldsymbol{r}\times\boldsymbol{p}$. (Here it is convenient to lower the indices of the coordinates r_i by the Euclidean metric on $\mathbb{R}^3$.) For the Hamiltonian

$$H = \frac{\boldsymbol{p}^2}{2m} + V(r)$$

we have

$$\{H, M_i\} = 0.$$

[2] The negative sign reflects the difference in definitions of X and X_ξ.

According to Theorem 6.9 and Corollary 6.10, Poisson brackets of the components of the angular momentum satisfy

$$\{M_i, M_j\} = -\varepsilon_{ijk} M_k,$$

which is also easy to verify directly using (5.5),

$$\{f, g\}(\boldsymbol{p}, \boldsymbol{r}) = \frac{\partial f}{\partial \boldsymbol{p}} \frac{\partial g}{\partial \boldsymbol{r}} - \frac{\partial f}{\partial \boldsymbol{r}} \frac{\partial g}{\partial \boldsymbol{p}}.$$

Example 6.2 (Kepler's problem). For every $\alpha \in \mathbb{R}$ the Lagrangian system on $\mathbb{R}^3$ with

$$L = \tfrac{1}{2} m \dot{\boldsymbol{r}}^2 + \frac{\alpha}{r}$$

has three extra integrals of motion — the components W_1, W_2, W_3 of the Laplace-Runge-Lenz vector, given by

$$\boldsymbol{W} = \frac{\boldsymbol{p}}{m} \times \boldsymbol{M} - \frac{\alpha \boldsymbol{r}}{r}$$

(see Section 3.3). Using Poisson brackets from the previous example, together with $\{r_i, M_j\} = -\varepsilon_{ijk} r_k$ and $\{p_i, M_j\} = -\varepsilon_{ijk} p_k$, we get by a straightforward computation,

$$\{W_i, M_j\} = -\varepsilon_{ijk} W_k \quad \text{and} \quad \{W_i, W_j\} = \frac{2H}{m} \varepsilon_{ijk} M_k,$$

where $H = \dfrac{\boldsymbol{p}^2}{2m} - \dfrac{\alpha}{r}$ is the Hamiltonian of Kepler's problem.

The Hamiltonian system $((\mathscr{M}, \omega), H)$, $\dim \mathscr{M} = 2n$, is called *completely integrable* if it has n independent integrals of motion $F_1 = H, \dots, F_n$ in involution. The former condition means that $dF_1(x), \dots, dF_n(x) \in T_x^* \mathscr{M}$ are linearly independent for almost all $x \in \mathscr{M}$. Hamiltonian systems with one degree of freedom such that dH has only finitely many zeros are completely integrable. Complete separation of variables in the Hamilton-Jacobi equation (see Section 5.4) provides other examples of completely integrable Hamiltonian systems.

Let $((\mathscr{M}, \omega), H)$ be a completely integrable Hamiltonian system. Suppose that the level set $\mathscr{M}_f = \{x \in \mathscr{M} : F_1(x) = f_1, \dots, F_n(x) = f_n\}$ is compact and tangent vectors $JdF_1, \dots, JdF_n$ are linearly independent for all $x \in \mathscr{M}_f$. Then by the Liouville-Arnold theorem, in a neighborhood of $\mathscr{M}_f$ there exist so-called *action-angle variables*: coordinates $\boldsymbol{I} = (I_1, \dots, I_n) \in \mathbb{R}_+^n = (\mathbb{R}_{>0})^n$ and $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_n) \in T^n = (\mathbb{R}/2\pi\mathbb{Z})^n$ such that $\omega = d\boldsymbol{I} \wedge d\boldsymbol{\varphi}$ and $H = H(I_1, \dots, I_n)$. According to Hamilton's equations,

$$\dot{I}_i = 0 \quad \text{and} \quad \dot{\varphi}_i = \omega_i = \frac{\partial H}{\partial I_i}, \quad i = 1, \dots, n,$$

so that action variables are constants, and angle variables change uniformly, $\varphi_i(t) = \varphi_i(0) + \omega_i t$, $i = 1, \dots, n$. The classical motion is almost-periodic with the frequencies $\omega_1, \dots, \omega_n$.

PROBLEM 6.1. Let $(\mathscr{M}, \omega)$ be a symplectic manifold. For $x \in \mathscr{M}$ choose a function q^1 on $\mathscr{M}$ such that $q^1(x) = 0$ and dq^1 does not vanish at x, and set $X = -X_{q^1}$. Show that there is a neighborhood U of $x \in \mathscr{M}$ and a function p_1 on U such that $X(q^1) = 1$ on U, and there exist coordinates $p_1, q^1, z^1, \dots, z^{2n-2}$ on U such that

$$X = \frac{\partial}{\partial p_1} \quad \text{and} \quad Y = X_{p_1} = \frac{\partial}{\partial q^1}.$$

PROBLEM 6.2. Continuing Problem 6.1, show that the 2-form $\omega - dp_1 \wedge dq^1$ on U depends only on coordinates $z^1, \dots, z^{2n-2}$ and is non-degenerate.

PROBLEM 6.3 (Coadjont orbits). Let G be a finite-dimensional Lie group, let $\mathfrak{g}$ be its Lie algebra, and let $\mathfrak{g}^*$ be the dual vector space to $\mathfrak{g}$. For $u \in \mathfrak{g}^*$ let $\mathscr{M} = \mathcal{O}_u$ be the orbit of u under the coadjoint action of G on $\mathfrak{g}^*$. Show that the formula

$$\omega(u_1, u_2) = u([x_1, x_2]),$$

where $u_1 = \mathrm{ad}^* x_1(u), u_2 = \mathrm{ad}^* x_2(u) \in T_u\mathscr{M}$, and ad^* stands for the coadjoint action of a Lie algebra $\mathfrak{g}$ on $\mathfrak{g}^*$, gives rise to a well-defined 2-form on $\mathscr{M}$, which is closed and non-degenerate. (The 2-form ω is called the *Kirillov-Kostant* symplectic form.)

PROBLEM 6.4 (Symplectic quotients). For a Poisson action of a Lie group G on a symplectic manifold $(\mathscr{M}, \omega)$, define the *moment map* $P : \mathscr{M} \to \mathfrak{g}^*$ by

$$P(x)(\xi) = \Phi_\xi(x), \ \xi \in \mathfrak{g}, \ x \in \mathscr{M},$$

where $\mathfrak{g}$ is the Lie algebra of G. For every $p \in \mathfrak{g}^*$ such that a stabilizer G_p of p acts freely and properly on $\mathscr{M}_p = P^{-1}(p)$ (such p is called *the regular value* of the moment map), the quotient $M_p = G_p \backslash \mathscr{M}_p$ is called a *reduced phase space*. Show that M_p is a symplectic manifold with the symplectic form uniquely characterized by the condition that its pull-back to $\mathscr{M}_p$ coincides with the restriction to $\mathscr{M}_p$ of the symplectic form ω.

PROBLEM 6.5 (Dual space to a Lie algebra). Let $\mathfrak{g}$ be a finite-dimensional Lie algebra with a Lie bracket $[\ ,\ ]$, and let $\mathfrak{g}^*$ be its dual space. For $f, g \in C^\infty(\mathfrak{g}^*)$ define

$$\{f, g\}(u) = u\left([df, dg]\right),$$

where $u \in \mathfrak{g}^*$ and $T_u^*\mathfrak{g}^* \simeq \mathfrak{g}$. Prove that $\{\ ,\ \}$ is a Poisson bracket. (It was introduced by Sophus Lie and is called a *linear*, or *Lie-Poisson* bracket.) Show that this bracket is degenerate and determine the center of $\mathcal{A} = C^\infty(\mathfrak{g}^*)$.

PROBLEM 6.6. A Poisson bracket $\{\ ,\ \}$ on $\mathscr{M}$ restricts to a Poisson bracket $\{\ ,\ \}_0$ on a submanifold $\mathscr{N}$, if the inclusion $\imath : \mathscr{N} \to \mathscr{M}$ is a Poisson mapping. Show that the Lie-Poisson bracket on $\mathfrak{g}^*$ restricts to a non-degenerate Poisson bracket on a coadjoint orbit, associated with the Kirillov-Kostant symplectic form.

PROBLEM 6.7. Do the computation in Example 6.2 and show that the Lie algebra of the integrals $M_1, M_2, M_3, W_1, W_2, W_3$ in Kepler's problem at $H(\boldsymbol{p}, \boldsymbol{r}) = E$ is isomorphic to the Lie algebra $\mathfrak{so}(4)$, if $E < 0$, to the Euclidean Lie algebra $\mathfrak{e}(3)$, if $E = 0$, and to the Lie algebra $\mathfrak{so}(1,3)$, if $E > 0$.

PROBLEM 6.8. Find the action-angle variables for a particle with one degree of freedom, when the potential $V(x)$ is a convex function on $\mathbb{R}$ satisfying $\lim_{|x|\to\infty} V(x) = \infty$. (*Hint:* Define $I = \oint p dx$, where integration goes over the closed orbit with $H(p, x) = E$.)

PROBLEM 6.9. Show that a Hamiltonian system describing a particle in $\mathbb{R}^3$ moving in a central field is completely integrable, and find the action-angle variables.

LECTURE 7

Hamiltonian systems with constraints

7.1. First order formalism

As in Lecture 1, consider Lagrangian system (M, L) with the Lagrangian function $L : TM \to \mathbb{R}$. If the Lagrangian L is non-degenerate, by doubling the number of degrees of freedom, we can replace (M, L) with another Lagrangian system $(\mathcal{M}, \mathcal{L})$, where the Lagrangian function $\mathcal{L} : T\mathcal{M} \to \mathbb{R}$ is linear in generalized velocities.

Namely, consider $\mathcal{M} = TM$ as new configuration space with generalized coordinates ξ and define the Lagrangian $\mathcal{L}$ on $T\mathcal{M}$ by the following formula

$$\mathcal{L}(\boldsymbol{\xi}, \dot{\boldsymbol{\xi}}) = \sum_{i=1}^{n} \frac{\partial L}{\partial v^i}(\dot{q}^i - v^i) + L(\boldsymbol{q}, \boldsymbol{v}) = \frac{\partial L}{\partial \boldsymbol{v}}(\dot{\boldsymbol{q}} - \boldsymbol{v}) + L(\boldsymbol{q}, \boldsymbol{v}). \tag{7.1}$$

Here $\boldsymbol{\xi} = (\boldsymbol{q}, \boldsymbol{v})$ are standard coordinates[1] on $\mathcal{M} = TM$, and $(\boldsymbol{\xi}, \dot{\boldsymbol{\xi}})$, where $\dot{\boldsymbol{\xi}} = (\dot{\boldsymbol{q}}, \dot{\boldsymbol{v}})$, are corresponding standard coordinates on $T\mathcal{M}$.

Lemma 7.1. *If Lagrangian function $L : TM \to \mathbb{R}$ is non-degenerate, Lagrangian systems (M, L) and $(\mathcal{M}, \mathcal{L})$ are equivalent — corresponding Euler-Lagrange equations coincide.*

Proof. The Euler-Lagrange equations

$$\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{\xi}}} - \frac{\partial \mathcal{L}}{\partial \boldsymbol{\xi}} = 0$$

for the Lagrangian system $(\mathcal{M}, \mathcal{L})$ reduce to

$$\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{q}}} - \frac{\partial \mathcal{L}}{\partial \boldsymbol{q}} = 0 \quad \text{and} \quad \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{v}}} - \frac{\partial \mathcal{L}}{\partial \boldsymbol{v}} = 0. \tag{7.2}$$

It follows from (7.1) that $\dfrac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{v}}} = 0$, and from the second equation in (7.2) we obtain

$$0 = \frac{\partial \mathcal{L}}{\partial \boldsymbol{v}} = \frac{\partial^2 L}{\partial \boldsymbol{v} \partial \boldsymbol{v}}(\dot{\boldsymbol{q}} - \boldsymbol{v}) - \frac{\partial L}{\partial \boldsymbol{v}} + \frac{\partial L}{\partial \boldsymbol{v}} = \frac{\partial^2 L}{\partial \boldsymbol{v} \partial \boldsymbol{v}}(\dot{\boldsymbol{q}} - \boldsymbol{v}).$$

Since Lagrangian L is non-degenerate, this implies

$$\dot{\boldsymbol{q}} = \boldsymbol{v}. \tag{7.3}$$

[1]To avoid confusion, here we do not denote standard coordinates on TM by $(\boldsymbol{q}, \dot{\boldsymbol{q}})$.

Using (7.1) we can rewrite the first equation in (7.2) as

$$0 = \frac{d}{dt}\frac{\partial L}{\partial \boldsymbol{v}} - \frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{q}\partial \boldsymbol{v}}(\dot{\boldsymbol{q}} - \boldsymbol{v}) - \frac{\partial L}{\partial \boldsymbol{q}}.$$

Using (7.3), we obtain Euler-Lagrange equations for the Lagrangian system (M, L),

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{\boldsymbol{q}}} - \frac{\partial L}{\partial \boldsymbol{q}} = 0. \quad \square$$

In general, Lagrangian system $(\mathcal{M}, \mathcal{L})$ in the *first order formalism* is defined by the Lagrangian function $\mathcal{L} : T\mathcal{M} \to \mathbb{R}$, which in standard coordinates on $T\mathcal{M}$ is given by

$$\mathcal{L}(\boldsymbol{\xi}, \dot{\boldsymbol{\xi}}) = \sum_{\alpha=1}^{N} f_\alpha(\boldsymbol{\xi})\dot{\xi}^\alpha - H(\boldsymbol{\xi}), \tag{7.4}$$

where H is a function on $\mathcal{M}$ and $N = \dim \mathcal{M}$. It is also said that Lagrangian $\mathcal{L}$ is *linear in generalized velocities*. It is associated with the 1-form $\vartheta_{\mathcal{L}}$ on $\mathcal{M} \times \mathbb{R}$,

$$\vartheta_{\mathcal{L}} = \sum_{\alpha=1}^{N} f_\alpha(\boldsymbol{\xi}) d\xi^\alpha - H(\boldsymbol{\xi}) dt. \tag{7.5}$$

It has the property that for every path $\gamma : [t_0, t_1] : \mathcal{M} \to \mathbb{R}$,

$$\int_{t_0}^{t_1} \mathcal{L}(\gamma'(t)) dt = \int_\sigma \vartheta_{\mathcal{L}},$$

where $\gamma'(t)$ is a vertical lift of $\gamma(t)$ to $T\mathcal{M}$ and $\sigma = \{(\gamma(t), t); t_0 \leq t \leq t_1\}$ is a 1-chain on $\mathcal{M} \times \mathbb{R}$ (cf. Problem 1.1 in Lecture 1).

Remark. In case when $\mathcal{M} = TM$ and Lagrangian $\mathcal{L}$ is given by (7.1),

$$\vartheta_{\mathcal{L}} = \theta_L - E dt,$$

where $\theta_L = \dfrac{\partial L}{\partial \boldsymbol{v}} d\boldsymbol{q}$ is the 1-form associated with the Lagrangian $L : TM \to \mathbb{R}$ (see formula (2.1) in Lecture 2), and $E = \dfrac{\partial L}{\partial \boldsymbol{v}}\boldsymbol{v} - L$ is the corresponding energy (see Sect. 2.1 in Lecture 2).

Definition. Lagrangian $\mathcal{L}$, given by (7.4), is called non-degenerate, if the 2-form

$$\omega = d\left(\sum_{\alpha=1}^{N} f_\alpha(\boldsymbol{\xi}) d\xi^\alpha\right) = \sum_{\alpha,\beta=1}^{N} \frac{\partial f_\beta}{\partial \boldsymbol{x}_\alpha}(\boldsymbol{\xi}) d\xi^\alpha \wedge d\xi^\beta$$

is non-degenerate on $\mathcal{M}$.

It follows from the previous remark and Problem 2.1 in Lecture 2, that for Lagrangians (7.1) this definition agrees with the one given in Lecture 1. If the Lagrangian $\mathcal{L}$ is non-degenerate, it follows from the Darboux theorem that $N = 2n$ is even and there exist local canonical coordinates $(\boldsymbol{p}, \boldsymbol{q}) = (p_1, \dots, p_n, q^1, \dots, q^n)$ on $\mathcal{M}$ such that

$$\vartheta_{\mathcal{L}} = \boldsymbol{p}d\boldsymbol{q} - H(\boldsymbol{p}, \boldsymbol{q})dt$$

and the Euler-Lagrange equations for the Lagrangian

$$\mathcal{L} = \boldsymbol{p}\dot{\boldsymbol{q}} - H(\boldsymbol{p}, \boldsymbol{q})$$

are Hamilton's equations

$$\dot{\boldsymbol{p}} = -\frac{\partial H}{\partial \boldsymbol{q}}, \quad \dot{\boldsymbol{q}} = \frac{\partial H}{\partial \boldsymbol{p}}$$

with the Hamiltonian function $H(\boldsymbol{p}, \boldsymbol{q})$. This repeats derivation of the Hamilton's equations given in Sect. 4.2 in Lecture 4, but without explicitly using Legendre transform.

Remark. In this case we trivially have

$$\boldsymbol{p} = \frac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{q}}} \quad \text{and} \quad H = \boldsymbol{p}\dot{\boldsymbol{q}} - \mathcal{L}.$$

7.2. Singular Lagrangians

Here we consider important case when Lagrangian (7.4) is singular. Darboux theorem is still applicable and guarantees existence of local coordinates $(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{\lambda}) = (p_1, \dots, p_n, q^1, \dots, q^n, \lambda_1, \dots, \lambda_m)$ on $\mathcal{M}$, where $N = 2n + m$, such that

$$\vartheta_{\mathcal{L}} = \boldsymbol{p}d\boldsymbol{q} - H(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{\lambda})dt + dS$$

for some (local) function $S(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{\lambda})$. Since addition of the exact form does not change equations of motion (see Problem 1.2 in Lecture 1), the Euler-Lagrange equations for the Lagrangian $\mathcal{L}$ have the following form

$$\dot{\boldsymbol{p}} = -\frac{\partial H}{\partial \boldsymbol{q}}, \quad \dot{\boldsymbol{q}} = \frac{\partial H}{\partial \boldsymbol{p}} \quad \text{and} \quad \frac{\partial H}{\partial \boldsymbol{\lambda}} = 0. \tag{7.6}$$

Now suppose that the $m \times m$ matrix $\left\{ \dfrac{\partial^2 H}{\partial \lambda_a \partial \lambda_b} \right\}_{a,b=1}^m$ has constant rank k on $\mathcal{M}$. If $k = m$, it follows from the implicit function theorem that the equations $\dfrac{\partial H}{\partial \boldsymbol{\lambda}} = 0$ in (7.6) determine a submanifold $\tilde{\mathcal{M}}$ in $\mathcal{M}$ of dimension $N - m = 2n$, given by the equations $\lambda_a = \lambda_a(\boldsymbol{p}, \boldsymbol{q})$, $a = 1, \dots, m$. In other words, in this case it is possible to exclude coordinates $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)$. Putting

$$\tilde{H}(\boldsymbol{p}, \boldsymbol{q}) = H(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{\lambda}(\boldsymbol{p}, \boldsymbol{q})) \quad \text{and} \quad \tilde{\mathcal{L}} = \boldsymbol{p}\dot{\boldsymbol{q}} - \tilde{H}(\boldsymbol{p}, \boldsymbol{q}),$$

we obtain a non-degenerate Lagrangian system $(\tilde{\mathcal{M}}, \tilde{\mathcal{L}})$ whose Euler-Lagrange equations coincide with the restriction to $\tilde{\mathcal{M}}$ of the first two equations in (7.6). Indeed, we have

$$\begin{aligned} \frac{\partial \tilde{H}}{\partial \boldsymbol{p}} &= \left(\frac{\partial H}{\partial \boldsymbol{p}} + \frac{\partial H}{\partial \boldsymbol{\lambda}}\frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{p}}\right)\bigg|_{\tilde{\mathcal{M}}} = \frac{\partial H}{\partial \boldsymbol{p}}\bigg|_{\tilde{\mathcal{M}}}, \\ \frac{\partial \tilde{H}}{\partial \boldsymbol{q}} &= \left(\frac{\partial H}{\partial \boldsymbol{q}} + \frac{\partial H}{\partial \boldsymbol{\lambda}}\frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{q}}\right)\bigg|_{\tilde{\mathcal{M}}} = \frac{\partial H}{\partial \boldsymbol{q}}\bigg|_{\tilde{\mathcal{M}}}. \end{aligned}$$

Correspondingly, $\tilde{\mathcal{M}}$ is a symplectic manifold with the symplectic form $d\boldsymbol{p}\wedge d\boldsymbol{q}$ and Hamiltonian $\tilde{H}(\boldsymbol{p},\boldsymbol{q})$, and Euler-Lagrange equations (7.6), restricted to $\tilde{\mathcal{M}}$, become Hamilton's equations.

In case $k < m$, by using appropriate change of coordinates $\boldsymbol{\lambda}$, we can exclude the first k coordinates $\lambda_1, \dots, \lambda_k$ (such coordinates are called *excludable*), while remaining $m-k$ coordinates $\lambda_{k+1}, \dots, \lambda_m$ satisfy

$$\frac{\partial^2 H}{\partial \lambda_a \partial \lambda_b} = 0, \quad a, b = k+1, \dots, m,$$

so that H is linear function of $\lambda_{k+1}, \dots, \lambda_m$. Thus from the very beginning we can assume that $\mathcal{M} = \mathcal{M}_0 \times \mathbb{R}^m$, where $\mathcal{M}_0$ has canonical coordinates $(\boldsymbol{p},\boldsymbol{q})$ and symplectic form $\omega = d\boldsymbol{p}\wedge d\boldsymbol{q}$, and consider singular Lagrangians on $T\mathcal{M}$ of the form

$$\mathcal{L} = \boldsymbol{p}\dot{\boldsymbol{q}} - H(\boldsymbol{p},\boldsymbol{q}) - \sum_{a=1}^{m} \lambda_a \varphi^a(\boldsymbol{p},\boldsymbol{q}), \tag{7.7}$$

where coordinates $\boldsymbol{\lambda}$ play the role of Lagrange multipliers. The Euler-Lagrange equations are

$$\dot{\boldsymbol{p}} = -\frac{\partial H}{\partial \boldsymbol{q}} - \sum_{a=1}^{m} \lambda_a \frac{\partial \varphi^a}{\partial \boldsymbol{q}}, \tag{7.8}$$

$$\dot{\boldsymbol{q}} = \frac{\partial H}{\partial \boldsymbol{p}} + \sum_{a=1}^{m} \lambda_a \frac{\partial \varphi^a}{\partial \boldsymbol{p}}, \tag{7.9}$$

and

$$\varphi^a(\boldsymbol{p},\boldsymbol{q}) = 0, \quad a = 1, \dots, m. \tag{7.10}$$

The functions $\varphi^a(\boldsymbol{p},\boldsymbol{q})$ are called *constraints* and equations (7.10) determine a subset $\mathcal{M}_0 \subset \mathcal{M}$. In case when $m \times 2n$ matrix $\left(\dfrac{\partial \varphi^a}{\partial p_i}, \dfrac{\partial \varphi^a}{\partial q^i}\right)$ has rank m on $\mathcal{M}_0$, the set $\mathcal{M}_0$ is a submanifold of $\mathcal{M}$ of dimension $2n - m$. Restricting the 1-form $\vartheta_{\mathcal{L}}$ on $\mathcal{M}_0 \times \mathbb{R}$ and using Darboux theorem, we either obtain a non-degenerate Lagrangian system that corresponds to the Hamiltonian system, or we get a singular Lagrangian. In that case we repeat the above procedure and obtain a Lagrangian as in (7.7), and corresponding constraints determine a submanifold $\mathcal{M}_1$ of $\mathcal{M}_0$. Iterating this procedure, in finitely many steps we obtain a non-degenerate Lagrangian.

At each step of this procedure one needs to solve equations (7.10) in order to apply the Darboux theorem to the restriction of the 1-form $\vartheta_{\mathcal{L}}$ to the submanifold $\mathcal{M}_0$. This could be a very difficult problem. However, there is important class of constraints for which one does need to solve equations (7.10).

7.3. First class constraints and reduced phase space

It seems natural to identify functions on $\mathcal{M}$ with the same restrictions on $\mathcal{M}_0$. Namely, let $\mathcal{I}$ be an ideal in the algebra $\mathcal{A} = C^\infty(\mathcal{M})$, consisting of functions that vanish on $\mathcal{M}_0$. Condition that the matrix $\left(\dfrac{\partial\varphi^a}{\partial p_i}, \dfrac{\partial\varphi^a}{\partial q^i}\right)$ has constant rank m on $\mathcal{M}_0$ leads to the following result.

Lemma 7.2. *Ideal $\mathcal{I}$ is generated by the constraints $\varphi^1, \dots, \varphi^m$.*

However, there are two questions one needs to address in order to formulated consistent dynamics for Hamiltonian systems with constraints.

- Whether trajectories $(\boldsymbol{p}(t), \boldsymbol{q}(t))$ of the Hamilton's equations (7.8)–(7.9) lie on $\mathcal{M}_0$ if $(\boldsymbol{p}(0), \boldsymbol{q}(0)) \in \mathcal{M}_0$.
- Describing the algebra of observables whose evolution does not depend on the arbitrary parameters $\lambda_1, \dots, \lambda_m$ in (7.8)–(7.9).

It is remarkable that, according to Dirac, the affirmative answer to both of these questions is obtained by using the following definition. Let $\{\ ,\ \}$ be the Poisson bracket on $\mathcal{M}$ associated with the symplectic form $d\boldsymbol{p} \wedge d\boldsymbol{q}$.

Definition. Constraints $\varphi^1, \dots, \varphi^m$ for the singular Lagrangian (7.7) are called *first class constraints* if $\{\varphi^a, \varphi^b\}, \{H, \varphi^a\} \in \mathcal{I}$, $a, b = 1, \dots, m$.

In other words, there are functions g_c^{ab} and h_b^a on $\mathcal{M}$ such that

$$\{\varphi^a, \varphi^b\} = \sum_{c=1}^m g_c^{ab}\varphi^c \quad \text{and} \quad \{H, \varphi^a\} = \sum_{b=1}^m h_b^a \varphi^b. \tag{7.11}$$

Lemma 7.3. *For the first class constraints trajectories $(\boldsymbol{p}(t), \boldsymbol{q}(t))$ of the Hamilton's equations (7.8)–(7.9) lie on $\mathcal{M}_0$ if $(\boldsymbol{p}(0), \boldsymbol{q}(0)) \in \mathcal{M}_0$.*

Proof. It follows from (7.8)–(7.9) that

$$\dot{\varphi}^a = \{H, \varphi^a\} + \sum_{b=1}^m \lambda_b\{\varphi^b, \varphi^a\},$$

and it follows from (7.11) that $\dot{\varphi}^a = 0$ on $\mathcal{M}_0$. Thus $\varphi^a(\boldsymbol{p}(t), \boldsymbol{q}(t)) = \varphi^a(\boldsymbol{p}(0), \boldsymbol{q}(0)) = 0$, $a = 1, \dots, m$. $\square$

In general, according to (7.8)–(7.9), the evolution of arbitrary $f \in \mathcal{A}$ is given by

$$\dot{f} = \{H, f\} + \sum_{a=1}^{m} \lambda_a \{\varphi^a, f\}, \tag{7.12}$$

and it follows from (7.11) that restriction of this equation to $\mathcal{M}_0$ does not depend on the choice of a representative in $f \bmod \mathcal{I}$ and defines the evolution in the algebra $\mathcal{A}/\mathcal{I}$. Still, this evolution depend on the choice of arbitrary parameters $\lambda_1, \dots, \lambda_m$.

DEFINITION. Admissible observables are functions f^* on $\mathcal{M}_0$ whose extensions f to $\mathcal{M}$ satisfy

$$\{f, \varphi^a\}|_{\mathcal{M}_0} = 0, \quad a = 1, \dots, m. \tag{7.13}$$

In particular, $H^* = H|_{\mathcal{M}_0}$ is an admissible observable. It follows from Lemma 7.2 and (7.11) that (7.13) is valid for any smooth extension f of a function f^* on $\mathcal{M}_0$. The Poisson bracket of admissible observables is defined by

$$\{f^*, g^*\}_0 = \{f, g\}|_{\mathcal{M}_0},$$

and it follows from (7.13) that admissible observables form a Poisson algebra $\mathcal{A}^*$. For admissible observables equation (7.13) takes the form

$$\dot{f}^* = \{H^*, f^*\}_0 \tag{7.14}$$

and no longer depends on on the choice of arbitrary parameters $\lambda_1, \dots, \lambda_m$.

Put

$$\mathcal{A}_0 = \{f \in \mathcal{A} : \{f, \varphi^a\}|_{\mathcal{M}_0} = 0, \quad a = 1, \dots, m.$$

LEMMA 7.4. *$\mathcal{A}_0$ is a Poisson subalgebra of $\mathcal{A}$: if $f, g \in \mathcal{A}_0$, then $fg \in \mathcal{A}_0$ and $\{f, g\} \in \mathcal{A}_0$. Moreover, $\mathcal{I} \subset \mathcal{A}_0$ is a Poisson algebra ideal of $\mathcal{A}_0$ and*

$$\mathcal{A}_0/\mathcal{I} \simeq \mathcal{A}^*.$$

PROOF. Follows from Lemma 7.2, equations (7.11) and Jacobi identity. □

The functions $f^* \in \mathcal{A}^*$ depend on $2n - m - m = 2(n - m)$ and in many cases can be thought of as functions on the *reduced phase space* — symplectic manifold Γ of dimension $2n - 2m$. This can be described geometrically as follows. Let $X_{\varphi^a} \in \mathrm{Vect}(\mathcal{M})$ be the Hamiltonian vector fields corresponding to the functions φ^a on $\mathcal{M}$. We have, according to formula (6.3) in Lecture 6,

$$[X_{\varphi^a}, X_{\varphi^b}] = X_{\{\varphi^a, \varphi^b\}}, \quad a, b = 1, \dots, m. \tag{7.15}$$

We also have

$$\omega(X_{\varphi^a}, X_{\varphi^b}) = \{\varphi^a, \varphi^b\},$$

so that

$$\omega_m(X_{\varphi^a}, X_{\varphi^b}) = 0 \quad \text{for all} \quad m \in \mathcal{M}_0.$$

Denote by Y_a the vector vector fields X_{φ^a} along $\mathcal{M}_0$. It follows from (7.10) that Y_a are tangent to $\mathcal{M}_0$ and

$$\omega|_{\mathcal{M}_0}(Y_a, Y_b) = 0, \quad a, b = 1, \dots, m.$$

Thus the closed 2-form ω_0 — a restriction of the symplectic form ω to $\mathcal{M}_0$ — has an m-dimensional kernel, generated by the vector fields $Y_a \in \mathrm{Vect}(\mathcal{M}_0)$. It follows from (7.11) and (7.15) that

$$[X_{\varphi^a}, X_{\varphi^b}] = \sum_{c=1}^{m} g_c^{ab} X_{\varphi^c},$$

This means that the vector fields $Y_1, \dots, Y_m$ generated a smooth involutive distribution on $\mathcal{M}_0$ — a subbundle $\mathcal{P}$ of the tangent bundle $T\mathcal{M}_0$, such that $[X, Y] \in \mathcal{P}$ if $X, Y \in \mathcal{P}$. By the Frobenius theorem, $\mathcal{M}_0$ is a foliation with m-dimensional leaves given by the integral manifolds of the distribution $\mathcal{P}$.

In case this foliation is a fibration with the base $\mathcal{M}^*$, a $2n-2m$ dimensional submanifold of $\mathcal{M}_0$, we have $\mathcal{A}^* = C^\infty(\mathcal{M}^*)$ and the closed 2-form ω^* — a restriction of the 2-form ω_0 to $\mathcal{M}^*$ is non-degenerate! Indeed, equations (7.13) imply that the functions f are constant along the fibers. Locally, $\mathcal{M}^*$ can be defined by the equations

$$\chi_a(\boldsymbol{p}, \boldsymbol{q}) = 0, \quad a = 1, \dots, m, \tag{7.16}$$

called *additional constraints*. Condition

$$\det\left(\{\chi_a, \varphi^b\}\right)_{a,b=1}^{m} \neq 0 \tag{7.17}$$

guarantees that the submanifold of $\mathcal{M}_0$, defined by (7.16), intersects transversally the integral manifolds of the distribution $\mathcal{P}$. If intersection of every integral manifold with this submanifold consists of only one point, equations (7.10) and (7.16) in $\mathcal{M}$ determine the reduced phase space — a $2n-2m$ manifold $\mathcal{M}^*$ with the symplectic form $\omega^* = \omega|_{\mathcal{M}^*}$.

In special case when

$$\{\chi_a, \chi_b\} = 0, \quad a, b = 1, \dots, m, \tag{7.18}$$

one can easily find canonical coordinates on $\mathcal{M}^*$. Indeed, put $p_a = \chi_a$, $a = 1, \dots, m$. By Darboux theorem, there are coordinates q^a and

$$(\boldsymbol{p}^*, \boldsymbol{q}^*) = \left(p_1^*, \dots, p_{2n-2m}^*, (q^*)^1, \dots, (q^*)^{2n-2m}\right)$$

such that

$$\omega = \sum_{a=1}^{m} dp_a \wedge dq^a + d\boldsymbol{p}^* \wedge d\boldsymbol{q}^*.$$

Transversality condition (7.17) becomes

$$\det\left(\frac{\partial\varphi^a}{\partial q^b}\right)_{a,b=1}^m \neq 0,$$

so that $q^a = q^a(\boldsymbol{p}^*, \boldsymbol{q}^*)$. Thus the reduced phase space $\mathcal{M}^*$ is given by the equations

$$p_a = 0, \quad q^a = q^a(\boldsymbol{p}^*, \boldsymbol{q}^*), \quad a = 1, \dots, m,$$

and

$$\omega^* = d\boldsymbol{p}^* \wedge d\boldsymbol{q}^*.$$

We also have $f^*(\boldsymbol{p}^*, \boldsymbol{q}^*) = f(0, \boldsymbol{p}^*, q^a(\boldsymbol{p}^*, \boldsymbol{q}^*), \boldsymbol{q}^*)$ and

$$\{f^*, g^*\} = \frac{\partial f^*}{\partial \boldsymbol{p}^*}\frac{\partial g^*}{\partial \boldsymbol{q}^*} - \frac{\partial f^*}{\partial \boldsymbol{q}^*}\frac{\partial g^*}{\partial \boldsymbol{p}^*}. \tag{7.19}$$

7.4. Second class constraints and Dirac bracket

Constraints (7.10), for which

$$\det\left(\{\varphi^a, \varphi^b\}\right) \neq 0,$$

are called the *second class constraints*. In this case m is necessarily even, $m = 2k$, and the submanifold $\mathcal{M}_0$, determined by equations (7.10) is a symplectic manifold with a symplectic form $\omega_0 = \omega|_{\mathcal{M}_0}$. In this case Poisson bracket $\{\ ,\ \}_0$ corresponding to ω_0 is obtained by the following construction. Let C_{ab} be the inverse matrix to $(\{\varphi^a, \varphi^b\})$, and let $\{\ ,\ \}$ be a Poisson bracket on $\mathcal{M}$ associated with the symplectic form ω.

Definition. *Dirac bracket* $\{\ ,\ \}_{\mathrm{DB}}$ on $\mathcal{M}$ is given by the following formula

$$\{f, g\}_{\mathrm{DB}} = \{f, g\} - \sum_{a,b=1}^{2k} \{f, \varphi^a\} C_{ab} \{\varphi^b, g\}. \tag{7.20}$$

It follows from this definition, that for all $f \in \mathcal{A}$,

$$\{f, \varphi^a\}_{\mathrm{DB}} = 0, \quad a = 1, \dots, 2k. \tag{7.21}$$

Lemma 7.5. *Dirac bracket is a degenerate Poisson bracket on $\mathcal{M}$ whose center consists of the functions $F(\varphi^1, \dots, \varphi^{2k})$, where $F : \mathbb{R}^{2k} \to \mathbb{R}$. Moreover, Dirac bracket restricts to $\mathcal{M}_0$ as a non-degenerate Poisson bracket that corresponds to the symplectic form ω_0.*

It follows from the transversality condition (7.17), that the first class constraints φ^a and additional constraints χ_a can be combined iinto the second class constraints $\varphi^1, \dots, \varphi^m, \chi_1, \dots, \chi_m$. Lemma 7.5 implies

Corollary 7.1. *Poisson bracket on the reduced phase space $\mathcal{M}^*$ for the first class constraints coincides with the Dirac bracket for the associated set of the second class constraints.*

PROBLEM 7.1. Prove that formula (7.1) gives a well defined function $\mathcal{L}$ on $T\mathcal{M}$.

PROBLEM 7.2. Prove Lemma 7.2.

PROBLEM 7.3. Prove that the symplectic quotient construction (see Problem 6.4 in Lecture 6) in case $p = 0$ is a particular case of the Dirac formalism, where φ^a are the Hamiltonian functions of the Hamiltonian vector fields X_{ξ^a} associated with a basis ξ^a of the Lie algebra $\mathfrak{g}$.

PROBLEM 7.4. Prove (7.19) by computing Poisson bracket $\{f, g\}$ on $\mathcal{M}$ in coordinates $\eta = (p_a, \boldsymbol{p}^*, \varphi^a, \boldsymbol{q}^*)$.

PROBLEM 7.5. Prove Lemma 7.5.

PROBLEM 7.6. Prove Corollary 7.1.

Notes and references

In addition to references in §3 in Chapter 1 in the book

- Leon A. Takhtajan, *Quantum Mechanics for Mathematicians*, GSM **95**, Amer. Math. Soc., Providence, RI, 2008,

there are the following monographs and lecture notes

- Jerrold E. Marsden and Tudor S. Ratiu, *Introduction to Mechanics and Symmetry*, 2nd Edition, Springer, 2002.

- Juan-Pablo Ortega and Tudor S. Ratiu, *Momentum Maps and Hamiltonian Reduction*, Birkhäuser, 2004.

- Alan Wienstein, *Lectures on symplectic manifolds*, Amer. Math. Soc., Providence, RI, 1977.

The paper

- J.-M. Lévy-Leblond, *Conservation Laws for Gauge-Variant Lagrangians in Classical Mechanics*, Am. J. Phys. **39** (1971), 502-506,

discusses the generalized Noether theorem and Laplace-Runge-Lenz vector. More details on Kepler problem can be found in Milnor's paper

- John Milnor, *On the geometry of Kepler problem*, The Amer. Math. Monthly, **90** (1983), 353-365

and in the book

- Victor Guiliemin and Shlomo Sternberg, *Variations on a Theme by Kepler*, Amer. Math. Soc., Providence, RI, 1990.

Symplectic quotients (Marsden-Weinstein quotients) construction was introduced in the paper

- J. E. Marsden and A. Weinstein, *Reduction of symplectic manifolds with symmetry*, Reports on Mathematical Physics, **5**(1) (1974), 121–130.

Generalized Hamiltonian dynamics for singular Lagrangians was developed by Dirac in the classics paper

• P.A.M. Dirac, *Generalized Hamiltonian Dynamics*, Proc. R. Soc. Lond. A **246** (1958) 326-332.

Our exposition follows Faddeev's paper

• L.D. Faddeev, *The Feynman integral for singular Lagrangians*, Theoret. and Math. Phys., **1**:1 (1969), 1–13.

More details on Hamiltonian systems with constraints can be found in the physics book

• Marc Henneaux and Claudio Teitelboim, *Quantization of Gauge Systems*, Princeton University Press, Princeton, NJ, 1992.

Part 2

Classical gauge theories

LECTURE 8

Maxwell equations

8.1. Physics formulation

The electromagnetic force is a fundamental force responsible for the interaction of electrically charged particles. Particles with positions $\boldsymbol{r}_a \in \mathbb{R}^3$, $a = 1, \dots, N$, may carry electric charges e_a with the density function

$$\rho(\boldsymbol{r}) = \sum_{a=1}^{N} e_a \delta(\boldsymbol{r} - \boldsymbol{r}_a).$$

In general one considers the charge density — a signed σ-additive measure, which is absolutely continuous with respect to the standard Lebesgue measure on $\mathbb{R}^3$, i.e., a signed measure $\rho(\boldsymbol{r})d^3\boldsymbol{r}$. Moving charges produce electric current. A single charge e_0 at a moving point $\boldsymbol{r}_0(t)$ produces a current

$$\boldsymbol{j}(\boldsymbol{r}, t) = e_0 \boldsymbol{v}(t) \delta(\boldsymbol{r} - \boldsymbol{r}_0(t)), \quad \text{where} \quad \boldsymbol{v}(t) = \frac{d\boldsymbol{r}_0(t)}{dt}.$$

In general, the current density is

$$\boldsymbol{j}(\boldsymbol{r}, t) = \rho(\boldsymbol{r}, t)\boldsymbol{v}(\boldsymbol{r}, t),$$

where $\boldsymbol{v}(\boldsymbol{r}, t)$ is a charge velocity at point $\boldsymbol{r} \in \mathbb{R}^3$ at time t.

An electric field $\boldsymbol{E}(\boldsymbol{r}, t)$, where $\boldsymbol{r} \in \mathbb{R}^3$, is generated by electric charge, and time-varying magnetic field $\boldsymbol{B}(\boldsymbol{r}, t)$, which is produced by moving electric charges. They satisfy Maxwell equations, which summarize the basic laws of electromagnetism. In a free space these equations have the following beautiful form[1]

$$\boldsymbol{\nabla} \cdot \boldsymbol{E} = \frac{1}{\varepsilon_0}\rho \qquad \text{(Gauss law)} \tag{8.1}$$

— the electric flux leaving a volume is proportional to the charge inside;

$$\boldsymbol{\nabla} \cdot \boldsymbol{B} = 0 \qquad \text{(Gauss law for magnetism)} \tag{8.2}$$

[1]We are using standard notations for the divergence and curl from the multivariable calculus.

— there are no magnetic charges, the total magnetic flux through a closed surface is zero;

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \qquad \text{(Faraday's induction law)} \tag{8.3}$$

— the voltage induced in a closed circuit is proportional to the rate of change of the magnetic flux it encloses;

$$\nabla \times \boldsymbol{B} = \mu_0 \boldsymbol{j} + \mu_0 \varepsilon_0 \frac{\partial \boldsymbol{E}}{\partial t} \qquad \text{(Ampère's circular law)} \tag{8.4}$$

— the magnetic field induced around a closed loop is proportional to the electric current plus displacement current (rate of change of electric field) it encloses.

Here the constant ε_0 is called a *permitivity of the free space* and the constant μ_0 is called *permeability of the free space* or *magnetic constant*. They satisfy

$$\mu_0 \varepsilon_0 = \frac{1}{c^2},$$

where c is the speed of light in the free space.[2] Maxwell equations imply all laws of the electromagnetism: Coulomb law, Bio-Laplace-Savart law, etc.

It follows from equation (8.2) that there is a vector-valued function $\boldsymbol{A}(\boldsymbol{r}, t)$, called *vector potential*, $\boldsymbol{A} = (A_x, A_y, A_z)$, such that

$$\boldsymbol{B} = \nabla \times \boldsymbol{A}. \tag{8.5}$$

Plugging (8.5) into (8.3) we get

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

so that there is a function $\varphi(\boldsymbol{r}, t)$, called *scalar potential*, such that

$$\boldsymbol{E} = -\nabla \varphi - \frac{\partial \boldsymbol{A}}{\partial t}. \tag{8.6}$$

Formulas (8.5) and (8.6) solve the first pair of Maxwell equations — equations (8.2)–(8.3).

8.2. Using differential forms

One can rewrite (8.5)–(8.6) as single equation by introducing the following four-dimensional notations[3]. Put $x^0 = ct, x^1 = x, x^2 = y, x^3 = z$ and consider four-vectors $\boldsymbol{x} \in \mathbb{R}^4$, $\boldsymbol{x} = (x^\mu)$, where $\mu = 0, 1, 2, 3$. Put[4]

$$A = A_\mu dx^\mu,$$

[2] In the SI system of units $\varepsilon_0 = 8.85 \times 10^{-12} \mathrm{C}^2 \mathrm{N}^{-1} \mathrm{m}^{-2}$, where C = Coulomb and N = Newton, and $\mu_0 = 4\pi \times 10^{-7} \mathrm{NA}^{-2}$, A = Ampère. In the Gaussian system of units (a part of CGS system of units based on centimetre-gram-second) $\varepsilon_0 = \frac{1}{4\pi c}$, $\mu_0 = \frac{4\pi}{c}$ and $\boldsymbol{E}_{\mathrm{CGS}} = c^{-1} \boldsymbol{E}_{\mathrm{SI}}$.

[3] No reference to the special relativity yet!

[4] Here and in what follows we always use summation over repeated indices.

where $A_0 = \frac{1}{c}\varphi, A_1 = -A_x, A_2 = -A_y, A_3 = -A_z$, and consider the 2-form $F = dA$. Explicitly,

$$F = \frac{1}{2}F_{\mu\nu}dx^\mu \wedge dx^\nu, \quad \text{where} \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{8.7}$$

and $\partial_\mu = \dfrac{\partial}{\partial x^\mu}$, $\mu = 0, 1, 2, 3$.

It follows from (8.5)–(8.6) that the skew-symmetric 2-tensor $F_{\mu\nu}$ is represented by the following 4×4 matrix

$$F = \begin{pmatrix} 0 & \frac{1}{c}E_x & \frac{1}{c}E_y & \frac{1}{c}E_z \\ -\frac{1}{c}E_x & 0 & -B_z & B_y \\ -\frac{1}{c}E_y & B_z & 0 & -B_x \\ -\frac{1}{c}E_z & -B_y & B_x & 0 \end{pmatrix}, \tag{8.8}$$

or

$$\begin{aligned} F = \frac{1}{c}E_x dx^0 \wedge dx^1 + \frac{1}{c}E_y dx^0 \wedge dx^2 + \frac{1}{c}E_z dx^0 \wedge dx^3 \\ -B_x dx^2 \wedge dx^3 - B_y dx^3 \wedge dx^1 - B_z dx^1 \wedge dx^2 \end{aligned}$$

The 2-tensor $F_{\mu\nu}$ is called the *electromagnetic field tensor*, or the *field strength tensor* or *Faraday tensor*.

Equation $F = dA$ gives expressions (8.5)–(8.6) for electric and magnetic fields in terms of the f*our-vector potential* A_μ. The first pair of Maxwell equations — equations (8.2)–(8.3) — directly follow from this representation, and can be written succinctly as

$$dF = 0,$$

or, equivalently,

$$\partial_\nu F_{\lambda\mu} + \partial_\lambda F_{\mu\nu} + \partial_\mu F_{\nu\lambda} = 0, \quad \lambda, \mu, \nu = 0, 1, 2, 3. \tag{8.9}$$

Indeed, by an elementary computation we have

$$\begin{aligned} dF = -\boldsymbol{\nabla}\cdot\boldsymbol{B}\, dx^1 \wedge dx^2 \wedge dx^3 - \left(\partial_0 B_x + \frac{1}{c}(\boldsymbol{\nabla}\times\boldsymbol{E})_x\right)dx^0 \wedge dx^2 \wedge dx^3 \\ - \left(\partial_0 B_y + \frac{1}{c}(\boldsymbol{\nabla}\times\boldsymbol{E})_y\right)dx^0 \wedge dx^3 \wedge dx^1 - \left(\partial_0 B_z + \frac{1}{c}(\boldsymbol{\nabla}\times\boldsymbol{E})_z\right)dx^0 \wedge dx^1 \wedge dx^2. \end{aligned}$$

To rewrite the second pair of Maxwell equations, equations (8.1) and (8.4), we observe that in the absence of the sources these equations can be obtained from the first pair (8.2)–(8.3) by the *electro-magnetic duality*

$$\frac{1}{c}\boldsymbol{E} \mapsto -\boldsymbol{B} \quad \text{and} \quad \boldsymbol{B} \mapsto \frac{1}{c}\boldsymbol{E}.$$

Under this transformation $F \mapsto *F$, the dual field strength 2-form, given by

$$\begin{aligned} *F = -B_x dx^0 \wedge dx^1 - B_y dx^0 \wedge dx^2 - B_z dx^0 \wedge dx^3 \\ -\frac{1}{c}E_x dx^2 \wedge dx^3 - \frac{1}{c}E_y dx^3 \wedge dx^1 - \frac{1}{c}E_z dx^1 \wedge dx^2, \end{aligned}$$

so that equations (8.1) and (8.4) can be written as a single equation

$$d * F = 0.$$

What is the geometric meaning of the dual 2-form $*F$? It is easy to check that it is a Hodge dual to the 2-form F with respect to the Minkowski metric

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu$$

on $\mathbb{R}^4$, given by the diagonal 4×4 matrix $\eta = \mathrm{diag}(1, -1, -1, -1)$! In other words, Minkowski metric is a pseudo-Riemannian metric on $\mathbb{R}^4$, given explicitly by

$$ds^2 = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2. \tag{8.10}$$

Indeed, let V be an oriented n-dimensional real vector space with a non-degenerate inner product $\langle\ ,\ \rangle$ (not necessarily positive-definite). The inner product on the vector spaces $\Lambda^k V$ is defined by

$$\langle u_1 \wedge \cdots \wedge u_k, v_1 \wedge \cdots \wedge v_k \rangle = \det(\langle u_i, v_j \rangle).$$

Let $\omega \in \Lambda^n V$ be the unit vector associated with the orientation of V — an element in $\Lambda^n V$, uniquely characterized by the property that its image is 1 under the isomorphism $\Lambda^n V \simeq \mathbb{R}$. Then for $v \in \Lambda^k V$ its Hodge dual is a vector $*v \in \Lambda^{n-k} V$, satisfying

$$u \wedge *v = \langle u, v \rangle \omega \quad \text{for all} \quad v \in \Lambda^k V.$$

Applying this definition to the vector space V with the basis dx^0, dx^1, dx^2, dx^3 and the inner product $\langle dx^\mu, dx^\nu \rangle = \eta^{\mu\nu}$, we get

$$*(a_{\mu\nu} dx^\mu \wedge dx^\nu) = b_{\mu\nu} dx^\mu \wedge dx^\nu,$$

where

$$b_{\mu\nu} = \frac{1}{2} \varepsilon_{\alpha\beta\mu\nu} \eta^{\alpha\lambda} \eta^{\beta\rho} a_{\lambda\rho}$$

and $\varepsilon_{\alpha\beta\gamma\delta}$ is totally antisymmetric tensor, $\varepsilon_{0123} = 1$. We have

$$\begin{aligned}
*(dx^0 \wedge dx^1) &= -dx^2 \wedge dx^3, \\
*(dx^0 \wedge dx^2) &= \quad dx^1 \wedge dx^3, \\
*(dx^0 \wedge dx^3) &= -dx^1 \wedge dx^2, \\
*(dx^1 \wedge dx^2) &= \quad dx^0 \wedge dx^3, \\
*(dx^3 \wedge dx^1) &= \quad dx^0 \wedge dx^2, \\
*(dx^2 \wedge dx^3) &= \quad dx^0 \wedge dx^1,
\end{aligned}$$

and the formula for $*F$ follows from the definition of the 2-form F.

To summarize, Maxwell equations in an empty space (without sources) can be written succinctly as

$$dF = 0 \quad \text{and} \quad d * F = 0. \tag{8.11}$$

Remark. The signs in Maxwell equations, reflected in electro-magnetic duality, forces the use of a pseudo-Riemannian metric (8.10). This may be considered as an alternative discovery of the Minkowski spacetime, without the reference to special relativity.

8.3. Maxwell's equations with sources

We have

$$\begin{gathered} d*F = \\ \left((\boldsymbol{\nabla}\times\boldsymbol{B})_x - \frac{1}{c}\partial_0 E_x\right) dx^0\wedge dx^2\wedge dx^3 - \left((\boldsymbol{\nabla}\times\boldsymbol{B})_y - \frac{1}{c}\partial_0 E_y\right) dx^0\wedge dx^1\wedge dx^3 \\ + \left((\boldsymbol{\nabla}\times\boldsymbol{B})_z - \frac{1}{c}\partial_0 E_z\right) dx^0\wedge dx^1\wedge dx^2 - \frac{1}{c}\boldsymbol{\nabla}\cdot\boldsymbol{E}\, dx^1\wedge dx^2\wedge dx^3. \end{gathered}$$

Define the four-current

$$J = J_\mu dx^\mu,$$

where $J_0 = -c\rho$ and $J_1 = j_x, J_2 = j_y, J_3 = j_z$. Using

$$\begin{aligned} *(dx^0\wedge dx^2\wedge dx^3) &= dx^1, \\ *(dx^0\wedge dx^1\wedge dx^3) &= -dx^2, \\ *(dx^0\wedge dx^1\wedge dx^2) &= dx^3, \\ *(dx^1\wedge dx^2\wedge dx^3) &= dx^0, \end{aligned}$$

we can succinctly rewrite equations (8.1) and (8.4) as

$$*d*F = \mu_0 J.$$

Equivalently, since $*^2 = -(-1)^k$ on the space of k-forms on $\mathbb{R}^4$, we have

$$d*F = \mu_0 *J,$$

so that $d*J = 0$, which is a *continuity equation*. Using that

$$\begin{aligned} *dx^0 &= dx^1\wedge dx^2\wedge dx^3, \\ *dx^1 &= dx^0\wedge dx^2\wedge dx^3, \\ *dx^2 &= -dx^0\wedge dx^1\wedge dx^3, \\ *dx^3 &= dx^0\wedge dx^1\wedge dx^2, \end{aligned}$$

we can rewrite it as follows

$$\frac{\partial J^\mu}{\partial x^\mu} = 0, \quad \text{where} \quad J^\mu = \eta^{\mu\nu} J_\nu.$$

Explicitly, the continuity equation has the form

$$\frac{\partial\rho}{\partial t} + \boldsymbol{\nabla}\cdot\boldsymbol{j} = 0.$$

REMARK. If J has compact support or is of rapid decay, the continuity equation leads to the total charge conservation. Namely, let

$$Q(t) = -\frac{1}{c}\int_{\{ct\}\times\mathbb{R}^3} *J = \int_{\mathbb{R}^3} \rho(t,\boldsymbol{r})d^3\boldsymbol{r}$$

be the total charge at time t. Then it follows from Stokes's theorem for $M = [ct_1, ct_2] \times \mathbb{R}^3$ that

$$0 = \int_M d * J = \int_{\partial M} *J = Q(t_2) - Q(t_1).$$

Also for any compact 3-manifold $V \subset \mathbb{R}^3$ we have

$$\frac{\partial}{\partial t}\int_V \rho(t,\boldsymbol{r})d^3\boldsymbol{r} = -\int_{\partial V} \boldsymbol{j}\cdot dS.$$

It is also convenient to introduce the tensor

$$F^{\mu\nu} = \eta^{\mu\alpha}\eta^{\nu\beta}F_{\alpha\beta},$$

which has the the same form as $F_{\mu\nu}$, where $\boldsymbol{E}$ is replaced by $-\boldsymbol{E}$. It is related to the dual strength field tensor by

$$(*F)_{\mu\nu} = \frac{1}{2}\varepsilon_{\mu\nu\alpha\beta}F^{\alpha\beta}.$$

Then the second pair of Maxwell equations can be written in the following form

$$\partial_\mu F^{\mu\nu} = J^\nu, \quad \nu = 0,1,2,3, \tag{8.12}$$

which is often used by physicists.

To summarize, the Maxwell's equations on $\mathbb{R}^4$ have the following form

$$dF = 0 \quad \text{and} \quad *\,d * F = J, \tag{8.13}$$

where the 4-current J satisfies the continuity equation. By Poincaré lemma, the first equation has a solution

$$F = dA \quad \text{where} \quad A = A_\mu dx^\mu.$$

Upon the identification $A_0 = \frac{1}{c}\varphi$ and $(A_1, A_2, A_3) = -\boldsymbol{A}$ we get expressions (8.5) and (8.6) for the magnetic and electric fields in terms of the vector and scalar potentials $\boldsymbol{A}$ and φ. Maxwell's equations are invariant under the *gauge transformations*

$$A \mapsto A + df,$$

where f is a smooth real-valued function on $\mathbb{R}^4$.

8.4. The principle of least action

The Maxwell equations (8.13) can be obtained from the principle of least action.

Namely, let $\mathcal{A} = \Omega^1(\mathbb{R}^4)$ be a vector space of smooth (C^∞) real-valued 1-forms $A = A_\mu dx^\mu$ on $\mathbb{R}^4$ such that corresponding 2-forms $F = dA$ have compact support (or decay sufficiently fast as $|\boldsymbol{x}| \to \infty$). Let J be a smooth real-valued 1-form on $\mathbb{R}^4$ with compact support (or decaying sufficiently fast as $|\boldsymbol{x}| \to \infty$) satisfying the continuity equation. Define the action functional $S : \mathcal{A} \to \mathbb{R}$ by

$$S(A) = -\frac{1}{4\pi}\int\limits_{\mathbb{R}^4}(F \wedge *F + 2A \wedge *J), \tag{8.14}$$

where $F = dA$.

Proposition 8.1. *The critical points of the action functional $S(A)$ are given by the Maxwell equations.*

Proof. For given $a \in \mathcal{A}$ put

$$\delta S(A) = \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} S(A + \varepsilon a).$$

We have, using the symmetry property of the Hodge star operator

$$\alpha \wedge *\beta = \beta \wedge *\alpha \tag{8.15}$$

and the Stokes theorem,

$$\begin{aligned}
\delta S(A) &= -\frac{1}{2\pi}\int\limits_{\mathbb{R}^4}(da \wedge *F + a \wedge *J) \\
&= -\frac{1}{2\pi}\int\limits_{\mathbb{R}^4}(a \wedge d * F + a \wedge *J) - \frac{1}{2\pi}\int\limits_{\mathbb{R}^4} d(a \wedge *F) \\
&= -\frac{1}{2\pi}\int\limits_{\mathbb{R}^4} a \wedge (d * F + *J).
\end{aligned}$$

Whence $\delta S(A) = 0$ for all $a \in \mathcal{A}$ yields

$$d * F = - * J. \quad \square$$

Remark. As in Sect. 1.2 in Lecture 1, one can consider a vector space $\mathcal{A}[^{A^1,t_1}_{A^0,t_0}]$ of real-valued 1 forms $A = A_\mu dx^\mu$ on $[ct_0, ct_1] \times \mathbb{R}^3$ satisfying $A_\mu(ct_0, \boldsymbol{r}) = A^0_\mu(\boldsymbol{r})$ and $A_\mu(ct_1, \boldsymbol{r}) = A^1_\mu(\boldsymbol{r})$. It follows from the above computation using the variations with fixed ends that critical points of the functional

$$-\frac{1}{4\pi}\int_{t_0}^{t_1}\int_{\mathbb{R}^3}(F \wedge *F + 2A \wedge *J)d^3\boldsymbol{r}dt$$

are given by the Maxwell equations.

Remark. In physics notation,

$$S(A) = \int\limits_{\mathbb{R}^4} \left(\mathscr{L}(A) - \frac{1}{2\pi} A_\mu J^\mu \right) d^4\boldsymbol{x},$$

where

$$\mathscr{L}(A) = -\frac{1}{16\pi} F_{\mu\nu} F^{\mu\nu} = \frac{c}{8\pi} \left(\frac{1}{c^2} \boldsymbol{E}^2 - \boldsymbol{B}^2 \right) \tag{8.16}$$

is the Lagrangian function of the free electromagnetic field.

LECTURE 9

Electrodynamics as U(1) gauge theory

Electrodynamics — theory of electromagnetism, described by Maxwell equations — is a gauge theory with the symmetry group $G = \mathrm{U}(1)$. To explain this fundamental fact, and to formulate the gauge theory with arbitrary compact symmetry group G — the celebrated Yang-Mills theory — one needs to use differential geometry of principal and vector bundles. It is succinctly summarized below.

9.1. Bundles, connections and curvature

Let G be a Lie group and M be a smooth manifold. A principal G-bundle over X is fiber bundle $\pi : P \to M$ with the smooth right G-action

$$P \times G \ni (p, g) \mapsto p \cdot g \in P,$$

which preserves the fibers and is free and transitive. By definition, there is an open covering $M = \bigcup_{\alpha \in A} U_\alpha$ such that over each U_α there is a local trivialization, a diffeomorphism

$$\varphi_\alpha : \pi^{-1}(U_\alpha) \to U_\alpha \times G$$

such that

$$\pi(\varphi_\alpha^{-1}(x, g)) = x \quad \text{and} \quad \varphi_\alpha^{-1}(x, g) = \varphi_\alpha^{-1}(x, e) \cdot g \quad \text{for all} \quad x \in U_\alpha,\ g \in G,$$

where e is identity in G. Putting

$$\lambda_{\alpha\beta} = \varphi_\alpha \circ \varphi_\beta^{-1} : U_{\alpha\beta} \times G \to U_{\alpha\beta} \times G,$$

where $U_{\alpha\beta} = U_\alpha \cap U_\beta$, introduces transition functions $t_{\alpha\beta} : U_{\alpha\beta} \to G$ by

$$\lambda_{\alpha\beta}(x, g) = (x, t_{\alpha\beta}(x)g).$$

The transition functions satisfy

$$t_{\alpha\beta} = t_{\beta\alpha}^{-1} \quad \text{on} \quad U_{\alpha\beta} \tag{9.1}$$

and

$$t_{\alpha\beta} t_{\beta\gamma} t_{\gamma\alpha} = e \quad \text{on} \quad U_{\alpha\beta\gamma} = U_\alpha \cap U_\beta \cap U_\gamma. \tag{9.2}$$

Conversely, a principal G-bundle P can be defined by transition functions, maps $t_{\alpha\beta} : U_{\alpha\beta} \to G$, satisfying (9.1)–(9.2) by

$$P = \bigsqcup_{\alpha \in A} (U_\alpha \times G)/\sim, \tag{9.3}$$

where $(x, g) \sim (y, h)$ if and only if $x = y \in U_{\alpha\beta}$ and $g = t_{\alpha\beta}(x)h$. Transition functions $t_{\alpha\beta}$ and $f_\alpha^{-1} t_{\alpha\beta} f_\beta$, where $f_\alpha : U \to G$, are arbitrary smooth functions, define the same bundle P. Sections of P over $U \subseteq M$ are the maps $\mathcal{s} : U \to P$ satisfying $\pi \circ \mathcal{s} = \mathrm{id}|_U$. They are determined by the maps $\mathcal{s}_\alpha : U_\alpha \to G$, satisfying

$$\mathcal{s}_\beta = \mathcal{s}_\alpha t_{\alpha\beta} \quad \text{on} \quad U_{\alpha\beta}. \tag{9.4}$$

The gauge group $\mathcal{G}(P)$ of a principal G-bundle P consists of bundle isomorphisms $f : P \to P$ that commute with right action. Such f can be uniquely written $f(p) = p \cdot f_*(p)$, where a function $f_* : P \to G$ satisfies

$$f_*(p \cdot g) = g^{-1} f_*(p) g \quad \text{for all} \quad p \in P,\ g \in G.$$

Elements of the gauge group $\mathcal{G}(P)$ are collections $\{f_\alpha\}_{\alpha \in A}$ of arbitrary smooth functions $f_\alpha : U_\alpha \to G$ that map sections to sections by the formula $\mathcal{s}' = \mathcal{s} \circ f$. Explicitly,

$$\mathcal{s}'_\alpha = \mathcal{s}_\alpha f_\alpha : U_\alpha \to G,$$

and $\mathcal{s}'_\alpha$ satisfy (9.4) with the transition functions

$$t'_{\alpha\beta} = f_\alpha^{-1} t_{\alpha\beta} f_\beta.$$

With every representation $R : G \to \mathrm{GL}(V)$ of a Lie group G in a complex vector space V there is a vector bundle $E \to M$ of rank $n = \dim V$, associated with a principal G-bundle $P \to M$. It has fiber V and the structure group G, as is defined as a quotient

$$E = (P \times V)/G, \tag{9.5}$$

where the right G-action is given by

$$(p, v) \cdot g = (p \cdot g, R(g^{-1})v), \quad p \in P,\ v \in V.$$

Equivalently, a vector bundle $E \to M$ can be defined by the transition functions $g_{\alpha\beta}$. Sections of E over $U \subseteq M$ are the functions $s_\alpha : U_\alpha \to V$, satisfying

$$s_\alpha = g_{\alpha\beta} s_\beta \quad \text{on} \quad U_{\alpha\beta}. \tag{9.6}$$

If a vector bundle $E \to M$ is associated with a principal G-bundle $P \to M$ through a representation $R : G \to \mathrm{GL}(V)$, then $g_{\alpha\beta} = R(t_{\alpha\beta})$.

Denote by $\Omega^0(E)$ the sheaf of smooth sections of a vector bundle E and by $\Omega^1(E)$ — a sheaf of 1-forms on M with values in E — a sheaf of smooth sections of $E \otimes T^*M \to M$. Connection[1] on E is a linear map $\nabla : \Omega^0(E) \to \Omega^1(E)$ satisfying the Leibniz rule

$$\nabla(f\zeta) = df \otimes \zeta + f(\nabla\zeta) \tag{9.7}$$

for all sections $\zeta \in \Omega^0(E)(U)$ and functions $f \in C^\infty(U)$, $U \subset M$. Connections can be thought of as a way of differentiating sections of E.

In terms of transition functions $g_{\alpha\beta}$ of the bundle E, connection ∇ is a collection $\{d + A_\alpha\}_{\alpha\in A}$, where d is the de Rham differential and A_α are $\operatorname{End} V$-valued 1-forms on U_α, satisfying the transformation law

$$A_\alpha = g_{\alpha\beta} A_\beta g_{\alpha\beta}^{-1} - dg_{\alpha\beta} g_{\alpha\beta}^{-1} \quad \text{on} \quad U_{\alpha\beta}. \tag{9.8}$$

Indeed, if s_α satisfy (9.6), then V-valued 1-forms on U_α

$$\nabla s_\alpha = (d + A_\alpha) s_\alpha \tag{9.9}$$

satisfy $\nabla s_\alpha = g_{\alpha\beta} \nabla s_\beta$ on $U_{\alpha\beta}$ if and only if

$$(d + A_\alpha)(g_{\alpha\beta} s_\beta) = dg_{\alpha\beta} s_\beta + g_{\alpha\beta} ds_\beta + A_\alpha g_{\alpha\beta} s_\beta$$

and

$$g_{\alpha\beta}(d + A_\beta) s_\beta = g_{\alpha\beta} ds_\beta + g_{\alpha\beta} A_\beta s_\beta$$

are equal for all $s_\beta|_{U_{\alpha\beta}}$, which is equation (9.8). Notation $\nabla_A = d + A$ we will used sometimes.

In local coordinates $x^1, \dots, x^n$ on a chart $U \subseteq M$,

$$\nabla s = \nabla_\mu(s) dx^\mu, \quad \text{where} \quad \nabla_\mu = \partial_\mu + A_\mu \quad \text{and} \quad A = A_\mu dx^\mu.$$

Operators ∇_μ are called *covariant derivatives*.

Remark. In physics literature the notation $\nabla_\mu = \partial_\mu + ieA_\mu$ is customary used, where e is the elementary charge — the magnitude of the electron charge $-e$. In quantum field theory the electromagnetic field describes photons, the exchange particles 'of light' for the electromagnetic interaction.

If a vector bundle $E \to M$ is associated with a principal G-bundle $P \to M$ through a representation $R : G \to \mathrm{GL}(V)$, denote by $\rho = d_e R$ the corresponding infinitesimal representation — a representation of a Lie algebra $\mathfrak{g}$ in $\operatorname{End} V$. Connections ∇ on E with the symmetry group[2] G have the property that A_α are 1-forms on U_α with values in $\rho(\mathfrak{g})$.

Connections ∇ form an affine space $\mathcal{A}(E)$ over the complex vector space $\Omega^1(M, \operatorname{End} E)$ of $\operatorname{End} E$-valued 1-forms on M. Here $\operatorname{End} E = E \otimes E^*$, where E^* is a dual bundle to E, is an endomorphism bundle of E with the transition functions $g_{\alpha\beta} \otimes g^*_{\alpha\beta}$, where $g^*_{\alpha\beta} = (g^{\mathrm{t}}_{\alpha\beta})^{-1}$. The gauge group $\mathcal{G}(E)$ consists of

[1] Here we use a definition that does not use a notion of a connection on a principal G-bundle.

[2] Such connections are obtained from connections on a principal G-bundle P.

maps $\phi = \{\phi_\alpha : U_\alpha \to \operatorname{End} V\}_{\alpha\in A}$, and it acts on $\mathcal{A}(E)$ by $A^\phi = \{A^\phi_\alpha\}_{\alpha\in A}$, where

$$A^\phi_\alpha = \phi_\alpha A_\alpha \phi_\alpha^{-1} - d\phi_\alpha \phi_\alpha^{-1} \quad \text{on} \quad U_\alpha. \tag{9.10}$$

Connection ∇_A on a bundle E determines a connection $\nabla_A^{\operatorname{End} E}$ on the bundle $\operatorname{End} E$,

$$\nabla^{\operatorname{End} E} \boldsymbol{s}_\alpha = d\boldsymbol{s}_\alpha + [A_\alpha, \boldsymbol{s}_\alpha], \tag{9.11}$$

where $\boldsymbol{s}_\alpha : U_\alpha \to \operatorname{End} V$ satisfy

$$\boldsymbol{s}_\alpha = g_{\alpha\beta} \boldsymbol{s}_\beta g_{\alpha\beta}^{-1} \quad \text{on} \quad U_{\alpha\beta}.$$

A linear map $\nabla : \Omega^0(E) \to \Omega^1(E)$ satisfying (9.7), by Leibniz rule extends to a map $\Omega^k(E) \to \Omega^{k+1}(E)$, which we continue to denote by ∇. Explicitly, it is determined by

$$\nabla(\psi \otimes \zeta) = d\psi \otimes \zeta + (-1)^k \psi \wedge \nabla\zeta,$$

where $\zeta \in \Omega^0(E)(U)$ and $\psi \in \Omega^k(M)(U)$. In particular, for the map

$$\nabla^2 : \Omega^0(E) \to \Omega^2(E)$$

we obtain

$$\begin{aligned} \nabla^2(f\zeta) &= \nabla(df \otimes \zeta + f\nabla\zeta) \\ &= -df \wedge \nabla\zeta + df \wedge \nabla\zeta + f\nabla^2\zeta = f\nabla^2\zeta. \end{aligned}$$

This means that $\nabla^2 : \Omega^0(E) \to \Omega^2(E)$ is determined by a 2-form F on M with values in $\operatorname{End} E$ — a global section of the bundle $\Lambda^2 T^*M \otimes \operatorname{End} E$ — by

$$\nabla^2 s_\alpha = F_\alpha s_\alpha \quad \text{on} \quad U_\alpha.$$

In terms of the transitions functions,

$$\begin{aligned} \nabla^2 s_\alpha &= (d + A_\alpha)(ds_\alpha + A_\alpha s_\alpha) \\ &= dA_\alpha s_\alpha - A_\alpha \wedge ds_\alpha + A_\alpha \wedge ds_\alpha + A_\alpha \wedge A_\alpha s_\alpha \\ &= (dA_\alpha + A_\alpha \wedge A_\alpha) s_\alpha, \end{aligned}$$

where $A_\alpha \wedge A_\alpha$ is understood as a product in $\operatorname{End} V$ together with the usual exterior multiplication. Thus $\operatorname{End} E$-valued 2-form F on M is a collection $\{F_\alpha\}_{\alpha\in A}$ of $\operatorname{End} V$-valued 2-forms on U_α,

$$F_\alpha = dA_\alpha + A_\alpha \wedge A_\alpha, \tag{9.12}$$

satisfying

$$F_\alpha = g_{\alpha\beta} F_\beta g_{\alpha\beta}^{-1} \quad \text{on} \quad U_{\alpha\beta}. \tag{9.13}$$

Transformation law (9.13) follows from (9.8)–(9.12). We will often use notation

$$F = F(A) = dA + A \wedge A.$$

If a vector bundle $E \to M$ is associated with a principal G-bundle P over M through a representation $R : G \to \mathrm{GL}(V)$, corresponding 2-forms F_α on U_α take values in $\rho(\mathfrak{g})$, where $\rho = d_e R$. It follows from (9.10) that the action of the gauge group $\mathcal{G}(E)$ on F is given by $F \mapsto F^\phi$, where

$$F^\phi_\alpha = \phi_\alpha F_\alpha \phi_\alpha^{-1} \quad \text{on} \quad U_\alpha. \tag{9.14}$$

In local coordinates $x^1, \dots, x^n$ on a chart $U \subseteq M$,

$$F = \frac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu, \quad \text{where} \quad F_{\mu\nu} = [\nabla_\mu, \nabla_\nu] = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu} + [A_\mu, A_\nu]$$

and $[A_\mu, A_\nu] = A_\mu \wedge A_\nu - A_\nu \wedge A_\mu$. It follows from (9.11) that curvature satisfies the Bianchi identity,

$$\nabla_A^{\mathrm{End}\,E}(F) = dF + A \wedge F - F \wedge A = 0, \tag{9.15}$$

which we will simply write as $\nabla_A F = 0$. It can also be obtained from the Jacobi identity

$$[\nabla_\mu, \nabla_\nu], \nabla_\sigma] + [\nabla_\nu, \nabla_\sigma], \nabla_\mu] + [\nabla_\sigma, \nabla_\mu], \nabla_\nu] = 0.$$

Remark. In general, for $B \in \Omega^k(M, \mathrm{End}\,E)$ we have

$$\nabla_A^{\mathrm{End}\,E}(B) = dB + A \wedge B - (-1)^k B \wedge A.$$

Let $\Phi : \mathrm{End}\,V \to \mathbb{C}$ be a homogeneous polynomial of order k, invariant under the adjoint action of $\mathrm{GL}(V)$ on $\mathrm{End}\,V$,

$$\Phi(B) = P(gBg^{-1}) \quad \text{for all} \quad B \in \mathrm{End}\,V \ \text{ and } \ g \in \mathrm{GL}(V).$$

It follows form (9.13)

$$\Phi(F_\alpha) = \Phi(F_\beta) \quad \text{on} \quad U_{\alpha\beta},$$

so that $\Phi(F) \in \Omega^{2k}(M)$. The Chern-Weil theory establishes the following facts.

1. The $2k$-form $\Phi(F)$ on M is closed,

$$d\Phi(F) = 0.$$

2. Cohomology class

$$[\Phi(F)] \in H^{2k}(M)$$

does not depend on a choice of a connection $d + A$ in a vector bundle E.

3. A map
$$\Phi \mapsto \Phi(F)$$
is a homomorphism of a commutative algebra of invariant polynomials on $\operatorname{End} V$ into the commutative algebra $H^{\mathrm{even}}(M)$ of differential forms of even degree on M.

The map $\Phi \mapsto \Phi(F)$ is called *Weil homomorphism*, and cohomology classes $[\Phi(F)]$ — *characteristic classes* of a bundle E, associated with the invariant polynomial Φ. Let P^i be elementary invariant polynomials of degree $i = 1, \dots, n$, defined by
$$\det(B + tI) = \sum_{k=0}^{n} P^{n-k}(B)t^k.$$
Forms $c_i(F) = P^i\left(\frac{\sqrt{-1}}{2\pi}F\right)$ are called *Chern forms*, and corresponding cohomology classes — *Chern classes*. It is a fundamental fact in the theory of characteristic classes, that
$$c_i(E) = \left[P^i\left(\frac{\sqrt{-1}}{2\pi}F\right)\right] \in \check{H}^{2i}(M, \mathbb{Z}), \quad i = 1, \dots, n,$$
where $\check{H}^{2i}(M, \mathbb{Z})$ stands for the Čech cohomology with coefficients in the constant sheaf $\mathbb{Z}$.

9.2. Line bundles and Maxwell equations

Let $L \to M$ be a complex line bundle over an n-dimensional manifold M associated with a principal U(1)-bundle P over M. Let $\{U_\alpha\}_{\alpha \in A}$ be an open cover of M and let $g_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{U}(1)$ be a transition functions for L, satisfying the cocycle condition
$$g_{\alpha\beta} g_{\beta\gamma} = g_{\alpha\gamma} \quad \text{on} \quad U_\alpha \cap U_\beta \cap U_\gamma.$$
A unitary connection ∇ — connection with symmetry group U(1) — is given by
$$\nabla = d + A_\alpha, \tag{9.16}$$
where $A_\alpha \in \Omega^1(U_\alpha)$ are 1-forms on U_α with values in the Lie algebra $\mathfrak{u}(1) \simeq \sqrt{-1}\,\mathbb{R}$ of the Lie group U(1), satisfying
$$A_\alpha = A_\beta - g_{\alpha\beta}^{-1} dg_{\alpha\beta} \quad \text{on} \quad U_\alpha \cap U_\beta.$$
Corresponding curvature 2-form $F = \nabla^2$ is a global 2-form on M given by
$$F = dA$$
and is a closed form, $dF = 0$.

Suppose that M carries either Riemannian or pseudo-Riemannian metric ds^2, and let $*$ be the corresponding Hodge star operator. As in Sect. 8.4 in Lecture 8, consider the functional

$$S(A) = -\frac{1}{4\pi}\int_M F \wedge *F, \tag{9.17}$$

defined on the affine space $\mathcal{A}$ of unitary connections on the line bundle L. In case M is non-compact it is assumed that connection $\nabla = d + A$ is such that integral (9.17) with $F = dA$ is convergent. As in Sect. 8.4, the critical points of the functional $S(A)$ are given by the equations

$$dF = 0 \quad \text{and} \quad d * F = 0. \tag{9.18}$$

In case $M = \mathbb{R}^4$ with the Minkowski metric, and $L = \mathbb{R}^4 \times \mathbb{C}$ is a trivial line bundle, these equations are Maxwell equations (8.11) in empty space[3].

The functional $S(A)$ is invariant under the action of a gauge group $\mathcal{G}(L)$ and defines a gauge theory with the symmetry group $\mathrm{U}(1)$. Corresponding equations of motions are given by (9.18), and in case when M is a four-manifold with the metric ds^2 of the signature $(+,-,-,-)$, generalize Maxwell equations (8.11) in empty space to a 'curved' spacetime.

9.3. Self-duality equations

In the Riemannian case equations (9.18) do not have physical interpretation. However, in case when M is compact Riemannian 4-manifold, they have extra mathematical structure, which will be very important for non-abelian gauge theories. Namely, in this case the first Chern form of the line bundle L with connection ∇ is

$$c_1(L,\nabla) = \frac{\sqrt{-1}}{2\pi}F, \quad c_1(L) = [c_1(L,\nabla)] \in \check{H}^2(M,\mathbb{Z}).$$

Due to the isomorphism

$$\mathrm{U}(1) \ni e^{i\theta} \mapsto \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \in \mathrm{SO}(2),$$

for a complex $\mathrm{U}(1)$-line bundle L there is a real rank 2 vector bundle $\mathcal{L}$ over M with the symmetry group $\mathrm{SO}(2)$. Its first Pontryagin class $p_1(\mathcal{L}) \in \check{H}^4(M,\mathbb{Z})$ is given by

$$p_1(\mathcal{L}) = -c_2(\mathcal{L}\otimes\mathbb{C}),$$

where $\mathcal{L}\otimes\mathbb{C}$ is a complexification of the real bundle $\mathcal{L}$ — rank 2 complex vector bundle over M, and $c_2(\mathcal{L}\otimes\mathbb{C})$ is its second Chern class. It is easy to see that $\mathcal{L}\otimes\mathbb{C} \simeq L \oplus \bar{L}$, where $\bar{L}$ is the line bundle with the transition functions

[3] Note that if $A = A_\mu dx^\mu$ is a real-valued 1-form on $M = \mathbb{R}^4$, used in Sect 8.2 in Lecture 8 in case L is a trivial line bundle, then in (9.16) we have $\nabla = d + \sqrt{-1}\,A$.

$\bar{g}_{\alpha\beta}$, so that $c_2(\mathcal{L} \otimes \mathbb{C})$ is represented by the differential form $\dfrac{1}{4\pi^2} F \wedge F$. The corresponding first Pontryagin number is

$$p_1 = -\frac{1}{4\pi^2} \int_M F \wedge F \in \mathbb{Z}.$$

In case when M is a Riemannian manifold with the metric ds^2, then the Maxwell's equations on M have the form

$$dF = 0 \quad \text{and} \quad d * F = 0,$$

where $F \in \Omega^2(M, \sqrt{-1}\,\mathbb{R})$ and $*$ is the Hodge star of the metric ds^2. They characterize curvature forms F as harmonic 2-forms. Since in the Riemannian case $*^2 = 1$ on 2-forms, and we have a decomposition

$$\Omega^2(M, \sqrt{-1}\,\mathbb{R}) = \Omega^2_+(M, \sqrt{-1}\,\mathbb{R}) \oplus \Omega^2_-(M, \sqrt{-1}\,\mathbb{R}) \tag{9.19}$$

according to the eigenspaces of the Hodge $*$-operator corresponding to the eigenvalues 1 and -1. The 2-form F on M is called self-dual or anti-self-dual, if $F \in \Omega^2_+(M, \sqrt{-1}\,\mathbb{R})$ or $F \in \Omega^2_-(M, \sqrt{-1}\,\mathbb{R})$ respectively,

$$*F = \pm F.$$

Correspondingly, connection $\nabla = d + A$ on a U(1)-line bundle L is called self-dual or anti-self-dual, if its curvature 2-form $F = dA$ is, respectively, self-dual or anti-self-dual. Curvature forms of self-dual or anti-self-dual connections satisfy Maxwell's equations on a Riemannian 4-manifold M automatically!

From the inequality

$$-\int_M \omega \wedge *\omega \geq 0$$

for all $\omega \in \Omega^2(M, \sqrt{-1}\,\mathbb{R})$, we get for a curvature 2-form F of a line bundle $L \to M$,

$$\begin{aligned} -\int_M F \wedge *F - 4\pi^2 p_1 &= -\int_M F \wedge *F + F \wedge F \\ &= -\frac{1}{2}\int_M (F - *F) \wedge *(F - *F) \geq 0 \end{aligned}$$

and

$$\begin{aligned} -\int_M F \wedge *F + 4\pi^2 p_1 &= -\int_M F \wedge *F - F \wedge F \\ &= -\frac{1}{2}\int_M (F + *F) \wedge *(F + *F) \geq 0. \end{aligned}$$

Thus we obtain the inequality

$$S(A) \geq \pi |p_1|,$$

where the absolute minima of the action are given by the self-dual connections in case $p_1 > 0$, by the anti-self-dual connections in case $p_1 < 0$ and by both these types in case $p_1 = 0$.

Remark. In the pseudo-Riemannian case $*^2 = -1$ on 2-forms and analog of decomposition (9.19) is valid only for complex-valued 2-forms. Corresponding self-duality equations take the form

$$*F = \pm\sqrt{-1}\,F$$

and have no solutions in $\Omega^2(M, \sqrt{-1}\,\mathbb{R})$. in other words, these equations have only "non-physical" solutions.

Problem 9.1. Find local trivializations for a vector bundle defined by (9.5) and show that in this case definition (9.7) reduces to (9.8).

Problem 9.2. Show that property **1** follows from the Bianchi identity (9.15).

Problem 9.3. Prove property **2**. (*Hint*: given two connections ∇_0 and ∇_1 on E, consider a homotopy $\nabla_t = (1-t)\nabla_0 + t\nabla_1$).

Problem 9.4. Prove that for every closed 2-form F on a compact manifold M with the property

$$\left[\frac{\sqrt{-1}}{2\pi}F\right] \in \check{H}^2(M, \mathbb{Z}),$$

there is a line bundle $L \to M$ and a connection $\nabla = d + A$ such that $F = dA$.

LECTURE 10

Yang-Mills theory

Here we consider the case when G is compact, connected, semi-simple Lie group.

10.1. Yang-Mills equations

Let $E \to M$ be a complex rank r vector bundle over an n-dimensional manifold M, which may be considered as a vector bundle with a non-compact symmetry group $G = \mathrm{GL}(r, \mathbb{C})$. There is a natural bundle map of the bundle $\operatorname{End} E$ — the endomorphism bundle of E — to the trivial line bundle over M, given by the trace map $\operatorname{tr} : \operatorname{End} V \to \mathbb{C}$ in the fibers. Explicitly,

$$\operatorname{End} V = V \otimes V^* \ni v \otimes w \mapsto \operatorname{tr}(v \otimes w) = w(v) \in \mathbb{C}.$$

This determines a map

$$\Omega^p(M, \operatorname{End} E) \otimes \Omega^q(M, \operatorname{End} E) \ni \omega_1 \otimes \omega_2 \mapsto \operatorname{tr}(\omega_1 \wedge \omega_2) \in \Omega^{p+q}(M). \tag{10.1}$$

Namely, if $\omega_1 = \psi_1 \otimes \zeta_1$, $\omega_2 = \psi_2 \otimes \zeta_2$, where $\psi_1 \in \Omega^p(M)$, $\psi_2 \in \Omega^q(M)$ and $\zeta_1, \zeta_2 \in \Omega^0(M, \operatorname{End} E)$, then

$$\operatorname{tr}(\omega_1 \wedge \omega_2) = \operatorname{tr}(\zeta_1 \zeta_2)\, \psi_1 \wedge \psi_2.$$

A choice of a Riemannian or pseudo-Riemannian metric ds^2 on M defines a Hodge star operator on the algebra $\Omega^\bullet(M)$ of differential forms on M. It extends to the operator

$$\star : \Omega^p(M, \operatorname{End} E) \to \Omega^{n-p}(M, \operatorname{End} E)$$

by

$$\star(\psi \otimes \zeta) = *\psi \otimes \zeta, \quad \psi \in \Omega^p(M),\ \zeta \in \Omega^0(M, \operatorname{End} E).$$

Denote by $\mathscr{A}_E$ the affine space of connections on E.

Definition. A *Yang-Mills action functional* $S : \mathscr{A}_E \to \mathbb{C}$ is given by

$$S(A) = -\frac{1}{4\pi} \int_M \operatorname{tr}(F \wedge \star F), \quad F = dA + A \wedge A, \ \ A \in \mathscr{A}_E. \tag{10.2}$$

If manifold M is non-compact, we assume that connections A are such that the integral in (10.2) is convergent (e.g., F has compact support). It follows from (9.14) that the functional S is invariant under the action of a gauge group $\mathscr{G}$ with the symmetry group $\mathrm{GL}(r, \mathbb{C})$.

PROPOSITION 10.1. *The critical points of the Yang-Mills action functional are given by the solutions of the Yang-Mills equations*

$$\nabla_A F = 0 \quad \textit{and} \quad \nabla_A \star F = 0. \tag{10.3}$$

PROOF. The first equation is just a Bianchi identity (9.15) in Lecture 9, while derivation of the second equation repeats the proof of Proposition 8.1 in Lecture 8. Namely, for $a \in \Omega^1(M, \operatorname{End} E)$ we have

$$\begin{aligned} F(A+a) &= F(A) + da + A \wedge a + a \wedge A + a \wedge a \\ &= F(A) + \nabla_A a + a \wedge a. \end{aligned}$$

Whence using the cyclic property of the trace, formula (8.15), Leibniz rule

$$d(a \wedge \star F) = da \wedge \star F - a \wedge d \star F$$

and Stokes theorem, we obtain

$$\begin{aligned} \delta S(A) &= \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} S(A + \varepsilon a) \\ &= -\frac{1}{4\pi} \int_M \operatorname{tr}\left((da + A \wedge a + a \wedge A) \wedge \star F + F \wedge \star(da + A \wedge a + a \wedge A)\right) \\ &= -\frac{1}{2\pi} \int_M \operatorname{tr}\left((da + A \wedge a + a \wedge A) \wedge \star F\right) \\ &= -\frac{1}{2\pi} \int_M \operatorname{tr}\left(a \wedge (d \star F + A \wedge \star F - \star F \wedge A)\right) \\ &= -\frac{1}{2\pi} \int_M \operatorname{tr}(a \wedge \nabla_A \star F). \end{aligned}$$

□

Suppose that the vector bundle E is associated with a principal G-bundle P over M through a representation $R : G \to \mathrm{GL}(V)$ of a compact Lie group G. When representation R is unitary with respect to Hermitian inner product in V, restriction of the Yang-Mills functional to the connections $\mathscr{A}_E^G$ with the symmetry group G gives a functional taking non-negative values. Indeed, in this case $\rho(\mathfrak{g})$ consists of skew-Hermitian endomorphisms V, and

$$-\operatorname{tr} B^2 \geq 0 \quad \text{for} \quad B = -B^* \in \operatorname{End} V,$$

where $*$ stands for the Hermitian conjugation.

Another important example is when a real vector space $V = \mathfrak{g}$ and representation R is given by the adjoint action Ad of G on $\mathfrak{g}$. A Lie algebra $\mathfrak{g}$ carries Ad-invariant symmetric bilinear form — the Killing form — given by

$$\langle x, y \rangle = -\operatorname{tr}(\operatorname{ad}_x \operatorname{ad}_y), \quad x, y \in \mathfrak{g},$$

where $\operatorname{ad}_x \in \operatorname{End} \mathfrak{g}$ is given by the adjoint action, $\operatorname{ad}_x(y) = [x, y]$. The Killing form defines positive-definite inner product if and only if a Lie group G is compact and semi-simple. Corresponding Lie algebra $\mathfrak{g}$ of a semi-simple Lie group G is characterized by the property that there is a basis $x_1, \dots, x_n$ of $\mathfrak{g}$, such that in the adjoint representation the $n \times n$ matrices $X_a = \mathrm{ad}_{x_a}$ satisfy

$$\operatorname{tr}(X_a X_b) = -2\delta_{ab}, \quad a, b = 1, \dots, n. \tag{10.4}$$

Equivalently, there is a basis $x_1, \dots, x_n$ of $\mathfrak{g}$ such that corresponding structure constants t_{ab}^c,

$$[x_a, x_b] = \sum_{c=1}^{n} t_{ab}^c x_c,$$

are totally anti-symmetric. In case $\mathfrak{g} = \mathfrak{su}(2)$ such basis in the defining two-dimensional representation is given by the matrices

$$x_1 = \frac{1}{2i}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad x_2 = \frac{1}{2i}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad x_3 = \frac{1}{2i}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

where $t_{abc} = t_{ab}^c$ is totally anti-symmetric and $t_{123} = 1$. Corresponding matrices X_1, X_2, X_3 in the adjoint representation of $\mathfrak{su}(2)$ are given by (see Example 2.2 in Lecture 2)

$$X_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad X_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and

$$\langle x, y \rangle = -4 \operatorname{tr}_{\mathbb{C}^2}(xy).$$

The real vector bundle associated with a principal G-bundle P through the adjoint representation of a Lie group G on its Lie algebra $\mathfrak{g}$ is called an adjoint bundle and is denoted by $\mathrm{ad}\, P$. In case when (M, ds^2) is a Riemannian manifold, the Killing form defines on $\Omega^p(M, \mathrm{ad}\, P)$ an inner product

$$(\omega_1, \omega_2) = \int_M \langle \omega_1 \wedge \star\, \omega_2 \rangle \tag{10.5}$$

with the L^2-norm

$$\|\omega\|^2 = \int_M \langle \omega \wedge \star\, \omega \rangle.$$

The symmetry

$$(\omega_1, \omega_2) = (\omega_2, \omega_1) \tag{10.6}$$

follows from (8.15) in Lecture 8 and the cyclic property of the trace. The Yang-Mills functional is the L^2-norm of the curvature form $F(A) \in \Omega^2(M, \mathrm{ad}\, P)$,

$$S(A) = \frac{1}{4\pi}\|F(A)\|^2.$$

In physics applications M is a four-manifold with pseudo-Riemannian metric of signature $(+,-,-,-)$ and $E = \operatorname{ad} P$ for some principal G-bundle P, where G is compact semi-simple Lie group. Of special importance is the case $M = \mathbb{R}^4$ with Minkowski metric, and $\operatorname{ad} P = M \times \mathfrak{g}$. Introduce of $A = A_\mu dx^\mu$ and

$$F = \frac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu, \quad F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu} + [A_\mu A_\nu], \tag{10.7}$$

where[1] $A_\mu = A_\mu^a X_a, F_{\mu\nu} = F_{\mu\nu}^a X_a \in \mathfrak{g}$ and generators X_a satisfy (10.4). Corresponding Yang-Mills functional (10.2) takes the form

$$S(A) = \frac{1}{16\pi} \int_{\mathbb{R}^4} \langle F_{\mu\nu}, F^{\mu\nu} \rangle d^4\boldsymbol{x} = -\frac{1}{8\pi} \int_{\mathbb{R}^4} F_{\mu\nu}^a (F^a)^{\mu\nu} d^4\boldsymbol{x}, \tag{10.8}$$

where $F^{\mu\nu} = (F^a)^{\mu\nu} X_a$, and Yang-Mills equations (10.3) become

$$\nabla_\mu F^{\mu\nu} = \frac{\partial F^{\mu\nu}}{\partial x^\mu} + [A_\mu, F^{\mu\nu}] = 0. \tag{10.9}$$

Yang-Mills equations (10.7) and (10.9) generalize U(1)-invariant Maxwell equations to the case of non-abelian symmetry group G. In terms of the components equation (10.9) takes the form

$$\frac{\partial (F^a)^{\mu\nu}}{\partial x^\mu} + t_{bc}^a A_\mu^b (F^c)^{\mu\nu} = 0,$$

where t_{bc}^a are totally anti-symmetric structure constants of $\mathfrak{g}$.

Remark. In physics one uses $\nabla_\mu = \partial_\mu - gA_\mu$ for the covariant derivative, where g is a *coupling constant* of the theory. In Quantum Chromodynamics (QCD) on considers $G = \mathrm{SU}(3)$ in the adjoint representation, and corresponding components $A_\mu^a(\boldsymbol{x})$, $a = 1, \dots, 8$, are the *gluon fields*; corresponding *quark fields* are in the fundamental representation of SU(3). In our notation gluon part of the QCD Lagrangian is

$$\mathscr{L}(A) = -\frac{1}{4g^2} F_{\mu\nu}^a (F^a)^{\mu\nu}, \tag{10.10}$$

where $F_{\mu\nu}^a$ plays the role of gluon field strength tensor. Corresponding elementary particle — a gluon (or gauge boson) — is the exchange particle for the strong force between quarks. In the Standard Model of electroweak and strong interactions one uses the symmetry group $G = \mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1)$.

10.2. Self-duality equations

In the Riemannian case Yang-Mills equations do not have direct physical interpretation. However, in case when M is compact Riemannian four-manifold, these equations have a fundamental mathematical significance.

[1] Here summation over repeated indices is understood.

Recall that the first Pontryagin class of a real vector bundle $\operatorname{ad} P$ over M is defined by

$$p_1(\operatorname{ad} P) = -c_2(\operatorname{ad}_{\mathbb{C}} P) \in \check{H}^4(M, \mathbb{Z}),$$

where $\operatorname{ad}_{\mathbb{C}} P = \operatorname{ad} P \otimes_{\mathbb{R}} \mathbb{C}$ is a complex vector bundle. Since $c_1(\operatorname{ad}_{\mathbb{C}} P) = 0$, it is easy to see that $c_2(\operatorname{ad}_{\mathbb{C}} P)$ is represented by the differential form $\dfrac{1}{8\pi^2}\operatorname{tr}(F \wedge F)$, and the corresponding first Pontryagin number is

$$p_1 = -\frac{1}{8\pi^2}\int_M \operatorname{tr}(F \wedge F) \in \mathbb{Z}.$$

In case when M is a Riemannian four-manifold with the metric ds^2, we have a decomposition

$$\Omega^2(M, \operatorname{ad} P) = \Omega^2_+(M, \operatorname{ad} P) \oplus \Omega^2_-(M, \operatorname{ad} P) \tag{10.11}$$

according to the eigenspaces of the Hodge star operator $\star$ corresponding to the eigenvalues 1 and -1. Since operator $\star$ is symmetric with respect to inner product (10.5) in $\Omega^2(M, \operatorname{ad} P)$, these subspaces are orthogonal. Equivalently, for $F = F_+ + F_-$, where $F_\pm \in \Omega^2_\pm(M, \operatorname{ad} P)$,

$$(F_+, F_-) = -\int_M \langle F_+ \wedge F_- \rangle = -\int_M \langle F_- \wedge \star F_+ \rangle = -(F_-, F_+),$$

and it follows from (10.6) that $(F_+, F_-) = 0$.

The curvature $F \in \Omega^2(M, \operatorname{ad} P)$ is called self-dual or anti-self-dual, if $F \in \Omega^2_+(M, \Omega^2(M, \operatorname{ad} P))$ or $F \in \Omega^2_-(M, \Omega^2(M, \operatorname{ad} P)$ respectively,

$$\star F = \pm F.$$

Correspondingly, connection $\nabla = d + A$ on a real vector bundle $\operatorname{ad} P$ is called self-dual or anti-self-dual, if its curvature is self-dual or anti-self-dual. Curvature forms of self-dual or anti-self-dual connections satisfy Yang-Mills equations on a Riemannian four-manifold M automatically!

Using the orthogonality of F_+ and F_-, we obtain

$$S(A) = \frac{1}{4\pi}\|F(A)\|^2 = \frac{1}{4\pi}\left(\|F_+\|^2 + \|F_-\|^2\right)$$

and

$$\begin{aligned} p_1 &= -\frac{1}{8\pi^2}\int_M \operatorname{tr}((F_+ + F_-) \wedge (F_+ + F_-)) \\ &= -\frac{1}{8\pi^2}(F_+ + F_-, F_+ - F_-) \\ &= -\frac{1}{8\pi^2}\left(\|(F_+\|^2 - \|F_-\|^2\right). \end{aligned}$$

From here we obtain the inequalities

$$S(A) - 2\pi p_1 \geq \frac{1}{2\pi}\|F_+\|^2 \quad \text{and} \quad S(A) + 2\pi p_1 \geq \frac{1}{2\pi}\|F_-\|^2.$$

Thus we see that the absolute minima of the Yang-Mills action on $\mathscr{A}^G_{\mathrm{ad}\,P}$ are given by the self-dual connections in case $p_1 > 0$, by the anti-self-dual connections in case $p_1 < 0$ and by both these types in case $p_1 = 0$. Number p_1 in called the *instanton number*. Solutions of the self-dual Yang-Mills equations for $M = S^4$ and $G = \mathrm{SU}(2)$ in case $p_1 = k > 0$ form he instanton moduli space $\mathscr{M}_k$, a smooth manifold of dimension $8k - 3$.

10.3. Hitchin equations

Let G be a compact real form of a complex Lie group, and denote by $*$ corresponding anti-involution on a complex Lie algebra $\mathfrak{g}_{\mathbb{C}} = \mathfrak{g} \otimes_{\mathbb{R}} \mathbb{C}$. Consider the self-duality equations in a trivial bundle $\mathrm{ad}\,P$ over $\mathbb{R}^4$ with Euclidean metric. Corresponding connections are $\mathfrak{g}$-valued 1-form $A = A_\mu d^\mu$ on $\mathbb{R}^4$ with the curvature 2-form

$$F = \frac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu, \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu].$$

The self-duality equations $F = \star F$ take a simple form

$$F_{12} = F_{34}, \quad F_{13} = F_{42}, \quad F_{14} = F_{23}.$$

Suppose that A_μ do not depend on variables x^3 and x^4. Introducing the so-called *Higgs fields* — $\mathfrak{g}$-valued functions $\phi_1 = A_3$, $\phi_2 = A_4$ on $\mathbb{R}^2$ — we can rewrite the self-duality equations as

$$\begin{aligned} F_{12} = [\phi_1, \phi_2] &= F_{24}, \\ F_{13} = [\nabla_1, \phi_1] = [\phi_2, \nabla_1] &= F_{42}, \\ F_{14} = [\nabla_1, \phi_2] = [\nabla_2, \phi_1] &= F_{23}. \end{aligned}$$

Let

$$F = F_{12} = \partial_1 A_2 - \partial_2 A_1 + [A_1, A_2]$$

be the curvature form of a connection $d + A_1 dx^1 + A_2 dx^2$ on a trivial $\mathrm{ad}\,P$ bundle over $\mathbb{R}^2$. Introducing the complex Higgs field $\phi = \phi_1 - \sqrt{-1}\phi_2$, the above equations can be written as

$$F = \frac{\sqrt{-1}}{2}[\phi, \phi^*] \quad \text{and} \quad [\nabla_1 + \sqrt{-1}\nabla_2, \phi] = 0. \tag{10.12}$$

Put $z = x^1 + \sqrt{-1}x^2$ and introduce

$$\Phi = \tfrac{1}{2}\phi\, dz \in \Omega^{1,0}(\mathbb{C}, \mathrm{ad}_{\mathbb{C}}P), \quad \Phi^* = \tfrac{1}{2}\phi\, d\bar{z} \in \Omega^{0,1}(\mathbb{C}, \mathrm{ad}_{\mathbb{C}}P).$$

Introducing connection 1-form

$$A = A_1 dx^1 + A_2 dx^2 = A^{1,0} dz + A^{0,1} d\bar{z}$$

in the complex vector bundle $\mathrm{ad}_{\mathbb{C}}P$ over $\mathbb{C}$, we can rewrite equations (10.12) as

$$F + [\Phi, \Phi^*] = 0, \tag{10.13}$$

$$\bar{\partial}_A \Phi = 0. \tag{10.14}$$

Here

$$[\Phi, \Phi^*] = \Phi \wedge \Phi^* + \Phi^* \wedge \Phi$$

is a graded Lie bracket on $\mathrm{ad}_{\mathbb{C}}P$-valued 1-forms, and $\bar{\partial}_A$ is a $(0,1)$-component of

$$\nabla_A = \partial + A^{1,0}dz + \bar{\partial} + A^{0,1}d\bar{z} = \partial_A + \bar{\partial}_A.$$

It is remarkable that equations (10.13)–(10.14) make sense over a Riemann surface M! Namely, consider a principal G-bundle P over M, a connection A in the adjoint bundle $\mathrm{ad}\, P$ and the Higgs field $\Phi \in \Omega^{1,0}(M, \mathrm{ad}_{\mathbb{C}}P)$. The pair (A, Φ) satisfies self-duality equations over a Riemann surface M, if

$$F(A) + [\Phi, \Phi^*] = 0 \quad \text{and} \quad \bar{\partial}_A \Phi = 0. \tag{10.15}$$

The second equation states that Φ is a holomorphic section of the complex vector bundle $\mathrm{ad}\, P \otimes \Omega^{1,0}(M, \mathbb{C})$ with respect to the complex structure in $\mathrm{ad}_{\mathbb{C}}P$ determined by the Cauchy-Riemann operator $\bar{\partial}_A = \bar{\partial} + A^{0,1}$ and the natural complex structure in $\Omega^{1,0}(M, \mathbb{C})$. Solution (A, Φ) of the self-duality equations (10.15) determines a flat complex connection $d + A + \Phi + \Phi^*$ on $\mathrm{ad}_{\mathbb{C}}P$.

LECTURE 11

Electromagnetic waves in a free space

11.1. Energy-momentum tensor

Suppose that F satisfies Maxwell equations without sources. Using equations (8.9) and (8.12) we have

$$\begin{aligned}\frac{\partial}{\partial x^\alpha}(F_{\mu\nu}F^{\mu\nu}) &= \frac{\partial F_{\mu\nu}}{\partial x^\alpha}F^{\mu\nu} + F_{\mu\nu}\frac{\partial F^{\mu\nu}}{\partial x^\alpha} = 2\frac{\partial F_{\mu\nu}}{\partial x^\alpha}F^{\mu\nu} \\ &= -2\left(\frac{\partial F_{\alpha\mu}}{\partial x^\nu} + \frac{\partial F_{\nu\alpha}}{\partial x^\mu}\right)F^{\mu\nu} = -4\frac{\partial}{\partial x^\mu}(F_{\nu\alpha}F^{\mu\nu}).\end{aligned}$$

Thus

$$\frac{\partial}{\partial x^\alpha}(F_{\mu\nu}F^{\mu\nu}) = -4\frac{\partial}{\partial x^\beta}(F_{\nu\alpha}F^{\beta\nu}),$$

and introducing

$$T_\alpha^\beta = F_{\nu\alpha}F^{\beta\nu} + \frac{1}{4}\delta_\alpha^\beta F_{\mu\nu}F^{\mu\nu},$$

we can rewrite this equation as a conservation law

$$\frac{\partial T_\alpha^\beta}{\partial x^\beta} = 0, \quad \alpha = 0, 1, 2, 3. \tag{11.1}$$

The tensor T_α^β is traceless $T_\alpha^\alpha = 0$ and symmetric, $T^{\alpha\beta} = T^{\beta\alpha}$, where

$$T^{\alpha\beta} = \eta^{\alpha\gamma}T_\gamma^\beta = -\eta_{\mu\nu}F^{\alpha\mu}F^{\beta\nu} + \frac{1}{4}\eta^{\alpha\beta}F_{\mu\nu}F^{\mu\nu}. \tag{11.2}$$

The tensor $T^{\alpha\beta}$ is called the *energy-momentum tensor*. Its components contain the *energy density*

$$T^{00} = \frac{1}{2}\left(\frac{1}{c^2}\boldsymbol{E}^2 + \boldsymbol{B}^2\right)$$

and the *momentum density*

$$T^{0i} = F^{0k}F^{ik} = \frac{1}{c}(\boldsymbol{E}\times\boldsymbol{B})_i, \quad i = 1, 2, 3.$$

The vector $\boldsymbol{S} = \boldsymbol{E}\times\boldsymbol{B}$ is called the *Poynting vector*.

Remark. The conservation law (11.1)

$$\frac{\partial T^{00}}{\partial t} = -\boldsymbol{\nabla}\cdot\boldsymbol{S}$$

can be verified directly using Maxwell's equations and the calculus formula

$$\boldsymbol{\nabla}\cdot(\boldsymbol{a}\times\boldsymbol{b})=\boldsymbol{b}\cdot(\boldsymbol{\nabla}\times\boldsymbol{a})-\boldsymbol{a}\cdot(\boldsymbol{\nabla}\times\boldsymbol{b}).$$

It also implies that implies that the total energy of the electromagnetic field

$$\mathscr{E}=\frac{1}{4\pi}\int_{\{ct\}\times\mathbb{R}^3}T^{00}d^3\boldsymbol{r}$$

does not depend on time.

11.2. Gauge fixing

Maxwell equations in the empty space

$$dF=0\quad\text{and}\quad d*F=0 \tag{11.3}$$

describe harmonic 2-forms on $\mathbb{R}^4$. Their general solution is given by $F=dA$, where

$$*\,d*dA=0. \tag{11.4}$$

This equation is not hyperbolic: if $A\in\Omega^1(\mathbb{R}^4)$ is a solution then $A+df$ for any smooth function f on $\mathbb{R}^4$ is also a solution. However, one can impose an additional condition

$$d*A=0, \tag{11.5}$$

which turns (11.4) into the hyperbolic equation

$$\Box A=0,$$

where

$$\Box=d*d*+*d*d$$

is the D'Alambertian — the Laplace operator of the Minkowski metric on $\mathbb{R}^4$, acting on 1-forms. In terms of $A=A_\mu dx^\mu$ equation (11.5) becomes

$$\partial_\mu A^\mu=0,\quad\text{where}\quad\partial_\mu=\frac{\partial}{\partial x^\mu}, \tag{11.6}$$

and is called the *Lorenz*[1] *gauge* condition. Since equation (11.4) can be written as

$$\partial_\mu F^{\mu\nu}=0,\quad\text{where}\quad F^{\mu\nu}=\partial^\mu A^\nu-\partial^\nu A^\mu$$

(see (8.12) in Lecture 8), we readily obtain that in the Lorenz gauge

$$\Box A^\mu=0,\quad\mu=0,1,2,3,$$

[1]Named after Danish physicist and mathematician Ludvig Lorenz, not to be confused with Dutch physicist Hedrick Lorentz!

where

$$\square = \partial_\mu \partial^\mu = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2}.$$

For every $A \in \Omega^1(\mathbb{R}^4)$ there is a gauge equivalent 1-form $A^f = A + df$ satisfying the Lorentz condition. Indeed, (11.5) gives

$$*d*df = -*d*A.$$

Using (11.6), for the function f we get the hyperbolic equation

$$\square f = -\partial_\mu A^\mu.$$

Remark. Maxwell equations with sources in the Lorentz gauge have the form

$$\square A^\mu = J^\mu.$$

The Lorenz gauge is not unique: if A satisfies (11.6), so does A^f, where $\square f = 0$. In free and empty space one can make a unique choice by imposing $A_0 = 0$. In general, this gauge condition is called *Hamilton gauge*. Together with Lorenz gauge it yields the *Coulomb gauge*,

$$\boldsymbol{\nabla}\cdot\boldsymbol{A} = 0. \tag{11.7}$$

Indeed, we can always make $A_0 = 0$ by using A^f, where $\partial_0 f = -A_0$. The remaining gauge transformations preserving $A_0 = 0$ are of the form $A \mapsto A + d\chi$, where χ is independent of x^0. In the free and empty space $\rho = 0$ and since $\varphi = cA_0 = 0$, it follows from equation (8.6) in Lecture 8 that

$$0 = \boldsymbol{\nabla}\cdot\boldsymbol{E} = -\frac{\partial}{\partial t}(\boldsymbol{\nabla}\cdot\boldsymbol{A}),$$

whence $\boldsymbol{\nabla}\cdot\boldsymbol{A}$ does not depend on t. Determining χ from the elliptic equation

$$\Delta\chi = -\boldsymbol{\nabla}\cdot\boldsymbol{A}, \quad \text{where} \quad \Delta = \boldsymbol{\nabla}\cdot\boldsymbol{\nabla} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2},$$

we arrive at (11.7). Not that in the free and empty space in the Coulomb gauge we also have $A_0 = 0$. In the presence of electric charges Coulomb gauge condition is

$$-\boldsymbol{\nabla}^2 A_0 = \frac{c}{\varepsilon_0}\rho \quad \text{and} \quad \boldsymbol{\nabla}\cdot\boldsymbol{A} = 0,$$

where $\rho(\boldsymbol{r}, t)$ is the electric charge density.

To summarize, in the Coulomb gauge Maxwell equations in free and empty space take the form

$$\square\boldsymbol{A} = 0 \quad \text{and} \quad \boldsymbol{E} = -\frac{\partial\boldsymbol{A}}{\partial t}, \quad \boldsymbol{B} = \boldsymbol{\nabla}\times\boldsymbol{A}, \tag{11.8}$$

so that also

$$\square\boldsymbol{E} = 0 \quad \text{and} \quad \square\boldsymbol{B} = 0. \tag{11.9}$$

11.3. Plane waves

In the Coulomb gauge consider the case when potential $\boldsymbol{A}$ depends only on the coordinate x. The wave equation reduces to

$$\frac{\partial^2 \boldsymbol{A}}{\partial t^2} - c^2 \frac{\partial^2 \boldsymbol{A}}{\partial x^2} = 0$$

and has a general solution

$$\boldsymbol{A}(t,x) = \boldsymbol{A}_1\left(t - \frac{x}{c}\right) + \boldsymbol{A}_2\left(t + \frac{x}{c}\right).$$

The wave moving in a positive direction on the x-axis is

$$\boldsymbol{A}\left(t - \frac{x}{c}\right),$$

and the Coulomb gauge condition gives

$$\frac{\partial A_x}{\partial x} = 0.$$

Thus $A_x = at$, where a is a constant, which gives rise to a constant electric field in the x-direction. Since such a field has nothing to do with the electromagnetic wave, we can set $A_x = 0$. Introducing the direction of the wave — the unit vector $\boldsymbol{n} = \boldsymbol{e}_x$ — we obtain that always $\boldsymbol{A} \perp \boldsymbol{n}$. Correspondingly,

$$\boldsymbol{E} = -\boldsymbol{A}' \quad \text{and} \quad \boldsymbol{B} = -\frac{1}{c}\boldsymbol{n} \times \boldsymbol{A}' = \frac{1}{c}\boldsymbol{n} \times \boldsymbol{E},$$

where the prime indicates t-derivative. Thus the electric and magnetic fields are perpendicular to the direction of propagation of the wave, and corresponding electromagnetic plane waves are *transverse*. Moreover, the electric and magnetic fields are orthogonal, and their strengths are related by $E = cB$. The vectors $\boldsymbol{n}, \dfrac{\boldsymbol{E}}{E}, \dfrac{\boldsymbol{B}}{B}$ form an orthonormal positively oriented basis of $\mathbb{R}^3$.

The components of the energy-momentum tensor of a plane wave are given by

$$T^{00} = \frac{E^2}{c^2} \quad \text{and} \quad \boldsymbol{S} = \frac{1}{c^2}\boldsymbol{E} \times \boldsymbol{n} \times \boldsymbol{E} = \frac{E^2}{c^2}\boldsymbol{n},$$

so that $(T^{00})^2 = \boldsymbol{S}^2$.

A *monochromatic* wave which is a simply periodic function of t with a vector potential

$$\boldsymbol{A} = \operatorname{Re}\left\{\boldsymbol{A}_0 e^{-i\omega(t - \frac{x}{c})}\right\}.$$

Here $\boldsymbol{A}_0 \in \mathbb{C}^3$ is a constant complex vector, ω is the *frequency*, $\lambda = \dfrac{2\pi c}{\omega}$ is the *wave length* and $\boldsymbol{k} = \dfrac{\omega}{c}\boldsymbol{n}$ is the *wave vector*, where $\boldsymbol{n}$ is a unit vector in the direction of propagation of the wave (in our case $\boldsymbol{n} = \boldsymbol{e}_x$). We have

$$\boldsymbol{A} = \operatorname{Re}\left\{\boldsymbol{A}_0 e^{i(\boldsymbol{k}\cdot\boldsymbol{r} - \omega t)}\right\},$$

where $\boldsymbol{k}\cdot\boldsymbol{r}-\omega t$ is the *phase* of the wave. Correspondingly,

$$\boldsymbol{E}=\operatorname{Re}\left\{\boldsymbol{E}_0 e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)}\right\} \quad \text{and} \quad \boldsymbol{B}=\operatorname{Re}\left\{\boldsymbol{B}_0 e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)}\right\},$$

where

$$\boldsymbol{E}_0=i\omega\boldsymbol{A}_0 \quad \text{and} \quad \boldsymbol{B}_0=i\boldsymbol{k}\times\boldsymbol{A}_0.$$

Consider the vector $\boldsymbol{E}_0\in\mathbb{C}^3$ and put $\boldsymbol{b}=\boldsymbol{E}_0 e^{i\alpha}$, where $\boldsymbol{E}_0^2=\boldsymbol{E}_0\cdot\boldsymbol{E}_0=|\boldsymbol{E}_0|^2e^{-2i\alpha}$. Then $\boldsymbol{b}^2=\boldsymbol{b}\cdot\boldsymbol{b}=|\boldsymbol{E}_0|^2$ and

$$\boldsymbol{E}=\operatorname{Re}\left\{\boldsymbol{b}e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t-\alpha)}\right\}.$$

Putting $\boldsymbol{b}=\boldsymbol{b}_1+i\boldsymbol{b}_2$, where $\boldsymbol{b}_1,\boldsymbol{b}_2\in\mathbb{R}^3$, we have

$$\boldsymbol{b}^2=\boldsymbol{b}_1^2-\boldsymbol{b}_2^2+2i\boldsymbol{b}_1\cdot\boldsymbol{b}_2\in\mathbb{R},$$

so that $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ are orthogonal. Since $\boldsymbol{A}_0$ is orthogonal to the wave vector $\boldsymbol{k}$, vectors $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ are also orthogonal to $\boldsymbol{k}$.

Choosing the xyz coordinate axes along positively oriented orthogonal basis $\boldsymbol{k},\boldsymbol{b}_1,\pm\boldsymbol{b}_2$, we get

$$\begin{aligned} E_y &= b_1\cos(\omega t-\boldsymbol{k}\cdot\boldsymbol{r}-\alpha),\\ E_z &= \pm b_2\sin(\omega t-\boldsymbol{k}\cdot\boldsymbol{r}-\alpha),\end{aligned}$$

where $b_1=|\boldsymbol{b}_1|$ and $b_2=|\boldsymbol{b}_2|$. If b_1,b_2 are non-zero, we have

$$\frac{E_x^2}{b_1^2}+\frac{E_z^2}{b_2^2}=1,$$

so that at each point of the space the electric field vector $\boldsymbol{E}$ rotates in the plane perpendicular to the direction of propagation and describes an ellipse. Such wave is called *elliptically polarized*. If $b_1=b_2$, the wave is called *circularly polarized*, and in case b_1 or b_2 is zero, the wave is called *linearly polarized*.

Remark. Introduce the 4-vector $(k^\mu)=\left(\frac{\omega}{c},\boldsymbol{k}\right)$ and $(k_\mu)=\left(\frac{\omega}{c},-\boldsymbol{k}\right)$ with the property $k_\mu k^\mu=0$. We have $k_\mu x^\mu=\omega t-\boldsymbol{k}\cdot\boldsymbol{r}$, so that

$$\boldsymbol{A}(x)=\operatorname{Re}\left\{\boldsymbol{A}_0 e^{-ik_\mu x^\mu}\right\}.$$

The electromagnetic waves describe *photons*, particles with 4-wave vector satisfying $k_0^2=\boldsymbol{k}^2$.

11.4. The general solution

The Cauchy problem for equation (11.8) has the form

$$\begin{aligned} \Box\boldsymbol{A}&=0,\\ \boldsymbol{A}(0,\boldsymbol{r})&=\boldsymbol{A}_0(\boldsymbol{r}),\\ \frac{\partial\boldsymbol{A}}{\partial t}(0,\boldsymbol{r})&=\boldsymbol{A}_1(\boldsymbol{r}),\end{aligned}$$

where Cauchy data $\boldsymbol{A}_0(\boldsymbol{r})$ and $\boldsymbol{A}_1(\boldsymbol{r})$ satisfy Coulomb gauge condition

$$\boldsymbol{\nabla}\cdot\boldsymbol{A}_0 = 0 \quad\text{and}\quad \boldsymbol{\nabla}\cdot\boldsymbol{A}_1 = 0$$

and rapidly decay as $|\boldsymbol{r}|\to\infty$.

Cauchy problem for the wave equation in $\mathbb{R}^4$ is solved by the Fourier transform. Namely, let

$$\begin{aligned}\boldsymbol{A}_0(\boldsymbol{r}) &= \frac{1}{(2\pi)^{\frac{3}{2}}}\int_{\mathbb{R}^3} e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\boldsymbol{a}_0(\boldsymbol{k})d^3\boldsymbol{k},\\ \boldsymbol{A}_1(\boldsymbol{r}) &= \frac{1}{(2\pi)^{\frac{3}{2}}}\int_{\mathbb{R}^3} e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\boldsymbol{a}_1(\boldsymbol{k})d^3\boldsymbol{k},\end{aligned}$$

where $\boldsymbol{a}_0(\boldsymbol{k}) = \bar{\boldsymbol{a}}_0(-\boldsymbol{k})$, $\boldsymbol{a}_1(\boldsymbol{k}) = \bar{\boldsymbol{a}}_1(-\boldsymbol{k})$ and $\boldsymbol{k}\cdot\boldsymbol{a}_0(\boldsymbol{k}) = \boldsymbol{k}\cdot\boldsymbol{a}_1(\boldsymbol{k}) = 0$. The solution is given by

$$\boldsymbol{A}(t,\boldsymbol{r}) = \frac{1}{(2\pi)^{\frac{3}{2}}}\int_{\mathbb{R}^3} e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\boldsymbol{a}(t,\boldsymbol{k})d^3\boldsymbol{k}, \tag{11.10}$$

where

$$\boldsymbol{a}(t,\boldsymbol{k}) = \cos(c|\boldsymbol{k}|t)\boldsymbol{a}_0(\boldsymbol{k}) + \frac{\sin(c|\boldsymbol{k}|t)}{c|\boldsymbol{k}|}\boldsymbol{a}_1(\boldsymbol{k}).$$

Introducing

$$\boldsymbol{a}(\boldsymbol{k}) = \frac{1}{2}\boldsymbol{a}_0(\boldsymbol{k}) + \frac{1}{2ic|\boldsymbol{k}|}a_1(\boldsymbol{k}),$$

we can rewrite (11.10) as

$$\boldsymbol{A}(t,\boldsymbol{r}) = \frac{1}{(2\pi)^3}\int_{\mathbb{R}^3}\left(e^{-i(\omega_{\boldsymbol{k}}t-\boldsymbol{k}\cdot\boldsymbol{r})}\boldsymbol{a}(\boldsymbol{k}) + e^{i(\omega_{\boldsymbol{k}}t-\boldsymbol{k}\cdot\boldsymbol{r})}\bar{\boldsymbol{a}}(\boldsymbol{k})\right)d^3\boldsymbol{k}, \tag{11.11}$$

where $\omega_{\boldsymbol{k}} = c|\boldsymbol{k}|$. For electric and magnetic fields we have

$$\begin{aligned}\boldsymbol{E} &= -\frac{\partial\boldsymbol{A}}{\partial t}\\ &= \frac{i}{(2\pi)^{\frac{3}{2}}}\int_{\mathbb{R}^3}\omega_{\boldsymbol{k}}\left(e^{-i(\omega_{\boldsymbol{k}}t-\boldsymbol{k}\cdot\boldsymbol{r})}\boldsymbol{a}(\boldsymbol{k}) - e^{i(\omega_{\boldsymbol{k}}t-\boldsymbol{k}\cdot\boldsymbol{r})}\bar{\boldsymbol{a}}(\boldsymbol{k})\right)d^3\boldsymbol{k}\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{B} &= \boldsymbol{\nabla}\times\boldsymbol{A}\\ &= \frac{i}{(2\pi)^{\frac{3}{2}}}\int_{\mathbb{R}^3}\boldsymbol{k}\times\left(e^{-i(\omega_{\boldsymbol{k}}t-\boldsymbol{k}\cdot\boldsymbol{r})}\boldsymbol{a}(\boldsymbol{k}) - e^{i(\omega_{\boldsymbol{k}}t-\boldsymbol{k}\cdot\boldsymbol{r})}\bar{\boldsymbol{a}}(\boldsymbol{k})\right)d^3\boldsymbol{k}.\end{aligned}$$

By Plancherel theorem we have for total energy of the electromagnetic field,

$$\begin{aligned}\mathscr{E} &= \frac{1}{8\pi}\int_{\mathbb{R}^3}\left(\frac{1}{c^2}\boldsymbol{E}^2 + \boldsymbol{B}^2\right)d^3\boldsymbol{r}\\ &= \frac{1}{4\pi}\int_{\mathbb{R}^3}(\omega_{\boldsymbol{k}}^2\,\boldsymbol{a}(\boldsymbol{k})\bar{\boldsymbol{a}}(\boldsymbol{k}) + (\boldsymbol{k}\times\boldsymbol{a}(\boldsymbol{k}))\cdot(\boldsymbol{k}\times\bar{\boldsymbol{a}}(\boldsymbol{k}))d^3\boldsymbol{k}\\ &= \frac{1}{2\pi c^2}\int_{\mathbb{R}^3}\omega_{\boldsymbol{k}}^2\,\boldsymbol{a}(\boldsymbol{k})\cdot\bar{\boldsymbol{a}}(\boldsymbol{k})d^3\boldsymbol{k},\end{aligned}$$

where we have used the identity $(\boldsymbol{k}\times\boldsymbol{a}(\boldsymbol{k}))\cdot(\boldsymbol{k}\times\bar{\boldsymbol{a}}(\boldsymbol{k})) = |\boldsymbol{k}|^2\boldsymbol{a}(\boldsymbol{k})\cdot\bar{\boldsymbol{a}}(\boldsymbol{k})$, which follows from $\boldsymbol{k}\cdot\boldsymbol{a}(\boldsymbol{k}) = 0$.

Similarly,

$$\begin{aligned}\frac{1}{4\pi}\int_{\mathbb{R}^3}\boldsymbol{S}d^3\boldsymbol{r} &= \frac{1}{4\pi c}\int_{\mathbb{R}^3}(\boldsymbol{E}\times\boldsymbol{B})d^3\boldsymbol{r}\\ &= \frac{1}{2\pi c}\int_{\mathbb{R}^3}\omega_{\boldsymbol{k}}\,\boldsymbol{a}(\boldsymbol{k})\times(\boldsymbol{k}\times\bar{\boldsymbol{a}}(\boldsymbol{k}))d^3\boldsymbol{k}\\ &= \frac{1}{2\pi}\int_{\mathbb{R}^3}\omega_{\boldsymbol{k}}\,(\boldsymbol{a}(\boldsymbol{k})\cdot\bar{\boldsymbol{a}}(\boldsymbol{k}))\boldsymbol{k}d^3\boldsymbol{k}.\end{aligned}$$

Finally, putting

$$\boldsymbol{P}(\boldsymbol{k}) = \frac{\omega_{\boldsymbol{k}}}{2c\sqrt{\pi}}(\boldsymbol{a}(\boldsymbol{k})+\bar{\boldsymbol{a}}(\boldsymbol{k}))\quad \boldsymbol{Q}(\boldsymbol{k}) = \frac{i}{2c\sqrt{\pi}}(\boldsymbol{a}(\boldsymbol{k})-\bar{\boldsymbol{a}}(\boldsymbol{k})),$$

we obtain a representation of the energy and momentum of electromagnetic field in terms of the oscillators

$$\begin{aligned}\frac{1}{8\pi}\int_{\mathbb{R}^3}\left(\frac{1}{c^2}\boldsymbol{E}^2+\boldsymbol{B}^2\right)d^3\boldsymbol{r} &= \frac{1}{2}\int_{\mathbb{R}^3}(\boldsymbol{P}^2(\boldsymbol{k})+\omega_{\boldsymbol{k}}^2\boldsymbol{Q}^2(\boldsymbol{k}))d^3\boldsymbol{k}\\ \frac{1}{4\pi c}\int_{\mathbb{R}^3}(\boldsymbol{E}\times\boldsymbol{B})d^3\boldsymbol{r} &= \frac{c}{2}\int_{\mathbb{R}^3}(\omega_{\boldsymbol{k}}^{-1}\boldsymbol{P}^2(\boldsymbol{k})+\omega_{\boldsymbol{k}}\boldsymbol{Q}^2(\boldsymbol{k}))\boldsymbol{k}d^3\boldsymbol{k},\end{aligned}$$

where the normal modes $\boldsymbol{P}(\boldsymbol{k})$ and $\boldsymbol{Q}(\boldsymbol{k})$ satisfy

$$\boldsymbol{k}\cdot\boldsymbol{P}(\boldsymbol{k}) = \boldsymbol{k}\cdot\boldsymbol{Q}(\boldsymbol{k}) = 0.$$

LECTURE 12

Hamiltonian formalism. Real scalar field

Here we consider four-dimensional spacetime $\mathbb{R}^4$ with coordinates $x = (x^0, x^1, x^2, x^3)$ and Minkowski metric $(dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2$. We put $c = 1$ so that $x^0 = t$.

12.1. Lagrangian formulation

The scalar field $\varphi(x)$ is a smooth real-valued function on $\mathbb{R}^4$ of the Schwartz class for each time slice $t = t_0$. The corresponding Lagrangian function has the form

$$\mathscr{L}(\varphi(x), \partial_\mu\varphi(x)) = \frac{1}{2}\left(\partial_\mu\varphi(x)\partial^\mu\varphi(x) - m^2\varphi(x)\right) - V_{\mathrm{int}}(\varphi(x)),$$

where

$$\partial_\mu\varphi = \frac{\partial\varphi}{\partial x^\mu}, \quad \mu = 0, 1, 2, 3.$$

In particular, $V_{\mathrm{int}}(\varphi) = 0$ corresponds to the Klein-Gordon model, and $V_{\mathrm{int}}(\varphi) = g\varphi^4/4!$ — to the φ^4–model.

The action functional

$$S(\varphi) = \int \mathscr{L}(\varphi, \partial_\mu\varphi)d^4x,$$

where integration goes over the part of $\mathbb{R}^4$ between the slices $t = t_0$ and $t = t_1$ with fixed $\varphi(t_0, \boldsymbol{x}) = \varphi_0(\boldsymbol{x})$ and $\varphi(t_1, \boldsymbol{x}) = \varphi_1(\boldsymbol{x})$, or over $\mathbb{R}^4$, where $\varphi(x)$ is assumed to be rapidly decaying as $|x| \to \infty$. Corresponding Euler-Lagrange equation $\delta S = 0$ takes the form

$$\frac{\partial\mathscr{L}}{\partial\varphi} - \frac{\partial}{\partial x^\mu}\frac{\partial\mathscr{L}}{\partial(\partial_\mu\varphi)} = 0 \tag{12.1}$$

and yields equation of motion of the massive real scalar field

$$(\Box + m^2)\varphi + V'_{\mathrm{int}}(\varphi) = 0. \tag{12.2}$$

For the φ^4–model this equation takes the form

$$(\Box + m^2)\varphi + g\frac{\varphi^3}{3!} = 0,$$

and is a nonlinear Klein-Gordon equation with cubic nonlinearity.

Remark. Let $\mathscr{F}$ be the space of scalar fields on $\mathbb{R}^4$. The Lagrangian L is map from $\mathscr{F}$ to the functions on $\mathbb{R}^4$ such that $L(\varphi)(x)$ depends only on the 1-jet of φ at $x \in \mathbb{R}^4$, i.e., $L(\varphi)(x) = \mathscr{L}(\varphi(x), \partial_\mu\varphi(x))$.

12.2. The energy-momentum tensor

Since the Lagrangian function does not depend explicitly on x, we have

$$\begin{aligned}\partial_\nu \mathscr{L} &= \frac{\partial \mathscr{L}}{\partial \varphi}\partial_\nu\varphi + \frac{\partial \mathscr{L}}{\partial(\partial_\mu\varphi)}\partial_\nu\partial_\mu\varphi \\ &= \left(\frac{\partial \mathscr{L}}{\partial \varphi} - \partial_\mu \frac{\partial \mathscr{L}}{\partial(\partial_\mu\varphi)}\right)\partial_\nu\varphi + \partial_\mu\left(\frac{\partial \mathscr{L}}{\partial(\partial_\mu\varphi)}\partial_\nu\varphi\right).\end{aligned}$$

Thus on the solutions of the Euler-Lagrange equation (12.1) we have

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

or

$$\partial_\mu T^\mu_\nu = 0, \tag{12.3}$$

where

$$T^\mu_\nu = \frac{\partial \mathscr{L}}{\partial(\partial_\mu\varphi)}\partial_\nu\varphi - \delta^\mu_\nu \mathscr{L}$$

is the *energy-momentum tensor*. The tensor $T^{\mu\nu} = \eta^{\nu\lambda}T^\mu_\lambda$ satisfies the conservation law

$$\partial_\mu T^{\mu\nu} = 0,$$

and is defined up to the addition of $\partial_\sigma \Psi^{\mu\nu\sigma}$, where $\Psi^{\mu\nu\sigma} = -\Psi^{\mu\sigma\nu}$.

For the scalar field the tensor $T^{\mu\nu} = \partial^\mu\varphi\partial^\nu\varphi - \eta^{\mu\nu}\mathscr{L}$ is symmetric and

$$\begin{aligned}T^{00} &= \frac{1}{2}\left((\partial_0\varphi)^2 + (\boldsymbol{\nabla}\varphi)^2 + m^2\varphi^2 + V_{\text{int}}(\varphi)\right), \\ T^{0k} &= \partial^0\varphi\partial^k\varphi, \quad T^{ij} = \partial^i\varphi\partial^j\varphi.\end{aligned}$$

Conservation law for the energy-momentum vector $(h, \boldsymbol{p})$, where $h = T^{00}$ and $\boldsymbol{p} = (T^{01}, T^{02}, T^{03})$ reads

$$\frac{\partial h}{\partial t} + \boldsymbol{\nabla}\cdot\boldsymbol{p} = 0.$$

For the electromagnetic field $\mathscr{L} = -\frac{1}{16\pi}F_{\mu\nu}F^{\mu\nu}$, and the tensor

$$\frac{\partial \mathscr{L}}{\partial(\partial_\mu A_\sigma)}\partial^\nu A_\sigma - \eta^{\mu\nu}\mathscr{L}$$

is no longer symmetric. Adding to it

$$-\frac{1}{4\pi}\partial_\sigma\left(A^\nu F^{\sigma\mu}\right) = -\frac{1}{4\pi}\partial_\sigma A^\nu F^{\sigma\mu}$$

(remember that equations of motion are used!), we get the energy-momentum tensor discussed in Sect. 11.1 of Lecture 11.

REMARK. In physics textbooks one proves (12.3) by using the invariance of the action functional under the translations $x \mapsto \tilde{x} = x + a$,

$$\int_{\tilde{V}} \mathscr{L}(\tilde{\varphi}, \partial_\mu \tilde{\varphi}) d^4\tilde{x} - \int_V \mathscr{L}(\varphi, \partial_\mu \varphi) d^4 x = 0,$$

where $\tilde{\varphi}(\tilde{x}) = \varphi(x)$, $\tilde{V} = V + a$ for arbitrary domain $V \subset \mathbb{R}^4$, and expressing the resulting zero as the variation of the action with $\delta\varphi = \partial_\mu a^\mu$ using the Stokes' theorem and that $\varphi(x)$ satisfies Euler-Lagrange equations.

12.3. Hamiltonian formulation

As in classical mechanics, let

$$\pi(x) = \frac{\partial \mathscr{L}}{\partial(\partial_0 \varphi(x))} = \partial_0 \varphi(x)$$

be canonically conjugated momentum to the field $\varphi(x)$, and define the Hamiltonian functional density $\mathscr{H}(\pi, \varphi)$ by the Legendre transform

$$\begin{aligned}\mathscr{H}(\pi(x), \varphi(x)) &= \pi^2(x) - \mathscr{L}(\varphi(x), \partial_\mu \varphi(x))\big|_{\partial_0\varphi = \pi} \\ &= \frac{1}{2}\left(\pi^2(x) + (\boldsymbol{\nabla}\varphi(x))^2 + m^2\varphi^2(x)\right) + V_{\mathrm{int}}(\varphi(x)).\end{aligned}$$

Equations of motion of the theory are Hamiltonian equations for the infinite-dimensional Hamiltonian system $(\mathscr{M}, \Omega, H)$ with the phase space $\mathscr{M} = \mathscr{S}(\mathbb{R}^3, \mathbb{R}) \times \mathscr{S}(\mathbb{R}^3, \mathbb{R})$, the symplectic form

$$\Omega = \int_{\mathbb{R}^3} \left(d\pi(\boldsymbol{x}) \wedge d\varphi(\boldsymbol{x})\right) d^3\boldsymbol{x},$$

and the Hamiltonian functional

$$H = \int_{\mathbb{R}^3} \mathscr{H} d^3\boldsymbol{x}.$$

REMARK. The Schwartz space $\mathscr{S}(\mathbb{R}^3)$ is a Fréchet space with the topology defined by the system of the semi-norms

$$\|f\|_{\alpha,\beta} = \sup_{\boldsymbol{x}\in\mathbb{R}^3} |\boldsymbol{x}^\alpha D^\beta f(\boldsymbol{x})|$$

for all multi-indices $\alpha, \beta \in \mathbb{Z}^3_{\geq 0}$. The symplectic form Ω is continuous skew-symmetric bilinear form on $\mathscr{M}$ defined by

$$\Omega\left((\pi_1, \varphi_1), (\pi_2, \varphi_2)\right) = \int_{\mathbb{R}^3} \left(\pi_1(\boldsymbol{x})\varphi_2(\boldsymbol{x}) - \pi_2(\boldsymbol{x})\varphi_1(\boldsymbol{x})\right) d^3\boldsymbol{x}.$$

The symplectic form Ω is (weakly) non-degenerate: $\Omega\left((\pi_1, \varphi_1), (\pi_2, \varphi_2)\right) = 0$ for all $(\pi_2, \varphi_2) \in \mathscr{M}$ implies $(\pi_1, \varphi_1) = 0$.

Darboux coordinates on $\mathscr{M}$ are $\pi(\boldsymbol{x}), \varphi(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^3$, and canonical Hamilton's equations

$$\partial_0 \pi(t, \boldsymbol{x}) = -\frac{\delta H}{\delta \varphi(\boldsymbol{x})}(\pi(t, \boldsymbol{x}), \varphi(t, \boldsymbol{x})), \tag{12.4}$$

$$\partial_0 \varphi(t, \boldsymbol{x}) = \frac{\delta H}{\delta \pi(\boldsymbol{x})}(\pi(t, \boldsymbol{x}), \varphi(t, \boldsymbol{x})) \tag{12.5}$$

give equation (12.2). Indeed, by calculus of variations we obtain

$$\frac{\delta H}{\delta \pi(\boldsymbol{x})}(\pi(x), \varphi(x)) = \pi(x)$$

and

$$\frac{\delta H}{\delta \varphi(\boldsymbol{x})}(\pi(x), \varphi(x)) = -\Delta\varphi(x) + m^2\varphi(x) + V'_{\text{int}}(\varphi(x)),$$

so that (12.4)–(12.5) yield

$$\partial_0^2 \varphi(x) = \Delta\varphi(x) - m^2\varphi(x) - V'_{\text{int}}(\varphi(x)).$$

To make these arguments rigorous, we need to define the algebra $\mathscr{A}$ of classical observables on $\mathscr{M}$.

Definition. A functional $F : \mathscr{M} \to \mathbb{R}$ is called *real-analytic* if $F(\varphi)$ for all $\varphi \in \mathscr{M}$ is represented by the absolutely convergent series

$$\begin{aligned} F(\pi, \varphi) = \sum_{m,n=0}^{\infty} \frac{1}{m!n!} \int_{\mathbb{R}^3} \cdots \int_{\mathbb{R}^3} c_{mn}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_m; \boldsymbol{y}_1, \dots, \boldsymbol{y}_n) \times \\ \times \pi(\boldsymbol{x}_1) \cdots \pi(\boldsymbol{x}_m) \varphi(\boldsymbol{y}_1) \cdots \varphi(\boldsymbol{y}_n) d^3\boldsymbol{x}_1 \cdots d^3\boldsymbol{x}_m d^3\boldsymbol{y}_1 \cdots d^3\boldsymbol{y}_n, \end{aligned}$$

where $c_{00} = c$ — a constant, and tempered distributions

$$c_{mn}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_m; \boldsymbol{y}_1, \dots, \boldsymbol{y}_n) \in \mathscr{S}(\underbrace{\mathbb{R}^3 \times \cdots \times \mathbb{R}^3}_{m+n})'$$

are independently symmetric with the respect to the variables $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m$ and $\boldsymbol{y}_1, \dots, \boldsymbol{y}_n$.

Definition. The real-analytic functional F is called *admissible*, if the variational derivatives

$$\begin{aligned} \frac{\delta F}{\delta \pi(\boldsymbol{x})} = \sum_{m=1}^{\infty} \sum_{n=0}^{\infty} \frac{1}{(m-1)!n!} \int_{\mathbb{R}^3} \cdots \int_{\mathbb{R}^3} c_{mn}(\boldsymbol{x}, \boldsymbol{x}_2, \dots, \boldsymbol{x}_m; \boldsymbol{y}_1, \dots, \boldsymbol{y}_n) \times \\ \times \pi(\boldsymbol{x}_2) \cdots \pi(\boldsymbol{x}_m) \varphi(\boldsymbol{y}_1) \cdots \varphi(\boldsymbol{y}_n) d^3\boldsymbol{x}_2 \cdots d^3\boldsymbol{x}_m d^3\boldsymbol{y}_1 \cdots d^3\boldsymbol{y}_n \end{aligned}$$

and

$$\begin{aligned} \frac{\delta F}{\delta \varphi(\boldsymbol{x})} = \sum_{m=0}^{\infty} \sum_{n=1}^{\infty} \frac{1}{m!(n-1)!} \int_{\mathbb{R}^3} \cdots \int_{\mathbb{R}^3} c_{mn}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_m; \boldsymbol{x}, \boldsymbol{y}_2, \dots, \boldsymbol{y}_n) \times \\ \times \pi(\boldsymbol{x}_1) \cdots \pi(\boldsymbol{x}_m) \varphi(\boldsymbol{y}_2) \cdots \varphi(\boldsymbol{y}_n) d^3\boldsymbol{x}_1 \cdots d^3\boldsymbol{x}_m d^3\boldsymbol{y}_2 \cdots d^3\boldsymbol{y}_n \end{aligned}$$

belong to the Schwarz class $\mathscr{S}(\mathbb{R}^3)$.

Clearly the product of admissible functionals is an admissible functional.

REMARK. For every real-analytic functional $F : \mathscr{M} \to \mathbb{R}$ its differential dF at every point $(\pi, \varphi) \in \mathscr{M}$ is a continuous linear map $dF : \mathscr{M} \to \mathbb{R}$, so that $dF \in \mathscr{S}(\mathbb{R}^3 \times \mathbb{R}^3)'$. A functional F is admissible if $dF \in \mathscr{S}(\mathbb{R}^3 \times \mathbb{R}^3)$, i.e. there exist Schwartz class functions, denoted by $\dfrac{\delta F}{\delta \pi(\boldsymbol{x})}$ and $\dfrac{\delta F}{\delta \varphi(\boldsymbol{x})}$, such that

$$dF(u, v) = \int_{\mathbb{R}^3} \left(\frac{\delta F}{\delta \pi(\boldsymbol{x})} u(\boldsymbol{x}) + \frac{\delta F}{\delta \varphi(\boldsymbol{x})} v(\boldsymbol{x}) \right) d^3\boldsymbol{x}$$

for all $(u, v) \in \mathscr{M}$.

REMARK. Condition that F is admissible means that for all $m, n \geq 0$ and $\pi_1, \dots, \varphi_m, \varphi_1, \dots, \varphi_n \in \mathscr{S}(\mathbb{R}^3)$ the distributions

$$c_{mn}(\pi_2 \otimes \cdots \otimes \pi_m \otimes \varphi_1 \otimes \cdots \otimes \varphi_n) \in \mathscr{S}(\mathbb{R}^3)'$$

and

$$c_{mn}(\pi_1 \otimes \cdots \otimes \pi_m \otimes \varphi_2 \otimes \cdots \otimes \varphi_n) \in \mathscr{S}(\mathbb{R}^3)'$$

are represented by the Schwarz class functions.

DEFINITION. The algebra $\mathscr{A}$ of classical observables on $\mathscr{M}$ is the algebra of all admissible functionals on $\mathscr{M}$.

The following result provides a rigorous foundation for the Hamiltonian mechanics with the infinite-dimensional phase space $\mathscr{M}$.

LEMMA 12.1. *The symplectic form Ω endows $\mathscr{A}$ with the Poisson algebra structure given by the Poisson bracket*

$$\{F, G\}(\pi, \varphi) = \int_{\mathbb{R}^3} \left(\frac{\delta F}{\delta \pi(\boldsymbol{x})} \frac{\delta G}{\delta \varphi(\boldsymbol{x})} - \frac{\delta F}{\delta \varphi(\boldsymbol{x})} \frac{\delta G}{\delta \pi(\boldsymbol{x})} \right) d^3\boldsymbol{x}, \tag{12.6}$$

where variational derivatives are evaluated at $(\pi, \varphi) \in \mathscr{M}$.

PROOF. It follows from the definition of real-analytic functionals and the above remark that $\{F, G\} \in \mathscr{A}$ for $F, G \in \mathscr{A}$. As in case of the canonical Poisson bracket on $\mathbb{R}^{2n}$ (see Sect. 5.3 in Lecture 5) the Jacobi identity for the bracket given by (12.6) is proved by a direct computation. $\square$

The Darboux coordinates $\pi(\boldsymbol{x}), \varphi(\boldsymbol{x})$, considered as evaluation functionals of (π, φ) at $\boldsymbol{x} \in \mathbb{R}^3$, do not belong to $\mathscr{A}$. Nevertheless, we have in the distributional sense,

$$\frac{\delta \pi(\boldsymbol{x})}{\delta \pi(\boldsymbol{y})} = \delta(\boldsymbol{x} - \boldsymbol{y}), \quad \frac{\delta \pi(\boldsymbol{x})}{\delta \varphi(\boldsymbol{y})} = 0 \quad \text{and} \quad \frac{\delta \varphi(\boldsymbol{x})}{\delta \pi(\boldsymbol{y})} = 0, \quad \frac{\delta \varphi(\boldsymbol{x})}{\delta \varphi(\boldsymbol{y})} = \delta(\boldsymbol{x} - \boldsymbol{y}),$$

and it follows from (12.6) that

$$\{F, \pi(\boldsymbol{x})\} = -\frac{\delta F}{\delta \varphi(\boldsymbol{x})} \quad \text{and} \quad \{F, \varphi(\boldsymbol{x})\} = \frac{\delta F}{\delta \pi(\boldsymbol{x})}.$$

Since for $F \in \mathscr{A}$

$$\partial_0 F(\pi, \varphi) = \int_{\mathbb{R}^3} \left(\frac{\delta F(\pi, \varphi)}{\delta \pi(\boldsymbol{x})} \partial_0 \pi(t, \boldsymbol{x}) + \frac{\delta F(\pi, \varphi)}{\delta \varphi(\boldsymbol{x})} \partial_0 \varphi(t, \boldsymbol{x}) \right) d^3\boldsymbol{x},$$

Hamilton's equations for smooth observables

$$\partial_0 F = \{H, F\}$$

are equivalent to canonical Hamilton's equations (12.4)–(12.5).

Remark. In physics textbooks, Poisson structure (12.6) on $\mathscr{A}$ is defined by the following Poisson brackets

$$\{\pi(\boldsymbol{x}), \pi(\boldsymbol{y})\} = \{\varphi(\boldsymbol{x}), \varphi(\boldsymbol{y})\} = 0 \quad \text{and} \quad \{\pi(\boldsymbol{x}), \varphi(\boldsymbol{y})\} = \delta(\boldsymbol{x} - \boldsymbol{y}), \tag{12.7}$$

understood in the distributional sense.

12.4. Fourier modes for the Klein-Gordon model

The Klein-Gordon equation

$$(\Box + m^2)\varphi(x) = 0 \tag{12.8}$$

in terms of the Fourier transform

$$\hat{\varphi}(k) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^4} e^{ik\cdot x} \varphi(x) d^4x, \quad \text{where} \quad k \cdot x = k^\mu x_\mu = k^0 x^0 - \boldsymbol{k}\boldsymbol{x},$$

takes the form

$$(k^2 - m^2)\hat{\varphi}(k) = 0.$$

Its general solution is a distribution supported on the two-sheeted mass hyperboloid $k^2 = (k^0)^2 - \boldsymbol{k}^2 = m^2$, which can be written as

$$\hat{\varphi}(k) = \delta(k^2 - m^2)\rho(k).$$

Here

$$\rho(k) = \theta(k^0)\rho_1(\boldsymbol{k}) + \theta(-k^0)\rho_2(\boldsymbol{k}),$$

where $\theta(k^0)$ is the Heavyside function and ρ_1, ρ_2 are distributions supported on $\mathbb{R}^3$. By definition of the distribution $\delta(k^2 - m^2) = \delta((k^0)^2 - \omega_{\boldsymbol{k}}^2)$, where $\omega_{\boldsymbol{k}} = \sqrt{\boldsymbol{k}^2 + m^2} > 0$, for a test function $u(\boldsymbol{k}) \in \mathscr{S}(\mathbb{R}^4)$ we have

$$(\theta(k^0)\rho_1(\boldsymbol{k})\delta(k^2 - m^2), u) = (\rho_1(\boldsymbol{k}), u_1),$$
$$(\theta(-k^0)\rho_2(\boldsymbol{k})\delta(k^2 - m^2), u) = (\rho_1(\boldsymbol{k}), u_2),$$

where

$$u_1(\boldsymbol{k}) = \frac{u(\omega_{\boldsymbol{k}}, \boldsymbol{k})}{2\omega_{\boldsymbol{k}}}, \quad u_2(\boldsymbol{k}) = \frac{u(-\omega_{\boldsymbol{k}}, \boldsymbol{k})}{2\omega_{\boldsymbol{k}}} \in \mathscr{S}(\mathbb{R}^3).$$

Whence

$$\hat{\varphi}(k)=\frac{1}{2\omega_{\boldsymbol{k}}}\rho_1(\boldsymbol{k})\delta(k^0-\omega_{\boldsymbol{k}})+\frac{1}{2\omega_{\boldsymbol{k}}}\rho_2(\boldsymbol{k})\delta(k^0+\omega_{\boldsymbol{k}}),$$

where reality condition $\overline{\rho(k)}=\rho(-k)$ gives $\rho_2(\boldsymbol{k})=\overline{\rho_1(-\boldsymbol{k})}$.

Substituting this $\hat{\varphi}(k)$ into the inverse Fourier transform

$$\varphi(x)=\frac{1}{(2\pi)^2}\int_{\mathbb{R}^4}e^{-ik\cdot x}\hat{\varphi}(k)d^4k,$$

introducing $a(\boldsymbol{k})=\sqrt{2\pi}\rho_1(\boldsymbol{k})$, $\bar{a}(\boldsymbol{k})=\overline{a(\boldsymbol{k})}$ and changing in the second integral $\boldsymbol{k}$ by $-\boldsymbol{k}$ we obtain

$$\varphi(x)=\frac{1}{(2\pi)^{\frac{3}{2}}}\int_{\mathbb{R}^3}\left(a(\boldsymbol{k})e^{-ik\cdot x}+\bar{a}(\boldsymbol{k})e^{ik\cdot x}\right)\frac{d^3\boldsymbol{k}}{2\omega_{\boldsymbol{k}}},\quad \text{where}\quad k^0=\omega_{\boldsymbol{k}}. \tag{12.9}$$

From this general distributional solution we can obtain a solution of the Cauchy problem for the Klein-Gordon equation, which consists in finding a solution $\varphi(x)$ of (12.8) satisfying

$$\varphi(0,\boldsymbol{x})=\varphi(\boldsymbol{x})\quad\text{and}\quad\partial_0\varphi(0,\boldsymbol{x})=\pi(\boldsymbol{x}).$$

Namely, from

$$\begin{aligned}\varphi(\boldsymbol{x})&=\frac{1}{(2\pi)^{\frac{3}{2}}}\int_{\mathbb{R}^3}\hat{\varphi}(\boldsymbol{k})e^{i\boldsymbol{kx}}d^3\boldsymbol{k}=\frac{1}{(2\pi)^{\frac{3}{2}}}\int_{\mathbb{R}^3}\left(a(\boldsymbol{k})e^{i\boldsymbol{kx}}+\bar{a}(\boldsymbol{k})e^{-i\boldsymbol{kx}}\right)\frac{d^3\boldsymbol{k}}{2\omega_{\boldsymbol{k}}},\\ \pi(\boldsymbol{x})&=\frac{1}{(2\pi)^{\frac{3}{2}}}\int_{\mathbb{R}^3}\hat{\pi}(\boldsymbol{k})e^{i\boldsymbol{kx}}d^3\boldsymbol{k}=\frac{-i}{(2\pi)^{\frac{3}{2}}}\int_{\mathbb{R}^3}\omega_{\boldsymbol{k}}\left(a(\boldsymbol{k})e^{i\boldsymbol{kx}}-\bar{a}(\boldsymbol{k})e^{-i\boldsymbol{kx}}\right)\frac{d^3\boldsymbol{k}}{2\omega_{\boldsymbol{k}}}\end{aligned}$$

we get

$$a(\boldsymbol{k})=\omega_{\boldsymbol{k}}\hat{\varphi}(\boldsymbol{k})+i\hat{\pi}(\boldsymbol{k})\in\mathscr{S}(\mathbb{R}^3),$$

so that (12.9) gives classical solution of the Cauchy problem.

It follows from Poisson brackets (12.7) that in the distributional sense

$$\{\hat{\pi}(\boldsymbol{k}),\hat{\pi}(\boldsymbol{l})\}=\{\hat{\varphi}(\boldsymbol{k}),\hat{\varphi}(\boldsymbol{l})\}=0$$

and

$$\begin{aligned}\{\hat{\pi}(\boldsymbol{k}),\hat{\varphi}(\boldsymbol{l})\}&=\frac{1}{(2\pi)^3}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\{\pi(\boldsymbol{x}),\varphi(\boldsymbol{y})\}e^{-i(\boldsymbol{kx}+\boldsymbol{ly})}d^3\boldsymbol{x}d^3\boldsymbol{y}\\ &=\frac{1}{(2\pi)^3}\int_{\mathbb{R}^3}e^{-i(\boldsymbol{k}+\boldsymbol{l})\boldsymbol{x}}d^3\boldsymbol{x}=\delta(\boldsymbol{k}+\boldsymbol{l}),\\ \{\hat{\pi}(\boldsymbol{k}),\overline{\hat{\varphi}(\boldsymbol{l})}\}&=\frac{1}{(2\pi)^3}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\{\pi(\boldsymbol{x}),\varphi(\boldsymbol{y})\}e^{-i(\boldsymbol{kx}-\boldsymbol{ly})}d^3\boldsymbol{x}d^3\boldsymbol{y}\\ &=\frac{1}{(2\pi)^3}\int_{\mathbb{R}^3}e^{-i(\boldsymbol{k}-\boldsymbol{l})\boldsymbol{x}}d^3\boldsymbol{x}=\delta(\boldsymbol{k}-\boldsymbol{l}).\end{aligned}$$

Thus we obtain

$$\{a(\boldsymbol{k}), a(\boldsymbol{l})\} = \{\bar{a}(\boldsymbol{k}), \bar{a}(\boldsymbol{l})\} = 0 \quad \text{and} \quad \{a(\boldsymbol{k}), \bar{a}(\boldsymbol{l})\} = 2i\omega_{\boldsymbol{k}}\delta(\boldsymbol{k}-\boldsymbol{l}). \tag{12.10}$$

Now it follows from Plancherel's theorem that

$$\begin{aligned} H &= \frac{1}{2}\int_{\mathbb{R}^3} \left(\pi^2(\boldsymbol{x}) + (\boldsymbol{\nabla}\varphi)^2(\boldsymbol{x}) + m^2\varphi^2(\boldsymbol{x})\right) d^3\boldsymbol{x} \\ &= \frac{1}{2}\int_{\mathbb{R}^3} \left(|\hat{\pi}(\boldsymbol{k})|^2 + \omega_{\boldsymbol{k}}^2|\hat{\varphi}(\boldsymbol{k})|^2\right) d^3\boldsymbol{k} \\ &= \int_{\mathbb{R}^3} \omega_{\boldsymbol{k}}\bar{a}(\boldsymbol{k})a(\boldsymbol{k})\frac{d^3\boldsymbol{k}}{2\omega_{\boldsymbol{k}}}. \end{aligned}$$

Similar computation gives for the total momentum

$$\begin{aligned} \boldsymbol{P} &= -\int_{\mathbb{R}^3} \pi(\boldsymbol{x})(\boldsymbol{\nabla}\varphi)(\boldsymbol{x})d^3\boldsymbol{x} \\ &= i\int_{\mathbb{R}^3} \hat{\pi}(\boldsymbol{k})\hat{\varphi}(-\boldsymbol{k})\boldsymbol{k}d^3\boldsymbol{k} \\ &= \int_{\mathbb{R}^3} \bar{a}(\boldsymbol{k})a(\boldsymbol{k})\boldsymbol{k}\frac{d^3\boldsymbol{k}}{2\omega_{\boldsymbol{k}}}. \end{aligned}$$

Thus we see that in terms of Fourier modes Hamilton's equations (12.4)–(12.5) decouple

$$\begin{aligned} \dot{a}(\boldsymbol{k}) &= \{H, a(\boldsymbol{k})\} = -i\omega_{\boldsymbol{k}}a(\boldsymbol{k}), \\ \dot{\bar{a}}(\boldsymbol{k}) &= \{H, \bar{a}(\boldsymbol{k})\} = \quad i\omega_{\boldsymbol{k}}\bar{a}(\boldsymbol{k}) \end{aligned}$$

and in accordance with (12.9)

$$a(t,\boldsymbol{k}) = e^{-i\omega_{\boldsymbol{k}}t}a(\boldsymbol{k}), \quad \bar{a}(t,\boldsymbol{k}) = e^{i\omega_{\boldsymbol{k}}t}\bar{a}(\boldsymbol{k}).$$

The real coordinates in the Fourier space

$$P(\boldsymbol{k}) = \frac{a(\boldsymbol{k}) + \bar{a}(\boldsymbol{k})}{2}, \quad Q(\boldsymbol{k}) = \frac{i(a(\boldsymbol{k}) - \bar{a}(\boldsymbol{k}))}{2\omega_{\boldsymbol{k}}}$$

are Darboux coordinates for the symplectic form Ω,

$$\Omega = \int_{\mathbb{R}^3} (dP(\boldsymbol{k}) \wedge dQ(\boldsymbol{k}))\, d^3\boldsymbol{k},$$

and the Hamiltonian of the Klein-Gordon model takes the form

$$H = \frac{1}{2}\int_{\mathbb{R}^3} \left(P^2(\boldsymbol{k}) + \omega_{\boldsymbol{k}}^2 Q^2(\boldsymbol{k})\right) d^3\boldsymbol{k}.$$

Thus in terms of Fourier modes the classical Klein-Gordon field is a collection of infinitely many non-interacting harmonic oscillators, parametrized by $\boldsymbol{k} \in \mathbb{R}^3$, with the frequencies $\omega_{\boldsymbol{k}} = \sqrt{\boldsymbol{k}^2 + m^2}$.

LECTURE 13

Hamiltonian formalism. Gauge theories.

13.1. Classical electrodynamics

Here we continue with $c = 1$ and use the Lagrangian function

$$\mathscr{L}(A) = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} = \frac{1}{2}(\boldsymbol{E}^2 - \boldsymbol{B}^2),$$

where

$$F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu}$$

(thus absorbing the factor $1/4\pi$ in (8.16) in Lecture 8). One can also rewrite Lagrangian function in the first order formalism (see Sect. 7.1 in Lecture 7),

$$\mathcal{L} = -\frac{1}{2}\left(\partial_\mu A_\nu - \partial_\nu A_\mu - \frac{1}{2}F_{\mu\nu}\right)F^{\mu\nu}, \tag{13.1}$$

where A_μ and $F_{\mu\nu}$ are considered to be independent. Indeed, corresponding Euler-Lagrange equations for $\mathcal{L}$ are Maxwell equations in free and empty space

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \quad \text{and} \quad \partial_\mu F^{\mu\nu} = 0.$$

Plugging formula for $F_{\mu\nu}$ back in (13.1), we obtain the Lagrangian function $\mathscr{L}(A)$. Using $A_\mu = (A_0, A_1, A_2, A_3) = (A_0, -\boldsymbol{A})$, formula (8.8) and equations $F_{ij} = \partial_i A_j - \partial_j A_i$, we can rewrite (13.1), up to a total divergence term $-\boldsymbol{\nabla}(A_0\boldsymbol{E})$, as

$$\mathcal{L} = -\boldsymbol{E}\cdot\partial_0\boldsymbol{A} - \frac{1}{2}(\boldsymbol{E}^2 + \boldsymbol{B}^2) + A_0\boldsymbol{\nabla}\cdot\boldsymbol{E}, \quad \text{where} \quad \boldsymbol{B} = \boldsymbol{\nabla}\times\boldsymbol{A}. \tag{13.2}$$

Thus for the electromagnetic field Lagrangian we obtain[1]

$$L = \int_{\mathbb{R}^3}(-\boldsymbol{E}(\boldsymbol{x})\cdot\partial_0\boldsymbol{A}(\boldsymbol{x}) - \tfrac{1}{2}(\boldsymbol{E}^2(\boldsymbol{x}) + \boldsymbol{B}^2(\boldsymbol{x})) + A_0(\boldsymbol{x})\boldsymbol{\nabla}\cdot\boldsymbol{E}(\boldsymbol{x}))d^3\boldsymbol{x}. \tag{13.3}$$

Comparison with formula (7.7) in Lecture 7 shows that Lagrangian of classical electrodynamics is singular. Namely, it follows from the first term in (13.3) that the phase space of the theory is the following infinite-dimensional real vector space[2]

$$\mathscr{M} = \mathscr{S}(\mathbb{R}^3, \mathbb{R}^3) \times \mathscr{S}(\mathbb{R}^3, \mathbb{R}^3)$$

[1] By the Stokes' theorem, contribution of the total divergence term is zero.

[2] Here $\mathscr{S}(\mathbb{R}^3, \mathbb{R}^3)$ stands for the $\mathbb{R}^3$-valued Schwartz functions on $\mathbb{R}^3$.

with the symplectic form Ω

$$\Omega = \int_{\mathbb{R}^3} (dE_i(\boldsymbol{x}) \wedge dA_i(\boldsymbol{x}))\, d^3\boldsymbol{x}, \tag{13.4}$$

so that the pairs $(E_i(\boldsymbol{x}), A_i(\boldsymbol{x}))$, are Darboux coordinates on $\mathscr{M}$ with the canonical Poisson brackets

$$\{E_i(\boldsymbol{x}), A_j(\boldsymbol{y})\} = \delta_{ij}\delta(\boldsymbol{x}-\boldsymbol{y}), \quad i,j = 1,2,3. \tag{13.5}$$

The second term in (13.3) is the Hamiltonian of the electromagnetic field,

$$H = \int_{\mathbb{R}^3} \mathscr{H}(\boldsymbol{x}) d^3\boldsymbol{x} = \frac{1}{2}\int_{\mathbb{R}^3} \left(\boldsymbol{E}^2 + \boldsymbol{B}^2\right) d^3\boldsymbol{x}, \quad \boldsymbol{B} = \boldsymbol{\nabla}\times\boldsymbol{A}. \tag{13.6}$$

Comparing the last term in (13.3) with the corresponding term in (7.7) we conclude that components $A_0(\boldsymbol{x})$ of the gauge field are the Lagrange multipliers, and the constraints $C(\boldsymbol{x})$ are given by the Gauss law,

$$C(\boldsymbol{x}) = \boldsymbol{\nabla}\cdot\boldsymbol{E}(\boldsymbol{x}) = 0, \quad \boldsymbol{x}\in\mathbb{R}^3.$$

It is instructive to analyze Hamilton's equations for the Hamiltonian system $(\mathscr{M}, \Omega, H)$. We have, using (13.5),

$$\dot{A}_i(\boldsymbol{x}) = \{H, A_i(\boldsymbol{x})\} = E_i(\boldsymbol{x}), \tag{13.7}$$

and since $\boldsymbol{A} = -(A_1, A_2, A_3)$, it gives

$$\boldsymbol{E} = -\frac{\partial \boldsymbol{A}}{\partial t},$$

which implies the Faraday law

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

Since the Gauss law for the magnetic field follows from the definition of $\boldsymbol{B} = \boldsymbol{\nabla}\times\boldsymbol{A}$, we get the first pair of Maxwell equations. Using $B_j = (\boldsymbol{\nabla}\times\boldsymbol{A})_j = -\varepsilon_{jkl}\partial_k A_l$ (note the negative sign!), we obtain

$$\begin{aligned}
\dot{E}_i(\boldsymbol{x}) = \{H, E_i(\boldsymbol{x})\} &= \int_{\mathbb{R}^3} B_j(\boldsymbol{y})\{(\boldsymbol{\nabla}\times\boldsymbol{A})_j(\boldsymbol{y}), E_i(\boldsymbol{x})\} d^3\boldsymbol{y} \\
&= -\varepsilon_{jkl}\int_{\mathbb{R}^3} B_j(\boldsymbol{y})\{\partial_k A_l(\boldsymbol{y}), E_i(\boldsymbol{x})\} d^3\boldsymbol{y} \\
&= \varepsilon_{jki}\int_{\mathbb{R}^3} B_j(\boldsymbol{y})\frac{\partial}{\partial y^k}\delta(\boldsymbol{x}-\boldsymbol{y}) d^3\boldsymbol{y} = \varepsilon_{ikj}\partial_k B_j(\boldsymbol{x}),
\end{aligned}$$

which gives the Ampére-Maxwell law

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

However, the remaining equation in the second pair of Maxwell equations — the Gauss law for the electric field $\boldsymbol{\nabla} \cdot \boldsymbol{E} = 0$ — is missing from Hamilton's equations and appears as constraints $C(\boldsymbol{x}) = 0$. This is a manifestation of the fact that Maxwell equations are described by a singular Lagrangian, and for their Hamiltonian formulation one needs to reduce the phase space $\mathscr{M}$.

It is easy to see that constraints $C(\boldsymbol{x})$ are the first class constraints (see Sect. 7.3 in Lecture 7). Indeed, it follows from (13.5) that

$$\{C(\boldsymbol{x}), C(\boldsymbol{y})\} = 0 \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^3$$

and also

$$\{H, C(\boldsymbol{x})\} = \{H, \boldsymbol{\nabla} \cdot \boldsymbol{E}(\boldsymbol{x})\} = \boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \times \boldsymbol{B})(\boldsymbol{x}) = 0, \quad \boldsymbol{x} \in \mathbb{R}^3.$$

According to Sect. 7.3 in Lecture 7, to determine the reduced phase space we need to introduce additional constraints $D(\boldsymbol{x}) = 0$ such that the integral operator with the kernel $\{C(\boldsymbol{x}), D(\boldsymbol{y})\}$ is non-degenerate in $L^2(\mathbb{R}^3)$. Convenient choice is

$$D(\boldsymbol{x}) = -\boldsymbol{\nabla} \cdot \boldsymbol{A}(\boldsymbol{x}),$$

which forces the Coulomb gauge! Indeed, it follows from (13.5) that,

$$\{C(\boldsymbol{x}), D(\boldsymbol{y})\} = \frac{\partial^2}{\partial x^i \partial y^i} \delta(\boldsymbol{x} - \boldsymbol{y}),$$

which is the integral kernel of the operator $-\Delta$, Laplace operator of the Euclidean metric on $\mathbb{R}^3$. Thus the reduced phase space $\mathscr{M}_0$ of classical electrodynamics is a linear subspace in $\mathscr{M}$ defined by

$$\mathscr{M}_0 = \left\{(\boldsymbol{E}(\boldsymbol{x}), \boldsymbol{A}(\boldsymbol{x})) \in \mathscr{M} : C(\boldsymbol{x}) = D(\boldsymbol{x}) = 0, \quad \boldsymbol{x} \in \mathbb{R}^3\right\}.$$

Since

$$\{D(\boldsymbol{x}), D(\boldsymbol{y})\} = 0,$$

Darboux coordinates for the symplectic form $\Omega_0 = \Omega|_{\mathscr{M}_0}$ can be found by the general procedure described in Sect. 7.3 in Lecture 7.

Using Corollary 7.1 in Lecture 7, the Poisson bracket $\{\ ,\ \}_0$ on $\mathscr{M}_0$, associated with the symplectic form Ω_0, can be written as a restriction of the Dirac bracket on $\mathscr{M}$, associated with the second class constraints $(C(\boldsymbol{x}), D(\boldsymbol{x}))$. Namely, it follows from (7.20) in Lecture 7 that

$$\begin{aligned} \{F, G\}_{\mathrm{DB}} = \{F, G\} + \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \Big(&\{F, C(\boldsymbol{x})\} G(\boldsymbol{y} - \boldsymbol{x}) \{D(\boldsymbol{y}), G\} - \\ &- \{F, D(\boldsymbol{x})\} G(\boldsymbol{x} - \boldsymbol{y}) \{C(\boldsymbol{y}), G\}\Big) d^3\boldsymbol{x} d^3\boldsymbol{y}, \end{aligned} \tag{13.8}$$

where $G(\boldsymbol{x} - \boldsymbol{y})$ is a distribution satisfying

$$\int_{\mathbb{R}^3} G(\boldsymbol{x} - \boldsymbol{z}) \{C(\boldsymbol{z}), D(\boldsymbol{y})\} d^3\boldsymbol{z} = \delta(\boldsymbol{x} - \boldsymbol{y}),$$

or

$$G(\boldsymbol{x}) = \frac{1}{(2\pi)^3}\int_{\mathbb{R}^3} \frac{e^{i\boldsymbol{k}\boldsymbol{x}}}{\boldsymbol{k}^2} d^3\boldsymbol{k}.$$

Using (13.5) we readily compute that

$$\{E_i(\boldsymbol{x}), E_j(\boldsymbol{y})\}_{\mathrm{DB}} = \{A_i(\boldsymbol{x}), A_j(\boldsymbol{y})\}_{\mathrm{DB}} = 0 \tag{13.9}$$

and

$$\{E_i(\boldsymbol{x}), A_j(\boldsymbol{y})\}_{\mathrm{DB}} = 4\pi\delta^{\perp}_{ij}(\boldsymbol{x}-\boldsymbol{y}), \quad \boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^3, \tag{13.10}$$

where the distribution $\delta^{\perp}_{ij}(\boldsymbol{x})$ is the *transverse δ-function*,

$$\delta^{\perp}_{ij}(\boldsymbol{x}) = \frac{1}{(2\pi)^3}\int_{\mathbb{R}^3}\left(\delta_{ij} - \frac{k_i k_j}{\boldsymbol{k}^2}\right) e^{i\boldsymbol{k}\boldsymbol{x}} d^3\boldsymbol{k}, \quad i,j = 1,2,3. \tag{13.11}$$

It satisfies

$$\partial_i \delta^{\perp}_{ij}(\boldsymbol{x}) = 0, \quad j = 1,2,3.$$

Thus Dirac bracket (13.8) yields a 'transverse' Poisson structure $\{\ ,\ \}$ on $\mathscr{M}$, determined by (13.9)–(13.10). It is degenerate and its center is generated by $C(\boldsymbol{x})$ and $D(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathbb{R}^3$. The Dirac bracket $\{\ ,\ \}_{\mathrm{DB}}$ restricts to $\mathscr{M}_0$ and yields a non-degenerate Poisson bracket $\{\ ,\ \}_0$ associated with the symplectic form Ω_0. Since

$$\int_{\mathbb{R}^3} \delta^{\perp}_{ij}(\boldsymbol{x}-\boldsymbol{y}) f_j(\boldsymbol{y}) d^3\boldsymbol{y} = f_i(\boldsymbol{x})$$

for any $\boldsymbol{f}(\boldsymbol{x}) \in \mathscr{S}(\mathbb{R}^3, \mathbb{R}^3)$ satisfying $\boldsymbol{\nabla}\cdot\boldsymbol{f}(\boldsymbol{x}) = 0$, it immediately follows from previous computations that Hamilton's equations on $\mathscr{M}_0$

$$\begin{aligned}\dot{\boldsymbol{E}}(\boldsymbol{x}) &= \{H, \boldsymbol{E}(\boldsymbol{x})\}_0,\\ \dot{\boldsymbol{A}}(\boldsymbol{x}) &= \{H, \boldsymbol{A}(\boldsymbol{x})\}_0,\end{aligned}$$

yield

$$\frac{\partial \boldsymbol{E}}{\partial t} = \boldsymbol{\nabla}\times\boldsymbol{B}, \quad \text{where} \quad \boldsymbol{B} = \boldsymbol{\nabla}\times\boldsymbol{A} \quad \text{and} \quad \frac{\partial \boldsymbol{A}}{\partial t} = -\boldsymbol{E}.$$

Together with the Gauss law, they give the full set of Maxwell equations in the Coulomb gauge.

In terms of the normal modes $\boldsymbol{P}(\boldsymbol{k})$ and $\boldsymbol{Q}(\boldsymbol{k})$ (see Sect. 11.4 in Lecture 11), satisfying

$$\boldsymbol{k}\cdot\boldsymbol{P}(\boldsymbol{k}) = \boldsymbol{k}\cdot\boldsymbol{Q}(\boldsymbol{k}) = 0,$$

the Poisson structure $\{\ ,\ \}_0$ is given by the transverse Poisson brackets

$$\{P_i(\boldsymbol{k}), Q_j(\boldsymbol{l})\}_0 = \left(\delta_{ij} - \frac{k_i l_j}{\boldsymbol{k}\cdot\boldsymbol{l}}\right)\delta(\boldsymbol{k}-\boldsymbol{l}).$$

This finishes Hamiltonian formulation of Maxwell's equations.

13.2. Yang-Mills equations

Let G be a semi-simple compact Lie group, $\mathfrak{g}$ be its Lie algebra with generators X_a satisfying

$$\operatorname{tr}(X_a X_b) = -2\delta_{ab}, \quad a, b = 1, \dots, n.$$

Let $A = A_\mu dx^\mu$ be a connection on a trivial bundle $\mathbb{R}^4 \times \operatorname{ad}\mathfrak{g}$ over $\mathbb{R}^4$, $A_\mu = A_\mu^a X_a$, and let

$$F = \frac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu, \quad \text{where} \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu],$$

be its curvature. Consider the Yang-Mills Lagrangian function (see Lecture 10),

$$\mathscr{L}(A) = \frac{1}{8} \operatorname{tr} F_{\mu\nu} F^{\mu\nu} = -\frac{1}{4} F_{\mu\nu}^a (F^a)^{\mu\nu},$$

where $F_{\mu\nu} = F_{\mu\nu}^a X_a$, and we put $g = 1$ in formula (10.10). As in case of classical electrodynamics, it can be written in the first order formalism

$$\mathcal{L} = \frac{1}{4} \operatorname{tr} \left(\partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu] - \frac{1}{2} F_{\mu\nu} \right) F^{\mu\nu}. \tag{13.12}$$

Put

$$\mathscr{E}_i = F_{0i} \quad \text{and} \quad \mathscr{B}_i = \varepsilon_{ijk} F^{jk}, \quad i =, 1, 2, 3.$$

Using equations $F_{ij} = \partial_i A_j - \partial_j A_i + [A_i, A_j]$ and the cyclic property of the trace, we can rewrite (13.12) as follows

$$\begin{aligned} \mathcal{L} &= -\frac{1}{2} \operatorname{tr} \left(\partial_0 A_k - \partial_k A_0 + [A_0, A_k] - \frac{1}{2} \mathscr{E}_k \right) \mathscr{E}_k + \frac{1}{8} \operatorname{tr} F_{ij} F^{ij} \\ &= -\frac{1}{2} \operatorname{tr} \left(\mathscr{E}_k \partial_0 A_k - \frac{1}{2} (\mathscr{E}_k^2 + \mathscr{B}_k^2) + A_0 \left(\partial_k \mathscr{E}_k + [A_k, \mathscr{E}_k] \right) \right) + \frac{1}{2} \partial_k \left(\operatorname{tr} A_0 \mathscr{E}_k \right). \end{aligned}$$

Thus up to a total divergence,

$$\mathcal{L} = -\frac{1}{2} \operatorname{tr} \left(\mathscr{E}_k \partial_0 A_k - \frac{1}{2} (\mathscr{E}_k^2 + \mathscr{B}_k^2) + A_0 C \right) \quad \text{where} \quad C = \partial_k \mathscr{E}_k + [A_k, \mathscr{E}_k],$$

or

$$\mathcal{L} = E_k^a \partial_0 A_k^a - \frac{1}{2} \left((E_k^a)^2 + (B_k^a)^2 \right) + A_0^a C^a, \tag{13.13}$$

where $\mathscr{E}_k = E_k^a X_a$, $\mathscr{B}_k = B_k^a X_a$ and $C = C^a X_a$. Thus for the Yang-Mills Lagrangian we obtain

$$L_{\mathrm{YM}} = \int_{\mathbb{R}^3} \left(E_k^a(\boldsymbol{x}) \partial_0 A_k^a(\boldsymbol{x}) - \tfrac{1}{2} \left((E_k^a)^2(\boldsymbol{x}) + (B_k^a)^2(\boldsymbol{x}) \right) + A_0^a(\boldsymbol{x}) C^a(\boldsymbol{x}) \right) d^3\boldsymbol{x}. \tag{13.14}$$

As formula (13.3) in case of classical electrodynamics, formula (13.14) shows that Yang-Mills Lagrangian is singular. Namely, it follows from the first term in (13.14) that the phase space of the theory is the following infinite-dimensional real vector space

$$\mathscr{M} = \mathscr{S}(\mathbb{R}^3, \mathbb{R}^{3n}) \times \mathscr{S}(\mathbb{R}^3, \mathbb{R}^{3n})$$

with the symplectic form Ω

$$\Omega = \int_{\mathbb{R}^3} (dE_k^a(\boldsymbol{x}) \wedge dA_k^a(\boldsymbol{x}))\, d^3\boldsymbol{x}, \tag{13.15}$$

so that the pairs $(E_k^a(\boldsymbol{x}), A_k^a(\boldsymbol{x}))$, are Darboux coordinates on $\mathscr{M}$ with the canonical Poisson brackets

$$\{E_k^a(\boldsymbol{x}), A_l^b(\boldsymbol{y})\} = \delta_{kl}\delta^{ab}\delta(\boldsymbol{x}-\boldsymbol{y}), \quad i,j=1,2,3 \quad \text{and} \quad a,b=1,\dots,n. \tag{13.16}$$

The second term in (13.14) is the Hamiltonian of the Yang-Mills field,

$$H = \int_{\mathbb{R}^3} \mathscr{H}(\boldsymbol{x}) d^3\boldsymbol{x} = \frac{1}{2}\int_{\mathbb{R}^3} \left((E_k^a)^2(\boldsymbol{x}) + (B_k^a)^2(\boldsymbol{x})\right) d^3\boldsymbol{x}. \tag{13.17}$$

Comparing the last term in (13.3) with the corresponding term in (7.7) we conclude that components $A_0^a(\boldsymbol{x})$ of the gauge field are the Lagrange multipliers, and the constraints $C^a(\boldsymbol{x})$ are given by the nonabelian Gauss law,

$$C^a(\boldsymbol{x}) = \partial_k E_k^a(\boldsymbol{x}) + t_{bc}^a A_k^b(\boldsymbol{x}) E_k^c(\boldsymbol{x}) = 0, \quad \boldsymbol{x} \in \mathbb{R}^3.$$

As in case of classical electrodynamics, these are the first class constraints, and we verify it by the following computation. Namely, it directly follows from (13.16) that

$$\{E_k^a(\boldsymbol{x}), C^b(\boldsymbol{y})\} = -t_c^{ab} E_k^c(\boldsymbol{x})\delta(\boldsymbol{x}-\boldsymbol{y})$$

and

$$\begin{aligned}\{A_k^a(\boldsymbol{x}), C^b(\boldsymbol{y})\} &= \{A_k^a(\boldsymbol{x}), \partial_l E_l^b(\boldsymbol{y}) + t_{cd}^b A_l^c(\boldsymbol{y}) E_l^d(\boldsymbol{y})\} \\ &= -\delta^{ab}\frac{\partial}{\partial y_k}\delta(\boldsymbol{x}-\boldsymbol{y}) - t_{ca}^b A_k^c(\boldsymbol{x})\delta(\boldsymbol{x}-\boldsymbol{y}) \\ &= -\left(\delta^{ab}\frac{\partial}{\partial y_k} + t_c^{ab} A^c(\boldsymbol{y})\right)\delta(\boldsymbol{x}-\boldsymbol{y}).\end{aligned}$$

Introducing

$$C(f) = -\frac{1}{2}\int_{\mathbb{R}^3} \operatorname{tr} C(\boldsymbol{x}) f(\boldsymbol{x}) d^3\boldsymbol{x} = \int_{\mathbb{R}^3} C^a(\boldsymbol{x}) f^a(\boldsymbol{x}) d^3\boldsymbol{x}$$

for a test function $f : \mathbb{R}^3 \to \operatorname{ad}\mathfrak{g}$, where $f(\boldsymbol{x}) = f^a(\boldsymbol{x}) X_a$ and $f^a(\boldsymbol{x}) \in \mathscr{S}(\mathbb{R}^3, \mathbb{R})$, we can succinctly rewrite these formulas as

$$\{\mathscr{E}_k(\boldsymbol{x}), C(f)\} \stackrel{\text{def}}{=} \{E_k^a(\boldsymbol{x}), C(f)\} X_a = [\mathscr{E}_k(\boldsymbol{x}), f(\boldsymbol{x})], \tag{13.18}$$

$$\{A_k(\boldsymbol{x}), C(f)\} \stackrel{\text{def}}{=} \{A_k^a(\boldsymbol{x}), C(f)\} X_a = (\nabla_k f)(\boldsymbol{x}). \tag{13.19}$$

From here we obtain

$$\{\partial_k \mathscr{E}_k(\boldsymbol{x}), C(g)\} = [\partial_k \mathscr{E}_k(\boldsymbol{x}), g(\boldsymbol{x})] + [\mathscr{E}_k(\boldsymbol{x}), \partial_k g(\boldsymbol{x})]$$

and

$$\begin{aligned}\{[A_k(\boldsymbol{x}), \mathscr{E}_k(\boldsymbol{x})], C(g)\} &= A_k(\boldsymbol{x})\{\mathscr{E}_k(\boldsymbol{x}), C(g)\} + \{A_k(\boldsymbol{x}), C(g)\}\mathscr{E}_k(\boldsymbol{x}) \\ &\quad - \mathscr{E}_k(\boldsymbol{x})\{A_k(\boldsymbol{x}), C(g)\} - \{\mathscr{E}_k(\boldsymbol{x}), C(g)\}A_k(\boldsymbol{x}) \\ &= [A_k(\boldsymbol{x}), [\mathscr{E}_k(\boldsymbol{x}), g(\boldsymbol{x})]] + [(\nabla_k g)(\boldsymbol{x}), \mathscr{E}_k(\boldsymbol{x})].\end{aligned}$$

Whence

$$\begin{aligned}\{C(\boldsymbol{x}), C(g)\} &= [\partial_k \mathscr{E}_k(\boldsymbol{x}), g(\boldsymbol{x})] + [A_k(\boldsymbol{x}), [\mathscr{E}_k(\boldsymbol{x}), g(\boldsymbol{x})]] + [[A_k(\boldsymbol{x}), g(\boldsymbol{x})], \mathscr{E}_k(\boldsymbol{x})] \\ &= [\partial_k \mathscr{E}_k(\boldsymbol{x}), g(\boldsymbol{x})] + [[A_k(\boldsymbol{x}), \mathscr{E}_k(\boldsymbol{x})], g(\boldsymbol{x})] \\ &= [C(\boldsymbol{x}), g(\boldsymbol{x})]\end{aligned}$$

and

$$\begin{aligned}\{C(f), C(g)\} &= -\frac{1}{2}\int_{\mathbb{R}^3} \operatorname{tr}\left([C(\boldsymbol{x}), g(\boldsymbol{x})]f(\boldsymbol{x})\right) d^3\boldsymbol{x} \\ &= \frac{1}{2}\int_{\mathbb{R}^3} \operatorname{tr}\left(C(\boldsymbol{x})[f(\boldsymbol{x}), g(\boldsymbol{x})]\right) d^3\boldsymbol{x}.\end{aligned}$$

Therefore, we finally obtain

$$\{C(f), C(g)\} = -C([f, g]). \tag{13.20}$$

Equivalently, (13.20) can be written as

$$\{C^a(\boldsymbol{x}), C^b(\boldsymbol{y})\} = -t^{ab}_c C^c(\boldsymbol{x})\delta(\boldsymbol{x} - \boldsymbol{y}).$$

Remark. One can also obtain formulas (13.18)–(13.20) using representation

$$C(f) = -\frac{1}{2}\int_{\mathbb{R}^3} \operatorname{tr}\left(\mathscr{E}_k(\boldsymbol{x})(\nabla_k f)(\boldsymbol{x})\right) d^3\boldsymbol{x}.$$

To compute $\{H, C^a(\boldsymbol{x})\}$ we observe that it follows from (13.18)

$$\{\mathscr{E}_k^2(\boldsymbol{x}), C(f)\} = [\mathscr{E}_k^2, f(\boldsymbol{x})].$$

Using (13.19), we also obtain

$$\begin{aligned}\{F_{ij}(\boldsymbol{x}), C(f)\} &= \{\partial_i A_j(\boldsymbol{x}) - \partial_j A_i(\boldsymbol{x}) + [A_i(\boldsymbol{x}), A_j(\boldsymbol{x})], C(f)\} \\ &= \partial_i(\nabla_j f)(\boldsymbol{x}) - \partial_j(\nabla_i f)(\boldsymbol{x}) + A_i(\boldsymbol{x})(\nabla_j f)(\boldsymbol{x}) + (\nabla_i f)(\boldsymbol{x})A_j(\boldsymbol{x}) \\ &\quad - (\nabla_j f)(\boldsymbol{x})A_i(\boldsymbol{x}) - A_j(\boldsymbol{x})(\nabla_i f)(\boldsymbol{x}) \\ &= [\partial_i A_j(\boldsymbol{x}) - \partial_j A_i(\boldsymbol{x}), f(\boldsymbol{x})] + [A_i(\boldsymbol{x}), [A_k(\boldsymbol{x}), f(\boldsymbol{x})] + [[A_i(\boldsymbol{x}), f(\boldsymbol{x})], A_k(\boldsymbol{x})] \\ &= [F_{ij}(\boldsymbol{x}), f(\boldsymbol{x})],\end{aligned}$$

so that

$$\{\mathscr{B}_k^2(\boldsymbol{x}), f(\boldsymbol{x})\} = [\mathscr{B}^2(\boldsymbol{x}), f(\boldsymbol{x})].$$

Whence

$$\{\mathscr{E}_k^2(\boldsymbol{x}) + \mathscr{B}_k^2, C(f)\} = [\mathscr{E}_k^2(\boldsymbol{x}) + \mathscr{B}_k^2, f(\boldsymbol{x})]$$

and for $\mathscr{H}(\boldsymbol{x}) = -\dfrac{1}{4}\operatorname{tr}\left(\mathscr{E}_k^2(\boldsymbol{x}) + \mathscr{B}_k^2(\boldsymbol{x})\right)$ we obtain

$$\{\mathscr{H}(\boldsymbol{x}), C(f)\} = 0.$$

Therefore

$$\{H, C(f)\} = 0$$

or equivalently,

$$\{H, C^a(\boldsymbol{x})\} = 0.$$

This finishes the proof that Yang-Mills theory is a Hamiltonian theory with first class constraints. As in the U(1)-case, for additional constraints one can use non-abelian Coulomb gauge

$$D(\boldsymbol{x}) = \partial_k A_k(\boldsymbol{x}) = 0.$$

Putting $D(\boldsymbol{x}) = D^a(\boldsymbol{x}) X_a$, we readily compute

$$\{C^a(\boldsymbol{x}), D^b(\boldsymbol{y})\} = \delta^{ab}\frac{\partial^2}{\partial x^k \partial y^k}\delta(\boldsymbol{x}-\boldsymbol{y}) + t_c^{ab} A_k^c(\boldsymbol{x})\frac{\partial}{\partial y^k}\delta(\boldsymbol{x}-\boldsymbol{y}). \tag{13.21}$$

Thus $M^{ab}(\boldsymbol{x},\boldsymbol{y}) = \{C^a(\boldsymbol{x}), D^b(\boldsymbol{y})\}$ is an integral kernel of the differential operator

$$M = -\Delta + \operatorname{ad} A_k(\boldsymbol{x})\partial_k, \tag{13.22}$$

acting on square summable $\operatorname{ad}\mathfrak{g}$-valued functions on $\mathbb{R}^3$. As in the U(1)-case, this operator is formally invertible, at least for small $A_k(\boldsymbol{x})$, which allows to define the reduced phase space of the theory.

Problem 13.1. The abelian group $C^\infty(\mathbb{R}^3, \mathbb{R})$ of gauge transformations acts on the phase space $\mathscr{M}$ by $f \cdot (\boldsymbol{E}, \boldsymbol{A}) = (\boldsymbol{E}, \boldsymbol{A} + \boldsymbol{\nabla} f)$. Prove that this action is Poisson and find the corresponding moment map (see Problems 6.4 and 7.3). Show that the reduced phase space for the regular value 0 is $\mathscr{M}_0$ and the corresponding symplectic structure is given by transverse Poisson brackets (13.10).

Problem 13.2. The nonabelian group $C^\infty(\mathbb{R}^3, G)$ of gauge transformations acts on the phase space $\mathscr{M}$ by $g \cdot (\mathscr{E}_k, A_k) = (g\mathscr{E}_k g^{-1}, gAg^{-1} - \partial_k g g^{-1})$. Prove that this action is Poisson and find the corresponding moment map.

Notes and references

Classic text

- L.D. Landau and E.M. Lifschitz, *The Classical Theory of Fields*, 4th edition, Butterworth-Heinemann, 1980

is the basic reference for classical electrodynamics, special relativity and theory of gravity. For more detailed exposition, including applications, see

- D.J. Griffiths, *Introduction to Electrodynamics*, Prentice-Hall, NJ, 1999
- J.D. Jackson, *Classical Electrodynamics*, 3rd edition, Wiley, 1998.

The book

- B.A. Dubrovin, A.T. Fomenko and S.P. Novikov, *Modern Geometry — Methods and Applications: Part II: The Geometry and Topology of Manifolds*, 2nd edition, Springer, 1991

is a good introduction to theory of connections on principal and vector bundles, see also

- T. Frankel, *The Geometry of Physics: an Introduction*, Cambridge University Press, 1999.

For concise exposition for vector bundles, see the books

- P. Griffiths and J. Harris, *Principles of algebraic geometry*, Wiley-Interscience, 1994.
- R.O. Wells, *Differential Analysis on Complex Manifolds*, Springer-Verlag New York, 2008.

For global existence and uniqueness theorems for the Yang-Mills equations on Minkowski spacetime see the papers

- D.M. Eardley and V. Moncrief, *The global existence of Yang-Mills-Higgs fields in 4-dimensional Minkowski space I. Local existence and smoothness properties* Comm. Math. Phys., **83** (1982), 171–191; *II. Completion of proof*, ibid. 193–212
- M.V. Goganov and L.V. Kapitanskiĭ, *Global solvability of the Cauchy problem for Yang-Mills-Higgs equations*, J. Sov. Math. **37** (1987), 802–822

and

- P.T. Chruściel and J. Shatah, *Global existence of solutions of the Yang-Mills equations on globally hyperbolic Lorentzian manifolds*, Asian J. Math. **1**:3 (1997), 530–548

for the general case.

For application of self-duality equations to the geometry of 4-manifolds, see the monograph

- S.K. Donaldson and P.B. Kronheimer, *The geometry of four-manifolds*, The Clarendon Press, Oxford University Press, New York, 1990.

The Hitchin equations were introduced in

- N. J. Hitchin, *The self-duality equations over a Riemann surface*, Proc. London Math. Soc. (3) **55** (1987) 59–126.

Our exposition of the Hamiltonian formalism for the Yang-Mills theory is based on

- L.D. Faddeev, *The Feynman integral for singular Lagrangians*, Theoret. and Math. Phys., **1**:1 (1969), 1–13
- L.D. Faddeev and A.A. Slavnov, *Gauge Fields. Introduction to Quantum Theory*, 2nd edition, Addison-Wesley, 1991.

The elegant proof in Lecture 13, that for the Yang-Mills theory $C^a(\boldsymbol{x})$ are the first class constraints, is based on Faddeev's lectures on Feynman path integral and gauge fields (unpublished, 1974).

Note that for large A_k Faddeev-Popov operator M, defined by formula (13.22), may have a zero eigenvalue, so that the Coulomb gauge condition intersects the orbits of gauge group more that once. This is the so-called *Gribov ambiguity*, which has been rigorously considered in

- I. M. Singer, *Some remarks on the Gribov ambiguity*, Commun. Math. Phys. **60** (1978), 7–12.

However, this problem does not affect the perturbation theory (Feynman rules) based on the Hamiltonian formulation of the Yang-Mills theory.

Part 3

Special relativity and theory of gravity

LECTURE 14

Special relativity

Maxwell's equations in vacuum are invariant with respect to the *Lorentz group* $\mathfrak{L} = \mathrm{O}(1,3)$ — the isometry group of Minkowski spacetime M_4 — the vector space $\mathbb{R}^4$ with Minkowski metric

$$ds^2 = \eta_{\mu\nu}dx^\mu dx^\nu = c^2dt^2 - dx^2 - dy^2 - dz^2.$$

Points in the spacetime are thought of as coordinates of *events* and the Minkowski distance between two events $P_1 = (ct_1, x_1, y_1, z_1)$ and $P_2 = (ct_2, x_2, y_2, z_2)$ is called the *interval*,

$$s_{12}^2 = c^2(t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2.$$

14.1. The relativity principle and the Lorentz group

The Minkowski structure of physical spacetime is a mathematical formulation of *Einstein's relativity principle*: "the speed of light is the same in all inertial frames of reference". If K and K' are two inertial reference frames, then the relativity principle is the statement that if $ds = 0$ in K then $ds' = 0$ in K'. From here it follows that

$$ds^2 = a(v)ds'^2,$$

where the constant $a(v)$ can depend only on the absolute value $v = |\boldsymbol{v}|$ of the relative velocity $\boldsymbol{v}$ of the inertial frames K and K'. Applying this to three reference frames K, K_1, K_2 we get

$$\frac{a(v_1)}{a(v_2)} = a(v_{12}),$$

where $v_{12} = |\boldsymbol{v}_2 - \boldsymbol{v}_1|$, which implies that $a(v) = 1$.

The Einstein relativity principle states that the physical laws are invariant with respect to the Lorentz group $\mathfrak{L}$, and replaces the Galilean relativity principle in Newtonian mechanics.

The orbits of the Lorentz group $\mathfrak{L}$ in M_4 have the form

$$\mathcal{O}_m = \{x \in M_4 : x^\mu x_\mu = c^2t^2 - x^2 - y^2 - z^2 = m^2\}$$

for all $m^2 \in \mathbb{R}$ and are two-sheeted hyperboloids when $m^2 > 0$, one-sheeted hyperboloids for $m^2 < 0$ and a cone $c^2t^2 = x^2 + y^2 + z^2$ for $m = 0$, the *light*

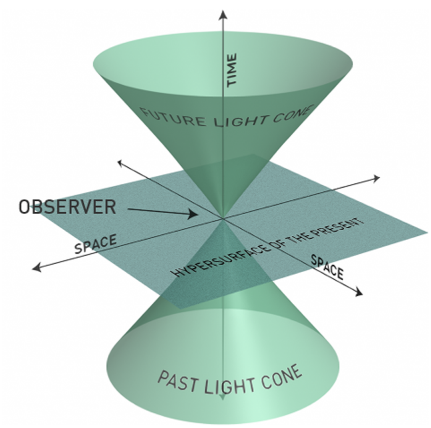

FIGURE 1. Light cone

cone (see Fig. 1). Correspondingly, two events $P_1, P_2 \in M_4$ are called *timelike* if $s_{12}^2 > 0$, *spacelike* if $s_{12}^2 < 0$ and *lightlike* if $s_{12} = 0$. It follows from the transitivity of the $\mathfrak{L}$-action on orbits that for two timelike events there is a Lorentz transformation such that they take place in the same point in space, $P_2 - P_1 = (t_2 - t_1, 0, 0, 0)$, while for the two spacelike events there is a Lorentz transformation such that they take place at the same time, $P_2 - P_1 = (0, \boldsymbol{x}_2 - \boldsymbol{x}_1)$. Clearly the spacelike events cannot be causally related. Correspondingly, the points inside the light cone with $t > 0$ represent the *absolute future* of the event at the origin O, while the points inside with $t < 0$ belong to the *absolute past*. The points outside the light cone are not causally related to the origin O and are *absolutely remote* relative to O. This means that the concepts "simultaneous", "earlier" and "later" are relative for these regions.

The Lorenz group $\mathfrak{L} = \mathrm{O}(1,3)$ consists of 4×4 matrices $\Lambda = \{\Lambda^\mu_\alpha\}$ satisfying

$$\Lambda^{\mathrm{t}}\eta\Lambda = \eta, \tag{14.1}$$

where $\eta = \mathrm{diag}\{1, -1, -1, -1\}$. Equivalently,

$$\Lambda^\mu_\alpha \Lambda^\nu_\beta \eta_{\mu\nu} = \eta_{\alpha\beta}.$$

The group $\mathfrak{L}$ acts linearly on M^4, $x \mapsto x' = \Lambda x$, where $x'^{\,\mu} = \Lambda^\mu_\nu x^\nu$. We have

$$(\Lambda^0_0)^2 - (\Lambda^1_0)^2 - (\Lambda^2_0)^2 - (\Lambda^i_0)^2 = 1,$$

so that $\Lambda^0_0 \geq 1$ or $\Lambda^0_0 \leq -1$. We also have $\det\Lambda = \pm 1$, so that the Lorentz group $\mathfrak{L}$ has four connected components.

The component of the identity $\mathfrak{L}^\uparrow_+$ preserves the future and past light cones and is called the *proper orthochronous Lorentz group* or *restricted Lorentz group*.

Other components are obtained from it by applying the *space inversion* $P = \mathrm{diag}\{1,-1,-1,-1\}$ or the *time reversal* $T = \mathrm{diag}\{-1,1,1,1\}$, or PT.

The restricted Lorentz group $\mathfrak{L}_+^{\uparrow}$ is six-dimensional connected Lie group generated rotations in $x^\mu x^\nu$-planes, $0 \leq \mu < \nu \leq 3$. Spacial rotations generated a subgroup SO(3), while rotations in x^0x^i-planes give *Lorentz boosts*. Explicitly, the rotation in x^0x^1-plane preserves $c^2t^2 - x^2$, where $x = x^1$. The corresponding transformation $x^\mu \mapsto x'^\mu$ can be written as

$$\begin{aligned} x &= x' \cosh\psi + ct' \sinh\psi, \\ ct &= x' \sinh\psi + ct' \cosh\psi. \end{aligned}$$

Putting

$$\cosh\psi = \frac{1}{\sqrt{1-\dfrac{v^2}{c^2}}}, \quad \sinh\psi = \frac{\dfrac{v}{c}}{\sqrt{1-\dfrac{v^2}{c^2}}},$$

where $|v| \leq c$, we get

$$x = \frac{x' + vt'}{\sqrt{1-\dfrac{v^2}{c^2}}}, \quad y = y', \quad z = z', \quad t = \frac{t' + \dfrac{v}{c^2}x'}{\sqrt{1-\dfrac{v^2}{c^2}}}. \tag{14.2}$$

This transformation relates coordinates (t,x,y,z) in the inertial reference frame K with the coordinates (t',x',y',z') in the inertial reference frame K' moving relative to K with velocity v along the x-axis. The formula for (t',x',y',z') in terms of (t,x,y,z) is given by replacing v by $-v$. wge When $|v| \ll c$ (or in the limit $c \to \infty$) Lorentz boost (14.2) becomes Galilean transformation (1.8) in Lecture 1,

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

Consider a particle in a reference frame K moving with velocity $\boldsymbol{v} = \dfrac{d\boldsymbol{r}}{dt}$. In the reference frame K' moving relative to K with velocity V in the x direction velocity of a particle is $\boldsymbol{v}' = \dfrac{d\boldsymbol{r}'}{dt'}$. Using

$$dx = \frac{dx' + V\,dt'}{\sqrt{1-\dfrac{V^2}{c^2}}}, \quad dy = dy', \quad dz = dz', \quad dt = \frac{dt' + \dfrac{V}{c^2}dx'}{\sqrt{1-\dfrac{V^2}{c^2}}}$$

we obtain

$$v_x = \frac{dx}{dt} = \frac{v'_x + V}{1 + \frac{v'_x V}{c^2}},$$
$$v_y = \frac{dy}{dt} = \frac{v'_y \sqrt{1 - \frac{V^2}{c^2}}}{1 + \frac{v'_x V}{c^2}},$$
$$v_z = \frac{dz}{dt} = \frac{v'_z \sqrt{1 - \frac{V^2}{c^2}}}{1 + \frac{v'_x V}{c^2}}.$$

When $|V| \ll c$ we get

$$v_x = v'_x + V, \quad v_y = v'_y \quad v_z = v'_z.$$

14.2. The Lorentz contraction and time delay

Consider a rod at rest in the K reference frame and suppose that it is parallel to the x-axis with the endpoints x_1 and x_2. The length of the rod, measured in the K reference frame, is just $\Delta x = x_2 - x_1$. To determine the length of the rode in the moving reference frame K', we need to find its endpoints x'_1 and x'_2 in K' at the same time t'. From (14.2) we obtain

$$x_1 = \frac{x'_1 + vt'}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad x_2 = \frac{x'_2 + vt'}{\sqrt{1 - \frac{v^2}{c^2}}}$$

and

$$\Delta x = \frac{\Delta x'}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

Denoting by $l_0 = \Delta x$ the *proper length* of the rod, the length in a reference frame where it is at rest, and by $l = \Delta x'$ its length in a moving reference frame K', we obtain the *Lorentz contraction*

$$l = l_0 \sqrt{1 - \frac{v^2}{c^2}},$$

so that $l < l_0$.

Next consider the clock which is at rest in the moving reference frame K'. Let (t'_1, x', y', z') and (t'_2, x', y', z') be two events occurring at the same point (x', y', z') in space in the reference frame K', so that the time between these events in K' is $\Delta t' = t'_2 - t'_1$. It follows from (14.2) that in the fixed reference frame K

$$t_1 = \frac{t_1' + \frac{v}{c^2}x'}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad t_2 = \frac{t_2' + \frac{v}{c^2}x'}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

Thus the time that elapses between these two events in the reference frame at rest K is

$$\Delta t = \frac{\Delta t'}{\sqrt{1 - \frac{v^2}{c^2}}},$$

so that $\Delta t' < \Delta t$. This is *time dilation* in special relativity: the time between events occurring at the same place in a moving reference frame is always smaller than the time between these events in a reference frame at rest. The time $\Delta t'$ is called a *proper time*.

Remark. Note that notion of being on the same point in space depends on the reference frame. Thus events (t_1', x', y', z') and (t_2', x', y', z') occur in the same point in space in the reference frame K', but in the reference frame K

$$x_1 = \frac{x' + vt_1'}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad x_2 = \frac{x' + vt_2'}{\sqrt{1 - \frac{v^2}{c^2}}},$$

and $x_1 \neq x_2$.

14.3. Lie algebra of the Lorentz group

The Lie algebra $\mathfrak{so}(1,3)$ of the Lorentz group is a Lie algebra of 4×4 matrices X satisfying

$$X^{\mathrm{t}}\eta + \eta X = 0,$$

which is obtained from (14.1) by setting $\Lambda = e^{sX} = I + sX + O(s^2)$. It is a semi-simple six-dimensional Lie algebra with the generators $M^{\lambda\mu}$, $0 \leq \lambda < \mu \leq 3$, and the Lie brackets

$$[M^{\lambda\mu}, M^{\rho\sigma}] = -\eta^{\lambda\rho}M^{\mu\sigma} + \eta^{\lambda\sigma}M^{\mu\rho} - \eta^{\mu\sigma}M^{\lambda\rho} + \eta^{\mu\rho}M^{\lambda\sigma}.$$

Here it is understood that $M^{\lambda\lambda} = 0$ (no summation over repeated indices!) and $M^{\lambda\mu} = -M^{\mu\lambda}$ for $\lambda > \mu$. The generators $M^{\lambda\mu}$ can be realized as the following 4×4 matrices

$$\left(M^{\lambda\mu}\right)^{\alpha}_{\beta} = \eta^{\alpha\lambda}\delta^{\mu}_{\beta} - \eta^{\alpha\mu}\delta^{\lambda}_{\beta}.$$

Introducing

$$J_i = \frac{1}{2}\varepsilon_{ikl}M^{kl} \quad \text{and} \quad K_i = M_{0i}, \quad i = 1, 2, 3,$$

we obtain the following Lie brackets

$$\begin{aligned}
[J_i, J_j] &= \varepsilon_{ijl} J_l, \\
[K_i, K_j] &= -\varepsilon_{ijl} J_l, \\
[J_i, K_j] &= \varepsilon_{ijl} K_l, \quad i, j = 1, 2, 3.
\end{aligned}$$

The generators J_1, J_2, J_3 correspond to the rotations in $\mathbb{R}^3$ and K_1, K_2, K_3 — to the Lorentz boosts. Explicitly[1], $J_1 = \begin{pmatrix} 0&0&0&0\\0&0&0&0\\0&0&0&-1\\0&0&1&0 \end{pmatrix}$, $J_2 = \begin{pmatrix} 0&0&0&0\\0&0&0&1\\0&0&0&0\\0&-1&0&0 \end{pmatrix}$, $J_3 = \begin{pmatrix} 0&0&0&0\\0&0&-1&0\\0&1&0&0\\0&0&0&0 \end{pmatrix}$ and $K_1 = \begin{pmatrix} 0&1&0&0\\1&0&0&0\\0&0&0&0\\0&0&0&0 \end{pmatrix}$, $K_2 = \begin{pmatrix} 0&0&1&0\\0&0&0&0\\1&0&0&0\\0&0&0&0 \end{pmatrix}$, $K_3 = \begin{pmatrix} 0&0&0&1\\0&0&0&0\\0&0&0&0\\1&0&0&0 \end{pmatrix}$.

Remark. Complexified Lie algebra $\mathfrak{so}(1,3)$ is isomorphic to $\mathfrak{so}(4,\mathbb{C})$ with the generators

$$J_i^{(\pm)} = \frac{1}{2}(J_i \pm \sqrt{-1}\, K_i)$$

satisfying

$$[J_i^{(+)}, J_j^{(+)}] = \varepsilon_{ijl} J_l^{(+)}, \quad [J_i^{(-)}, J_j^{(-)}] = \varepsilon_{ijl} J_l^{(-)}, \quad [J_i^{(+)}, J_j^{(-)}] = 0,$$

which establishes the Lie algebra isomorphism $\mathfrak{so}(4) \cong \mathfrak{so}(3) \oplus \mathfrak{so}(3)$. Note that over $\mathbb{R}$ there is a Lie group isomorphism

$$\mathrm{SO}(3) \times \mathrm{SO}(3) \cong \mathrm{SO}(4)/\{I, -I\}.$$

Remark. Replacing $\eta = \mathrm{diag}(1, -1, -1, -1)$ by $\eta_c = \mathrm{diag}(c, -1, -1, -1)$, we get generators J_i and K_i^c, and since $\eta_c^{-1} = \mathrm{diag}(1/c, -1-1, -1)$ we obtain

$$[K_i^c, K_j^c] = -\frac{1}{c^2} \varepsilon_{ijl} J_l.$$

Thus in the non-relativistic limit $c \to \infty$ for the generators J_i and $\tilde{K}_i = \lim_{c\to\infty} K_i^c$ we obtain the relations

$$\begin{aligned}
[J_i, J_j] &= \varepsilon_{ijl} J_l, \\
[J_i, \tilde{K}_j] &= \varepsilon_{ijl} K_l, \\
[\tilde{K}_i, \tilde{K}_j] &= 0,
\end{aligned}$$

which characterize the Lie algebra $\mathfrak{se}(3)$ of the Euclidean group $E(3)$ — the homogenous Galilean group G_0 — discussed in Sect. (1.3) in Lecture 1! Thus we see that Euclidean Lie algebra $\mathfrak{se}(3)$ is a *contraction* of the Lorentz Lie algebra $\mathfrak{so}(1,3)$.

[1]Compare with formulas for X_1, X_2 and X_3 in Example 2.2 in Lecture 2.

14.4. Lorentz group as deformation of the Galilean group

Specifically, the Lorentz Lie algebra $\mathfrak{so}(1,3)$ can be considered as a *deformation* of the Galilean Le algebra $\mathfrak{se}(3)$, with the deformation parameter being the inverse square of the speed of light c.

Namely recall that a formal deformation of a Lie algebra $\mathfrak{g}$ with a Lie bracket $[\ ,\ ]$ is a Lie algebra $\tilde{\mathfrak{g}}$ over $R[[t]]$, a ring of formal power series in variable[2] t, with the Lie bracket

$$[x,y]_t = [x,y] + tm_1(x,y) + t^2 m_2(x,y) + \cdots$$

The Jacobi identity for the bracket $[\ ,\ ]_t$ implies that the linear map $m_1 : \Lambda^2\mathfrak{g} \to \mathfrak{g}$ satisfies

$$\begin{aligned} &[m_1(x,y),z] + m_1([z,x],y) + [m_1(y,z),x] \\ &+[m_1(z,x),y] + m_1([z,x],y) + m_1([y,z],x) = 0 \end{aligned} \tag{14.3}$$

for all $x,y,z \in \mathfrak{g}$. This is the equation of 2-cocycle in the Shevalley-Eilenberg complex $\mathrm{Hom}(\Lambda^\bullet\mathfrak{g},\mathfrak{g})$, where $\mathfrak{g}$ is considered as a left $\mathfrak{g}$-module with respect to the adjoint action. Specifically, for any $\mathfrak{g}$-module M the coboundary map $\delta_k : \mathrm{Hom}(\Lambda^k\mathfrak{g}, M) \to \mathrm{Hom}(\Lambda^{k+1}\mathfrak{g}, M)$ is defined by

$$\begin{aligned} (\delta_k f)(x_1,\dots,x_{k+1}) &= \sum_{i=1}^{k+1}(-1)^{i+1}x_i\cdot f(x_1,\dots,\hat{x}_i,\dots,x_{k+1}) + \\ &+ \sum_{1\le i<j\le k+1}(-1)^{i+j}f([x_i,x_j],x_1,\dots,\hat{x}_i,\dots,\hat{x}_j,\dots x_{k+1}). \end{aligned}$$

Remark. In case when $\mathfrak{g} = \mathrm{Vect}(X)$, where X is a smooth manifold, and $M = C^\infty(X)$, the Chevalley-Eilenberg complex $\mathrm{Hom}(\Lambda^\bullet\mathfrak{g}, M)$ becomes the de Rham complex $\Omega^\bullet_{\mathrm{dR}}(X,\mathbb{R})$.

Equation (14.3) for m_1 can be written as $\delta_2 m_1 = 0$. Coboundaries

$$(\delta_1 f)(x,y) = [x,f(y)] - [y,f(x)] - f([x,y])$$

give infinitesimally trivial deformations: the linear map $F_t(x) = x + tf(x)$ establishes the infinitesimal isomorphism

$$F_t([x,y]_t) = [F_t(x),F_t(y)] + O(t^2).$$

Thus nontrivial infinitesimal deformations are in one-to-one correspondence with the second cohomology group $H^2(\mathfrak{g},\mathfrak{g})$.

Definition. The Lie algebra is called *stable* if $H^2(\mathfrak{g},\mathfrak{g}) = 0$.

The semi-simple Lie algebras are stable. However, for the Lie algebra $\mathfrak{g} = \mathfrak{se}(3)$ we have $H^2(\mathfrak{g},\mathfrak{g}) = \mathbb{R}$ and for the 2-cocycle m_1 with the only non-zero values $m_1(\tilde{K}_i,\tilde{K}_j) = -\varepsilon_{ijk}J_k$ we obtain that the bracket

$$[x,y]_t = [x,y] + tm_1(x,y)$$

[2] Should not be confused with the time variable!

is a Lie bracket (contribution of the terms proportional to t^2 to the Jacobi identity is zero). Putting $t = c^{-2}$ we obtain the Lorentz Lie algebra!

The Lorentz algebra is semi-simple and therefore is stable. Whence the passage from the Newtonian spacetime to the Minkowski spacetime is a deformation from the unstable structure to the stable one and the special relativity is natural deformation of the Newtonian mechanics.

LECTURE 15

Relativistic particle

A motion of a particle in M_4 is described by a *world line*. By definition, it is a map $\gamma : [t_1, t_2] \to M^4$, $\gamma(t) = x^\mu(t)$, such that at each $t \in [t_1, t_2]$ the tangent vector $\gamma'(t)$ is timelike. Explicitly, $\gamma(t) = (ct, \boldsymbol{r}(t))$ where $\boldsymbol{v}(t) = \dot{\boldsymbol{r}}(t)$ satisfies $|v(t)| < c$, where $v = |\boldsymbol{v}|$. In terms of the natural parameter s on the world line,

$$ds = c\sqrt{1 - \frac{v^2}{c^2}}\, dt,$$

the unit tangent vector is given by

$$u^\mu = \frac{dx^\mu}{ds} = \left(\frac{1}{\sqrt{1 - \dfrac{v^2}{c^2}}}, \frac{\boldsymbol{v}}{c\sqrt{1 - \dfrac{v^2}{c^2}}} \right), \quad u_\mu u^\mu = 1,$$

and the acceleration is

$$a^\mu = \frac{du^\mu}{ds}, \quad a^\mu u_\mu = 0.$$

Remark. The natural parameter is c times the proper time along the world line,

$$s(t) = c\int_{t_1}^{t} \sqrt{1 - \frac{v^2(\tau)}{c^2}}\, d\tau.$$

15.1. The principle of the least action

Let $a, b \in M_4$ be two events with a timelike interval $s_{ab}^2 > 0$. It is natural to define the action of the a relativistic particle along the world line $\gamma : [t_0, t_1] \to M_4$, $\gamma(t_0) = a$ and $\gamma(t_1) = b$, by the following expression

$$S(\gamma) = -\alpha \int_a^b ds.$$

Here integration goes over the world line γ and α is a constant.

It follows from the pseudo-Euclidean structure of the Minkowski spacetime that the integral $\int_a^b ds$ takes a maximal value when it is taken along a straight world line connecting a and b. Indeed, applying a Lorentz transformation, we can assume that $a = (ct_0', x', y', z')$ and $b = (ct_1', x', y', z')$, so that along a world line γ

$$\int_a^b ds \leq c(t_1' - t_0')$$

and the equality occurs for γ being a straight line connecting a and b with zero velocity.

Thus to have a minimum of the action we have $\alpha > 0$, so that or $\gamma(t) = (ct, \boldsymbol{r}(t))$,

$$S(\gamma) = \int_{t_0}^{t_1} L(\gamma'(t))dt, \quad \text{where} \quad L = -\alpha\sqrt{1 - \frac{v^2}{c^2}} \quad \text{and} \quad v = |\dot{\boldsymbol{r}}|.$$

The quantity α characterizes the particle. In classical mechanics a particle is characterized by its mass m (see Lecture 1). Whence in the non-relativistic limit $c \to \infty$ we should recover the Lagrangian of a free particle $mv^2/2$, and this comparison yields a relation between α and m. Namely, we have as $c \to \infty$

$$L = -\alpha c\sqrt{1 - \frac{v^2}{c^2}} = -\alpha c + \frac{\alpha v^2}{2c} + O(c^{-3}).$$

Omitting the constant term $-\alpha c$, which does not affect the equations of motion, we obtain $\alpha = mc$. Thus the action of a free relativistic particle of mass m is

$$S(\gamma) = -mc\int_a^b ds = \int_{t_0}^{t_1} L(\gamma'(t))dt \tag{15.1}$$

with the Lagrangian function

$$L = -mc^2\sqrt{1 - \frac{v^2}{c^2}}. \tag{15.2}$$

PROPOSITION 15.1. *The Euler-Lagrange equations of a free relativistic particle are*

$$\frac{du^\mu}{ds} = 0$$

and describe a motion with constant velocity.

PROOF. Since $ds = \sqrt{dx_\mu dx^\mu}$, we have along the world-line γ,

$$\begin{aligned}\delta(ds) &= \frac{1}{2}\left(\frac{dx_\mu}{ds}\delta dx^\mu + \delta dx_\mu \frac{dx^\mu}{ds}\right)\\ &= u^\mu d\delta x_\mu\\ &= d(u^\mu \delta x_\mu) - \frac{du^\mu}{ds}\delta x_\mu ds,\end{aligned}$$

and using $\delta x_\mu(a) = \delta x_\mu(b) = 0$, we obtain

$$\delta S = -mc\int_a^b \delta(ds) = mc\int_a^b \frac{du^\mu}{ds}\delta x_\mu ds. \qquad \square$$

15.2. Energy-momentum vector

Canonically conjugated momentum $\boldsymbol{p}$ to the position $\boldsymbol{r}$ of the particle is given by

$$\boldsymbol{p} = \frac{\partial L}{\partial \boldsymbol{v}} = \frac{m\boldsymbol{v}}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

The corresponding energy is

$$\mathscr{E} = \boldsymbol{p} \cdot \boldsymbol{v} - L = \frac{mv^2}{\sqrt{1 - \frac{v^2}{c^2}}} + mc^2\sqrt{1 - \frac{v^2}{c^2}} = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

At $v = 0$ we obtain the *rest energy* $\mathscr{E}_0$ of the particle,

$$\mathscr{E}_0 = mc^2.$$

At small velocities we obtain

$$\mathscr{E} = \mathscr{E}_0 + \frac{mv^2}{2} + O(v^4)$$

which, except for the rest energy, is the classical expression for the kinetic energy of a free particle. We have

$$\frac{\mathscr{E}^2}{c^2} = p^2 + m^2c^2, \quad p^2 = \boldsymbol{p} \cdot \boldsymbol{p},$$

so that the corresponding Hamiltonian function is

$$\mathscr{H} = c\sqrt{p^2 + m^2c^2},$$

and Hamilton's equations

$$\dot{\boldsymbol{p}} = -\frac{\partial \mathscr{H}}{\partial \boldsymbol{r}}, \quad \dot{\boldsymbol{r}} = \frac{\partial \mathscr{H}}{\partial \boldsymbol{p}}$$

give Euler-Lagrange equations of a free relativistic particle (see Proposition 15.1). Introducing the energy-momentum four vector $p^\mu = (\mathscr{E}/c, \boldsymbol{p})$, so that $p_\mu = (\mathscr{E}/c, -\boldsymbol{p})$, we have

$$p_\mu p^\mu = m^2c^2.$$

Note that $\boldsymbol{p} = -(p_1, p_2, p_3)$ and

$$p_\mu = -\frac{\partial L}{\partial \dot{x}^\mu}.$$

15.3. Charged particle in the electromagnetic field

Here we consider the interaction of a free relativistic particle of mass m and charge e with the external electromagnetic field with a potential $A = A_\mu dx^\mu$, where $A_\mu = (c\varphi, -\boldsymbol{A})$. To every world line $\gamma : [t_0, t_1] \to M_4$ one associates a holonomy of the connection $d + A$ along γ, the integral

$$\int_a^b A_\mu dx^\mu$$

of A along γ. It is natural to define the action of a free particle in the electromagnetic field as a linear combination of the action of a free particle and the holonomy, and we put

$$\begin{aligned} S(\gamma) &= -mc\int_a^b ds - \frac{e}{c}\int_a^b A_\mu dx^\mu \\ &= \int_{t_0}^{t_1}\left(-mc^2\sqrt{1-\frac{v^2}{c^2}} + \frac{e}{c}\boldsymbol{A}\cdot\boldsymbol{v} - e\varphi\right)dt. \end{aligned} \tag{15.3}$$

Proposition 15.2. *The Euler-Lagrange equations for the action functional* (15.3) *have the form*

$$\frac{d\boldsymbol{p}}{dt} = \boldsymbol{F},$$

where F is the Lorentz force,

$$\boldsymbol{F} = e\left(\boldsymbol{E} + \frac{\boldsymbol{v}}{c}\times\boldsymbol{B}\right).$$

Proof. We have

$$\begin{aligned} \delta\int_a^b A_\mu dx^\mu &= \int_a^b\left(\frac{\partial A_\mu}{\partial x^\nu}\frac{dx^\mu}{ds}\delta x^\nu + A_\mu\frac{d\delta x^\mu}{ds}\right)ds \\ &= \int_a^b\left(\frac{\partial A_\mu}{\partial x^\nu}\frac{dx^\mu}{ds}\delta x^\nu - \frac{\partial A_\mu}{\partial x^\nu}\frac{dx^\nu}{ds}\delta x^\mu\right)ds \\ &= -\int_a^b F_{\mu\nu}\frac{dx^\mu}{ds}\delta x^\nu ds. \end{aligned}$$

Now using Proposition 15.1 we obtain

$$\delta S = \int_a^b\left(mc\frac{du_\nu}{ds} + \frac{e}{c}F_{\mu\nu}\frac{dx^\mu}{ds}\right)\delta x^\nu ds,$$

and the Euler-Lagrange equations take the following invariant form

$$mc\frac{du_\nu}{ds} + \frac{e}{c}F_{\mu\nu}\frac{dx^\mu}{ds} = 0. \tag{15.4}$$

Using formula (8.8) in Lecture 8, relation $mcu_\nu = p_\nu$ and equation (15.4) for $\nu = 1, 2, 3$, we readily obtain

$$\frac{d\boldsymbol{p}}{dt} = e\boldsymbol{E} + \frac{e}{c}\boldsymbol{v}\times\boldsymbol{B}. \tag{15.5}$$

□

REMARK. Since $mcu_0 = \sqrt{m^2c^2 + \boldsymbol{p}^2}$, equation (15.4) for $\nu = 0$ follows from (15.5).

REMARK. In the non-relativistic limit $|\boldsymbol{v}| \ll c$ equation (15.5) turns into

$$m\frac{d\boldsymbol{v}}{dt} = e\left(\boldsymbol{E} + \frac{\boldsymbol{v}}{c} \times \boldsymbol{B}\right)$$

— Newton's equation with the Lorentz force.

The Lagrangian of a charged particle in electromagnetic field is

$$L = -mc^2\sqrt{1 - \frac{v^2}{c^2}} + \frac{e}{c}\boldsymbol{A} \cdot \boldsymbol{v} - e\varphi.$$

The canonically conjugated to $\boldsymbol{r}$ momentum of the charged particle, the generalized momentum, is defined by

$$\boldsymbol{P} = \frac{\partial L}{\partial \boldsymbol{v}} = \frac{m\boldsymbol{v}}{\sqrt{1 - \frac{v^2}{c^2}}} + \frac{e}{c}\boldsymbol{A} = \boldsymbol{p} + \frac{e}{c}\boldsymbol{A},$$

and the corresponding energy is

$$\begin{aligned}\mathscr{E} = \boldsymbol{v}\frac{\partial L}{\partial \boldsymbol{v}} - L &= \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} + e\varphi,\\ &= c\sqrt{m^2c^2 + \boldsymbol{p}^2} + e\varphi.\end{aligned}$$

The Hamiltonian function is obtained from the energy $\mathscr{E}$ by replacing $\boldsymbol{p} = \boldsymbol{P} - \frac{e}{c}\boldsymbol{A}$ and is given by

$$\mathscr{H} = c\sqrt{m^2c^2 + \left(\boldsymbol{P} - \frac{e}{c}\boldsymbol{A}\right)^2} + e\varphi.$$

Hamilton's equations of motion

$$\dot{\boldsymbol{P}} = -\frac{\partial \mathscr{H}}{\partial \boldsymbol{r}}, \quad \dot{\boldsymbol{r}} = \frac{\partial \mathscr{H}}{\partial \boldsymbol{P}},$$

together with the definitions

$$\boldsymbol{E} = -\boldsymbol{\nabla}\varphi - \frac{\partial \boldsymbol{A}}{\partial t}, \quad \boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{A},$$

give Euler-Lagrange equations for a charged particle in the electromagnetic field.

LECTURE 16

Hamiltonian formulation

16.1. Poincaré group and Noether integrals

The Poincaré group is a ten-dimensional Lie group, the group of isometries

$$x^\mu \mapsto x'^\mu = \Lambda^\mu_\nu x^\nu + a^\mu \tag{16.1}$$

of Minkowski spacetime M_4. The group multiplication in $\mathfrak{P}$ is given by

$$(\Lambda_1, a_1)(\Lambda_1, a_2) = (\Lambda_1\Lambda_2, a_1 + \Lambda_1 a_2), \quad \Lambda_{1,2} \in \mathfrak{L},\ a_{1,2} \in \mathbb{R}^4.$$

There is an embedding $\mathfrak{P} \hookrightarrow \mathrm{GL}(5, \mathbb{R})$ given by

$$(\Lambda, a) \mapsto \begin{pmatrix} \Lambda & a \\ 0 & 1 \end{pmatrix}.$$

The Lie algebra $\mathfrak{p}$ of the Poincaré group $\mathfrak{P}$ is a ten-dimensional Lie algebra, a semi-direct sum of the abelian Lie algebra $\mathbb{R}^4$ and the Lorentz Lie algebra $\mathfrak{so}(1,3)$. Denoting by P^μ the generators of $\mathfrak{p}$ corresponding to space-time translations we obtain the following set of relations:

$$\begin{aligned}
[P^\mu, P^\nu] &= 0, \\
[M^{\lambda\mu}, P^\sigma] &= \eta^{\lambda\sigma} P^\mu - \eta^{\mu\sigma} P^\lambda, \\
[M^{\lambda\mu}, M^{\rho\sigma}] &= -\eta^{\lambda\rho} M^{\mu\sigma} + \eta^{\lambda\sigma} M^{\mu\rho} - \eta^{\mu\sigma} M^{\lambda\rho} + \eta^{\mu\rho} M^{\lambda\sigma}.
\end{aligned}$$

The Lagrangian function of a free relativistic particle

$$L = -mc\sqrt{\dot{x}^\mu \dot{x}_\mu}, \quad \dot{x}^\mu = \frac{dx^\mu}{dt}, \quad t = \frac{x^0}{c},$$

is invariant under the action (16.1) of the Poincaré group on M_4,

$$L dt = L' dt', \quad \text{where} \quad L' = -mc\sqrt{\dot{x}'^\mu \dot{x}'_\mu}, \quad \dot{x}'^\mu = \frac{dx'^\mu}{dt}, \quad t' = \frac{x'^0}{c}.$$

According to Noether theorem in Sect. 2.2 in Lecture 2, there are ten integrals of motion corresponding to the generators P^μ and $M^{\lambda\mu}$. The integrals of motion for the abelian Lie algebra $\mathbb{R}^4$ are

$$p_\mu = -\frac{\partial L}{\partial \dot{x}^\mu},$$

that is,

$$p_0 = \frac{\mathscr{H}}{c} = \sqrt{p^2 + m^2c^2}, \quad \boldsymbol{p} = \frac{m\boldsymbol{v}}{\sqrt{1 - \dfrac{v^2}{c^2}}}$$

(recall that $p_\mu = (p_0, -\boldsymbol{p})$, see Sect. 15.2 in Lecture 15). The vector fields on $\mathbb{R}^4$ which corresponds to the one-parameter subgroups $e^{sM^{\mu\nu}}$ of the Lorentz group generated by $M^{\mu\nu}$ are

$$X^{\mu\nu} = (M^{\mu\nu} \cdot x)^\sigma \frac{\partial}{\partial x^\sigma} = (\eta^{\sigma\nu}x^\mu - \eta^{\sigma\mu}x^\nu)\frac{\partial}{\partial x^\sigma}.$$

The corresponding Noether integrals are given by

$$J^{\mu\nu} = (\eta^{\sigma\nu}x^\mu - \eta^{\sigma\mu}x^\nu)\frac{\partial L}{\partial \dot{x}^\sigma} = x^\mu p^\nu - x^\nu p^\mu.$$

Thus we obtain components of the total angular momentum

$$J_x = J^{23} = x^2p^3 - x^3p^2, \quad J_y = J^{31} = x^3p^1 - x^1p^3, \quad J_z = J^{12} = x^1p^2 - p^1x^2$$

and integrals of motion corresponding to Lorentz boosts

$$K_x = J^{01} = x^0p^1 - x^1p^0, \quad K_y = J^{02} = x^0p^2 - x^2p^0, \quad K_z = J^{01} = x^0p^3 - x^1p^3.$$

Of course it is easy to verify directly that these functions are integrals of motion. Thus we have

$$\dot{J}^{0i} = cp^i - \dot{x}^ip^0 = 0$$

due to the relation

$$\boldsymbol{v} = \frac{c\boldsymbol{p}}{\sqrt{p^2 + m^2c^2}},$$

which follows from

$$\boldsymbol{p} = \frac{m\boldsymbol{v}}{\sqrt{1 - \dfrac{v^2}{c^2}}}.$$

16.2. Hamiltonian action of the Poincaré group

The Legendre transform

$$\boldsymbol{p} = \frac{m\boldsymbol{v}}{\sqrt{1 - \dfrac{v^2}{c^2}}}. \tag{16.2}$$

maps $\mathbb{B}(0, c)$, the ball of radius c in $\mathbb{R}^3$, onto $\mathbb{R}^3$ and the phase space of a free relativistic particle of mass m is $\mathbb{R}^6$. The inverse transform is

$$\boldsymbol{v} = \frac{c\boldsymbol{p}}{\sqrt{p^2 + m^2c^2}} = \frac{c\boldsymbol{p}}{p^0}. \tag{16.3}$$

The symplectic form is given by

$$\omega = d\boldsymbol{p} \wedge d\boldsymbol{r} = dp^1 \wedge dx^1 + dp^2 \wedge dx^2 + dp^3 \wedge dx^3$$

with Darboux coordinates[1] $(\boldsymbol{p}, \boldsymbol{r}) = (p^1, p^2, p^3, x^1, x^2, x^3)$.

It is remarkable that there is a Hamiltonian action of the Poincaré group $\mathfrak{P}$ on $\mathbb{R}^6$!

Indeed, let $\mathscr{L}$ be the set of all timelike straight line in $\mathbb{R}^4$. Every $l \in \mathscr{L}$ has the form $l = \{x + sv,\ s \in \mathbb{R}\}$, where $x, v \in \mathbb{R}^4$ and v is timelike, $v^\mu v_\mu > 0$. The Poincaré group $\mathfrak{P}$ acts on $\mathscr{L}$ by

$$(\Lambda, a)(l) = \{\Lambda x + a + s\Lambda v\}.$$

Each timelike l admits a unique representation $l = \{x + sv,\ s \in \mathbb{R}\}$ where $x = (0, \boldsymbol{r})$ and $v = (c, \boldsymbol{v})$ with $v = |\boldsymbol{v}| < c$. Thus $\mathscr{L} \cong \mathbb{R}^3 \times \mathbb{B}(0, c)$, which is isomorphic to $\mathbb{R}^6$ by the Legendre transform $\boldsymbol{v} \mapsto \boldsymbol{p}$, and we obtain the Poincaré group action on $\mathbb{R}^6$.

This action preserves the symplectic form and is Hamiltonian. Specifically, the action of the Euclidean group $E(3) < \mathfrak{P}$ on $\mathbb{R}^6 \cong \mathbb{R}^3 \times \mathbb{B}(0, c)$ is Hamiltonian with the Hamiltonian functions

$$J_1 = x^2p^3 - x^2p^3, \quad J_2 = x^3p^1 - x^1p^3 \quad J_3 = x^1p^2 - x^2p^1$$

(see Example 6.1 in Lecture 6) and $P_i = -p^i$. Indeed, abelian group of translations of $\mathbb{R}^3$ acts on $\mathbb{R}^6$ by $(\boldsymbol{p}, \boldsymbol{r}) \mapsto (\boldsymbol{p}, \boldsymbol{r} + \boldsymbol{a})$ and the corresponding vector field $X_{\boldsymbol{a}}$ is given by

$$X_{\boldsymbol{a}}(f)(\boldsymbol{p}, \boldsymbol{r}) = \left.\frac{d}{du}\right|_{u=0} f(\boldsymbol{p}, \boldsymbol{r} - \boldsymbol{a}) = -a^i \frac{\partial f}{\partial x^i}(\boldsymbol{p}, \boldsymbol{r}).$$

Thus the vector fields $X_{\boldsymbol{e}_i}$ are Hamiltonian vector fields with Hamiltonian functions $-p^i$, i.e.,

$$X_{\boldsymbol{e}_i} = -\frac{\partial}{\partial x^i} = -J(dp^i), \quad i = 1, 2, 3.$$

The one-parameter subgroup T of time translations acts on $\mathscr{L}$ by $l \mapsto l + (x^0, 0, 0, 0)$ with the representative $(\boldsymbol{r} - x^0\boldsymbol{v}/c, \boldsymbol{v})$. Thus T acts on $\mathbb{R}^6$ by

$$\boldsymbol{r} \mapsto \boldsymbol{r} - \frac{x^0\boldsymbol{p}}{p^0}, \quad \boldsymbol{p} \mapsto \boldsymbol{p}$$

and the corresponding vector field is $X = \dfrac{p^i}{p^0}\dfrac{\partial}{\partial x^i}$. Using that

$$J(d\boldsymbol{p}) = \frac{\partial}{\partial \boldsymbol{r}} \quad \text{and} \quad J(d\boldsymbol{r}) = -\frac{\partial}{\partial \boldsymbol{p}},$$

[1] Note that in accordance with Sect. 15.2 in Lecture 18 we have $\boldsymbol{p} = (p^1, p^2, p^3)$.

(see Sect. 4.3 in Lecture 4) we obtain that $X = J(dp^0)$, i.e., X is a Hamiltonian vector with with the Hamiltonian function is $p^0 = \sqrt{p^2 + m^2c^2}$, i.e., is $1/c$ times the Hamiltonian of a free relativistic particle of mass m.

Next, consider the one-parameter subgroup $\mathscr{K}_1$ of $\mathscr{P}$ which consists on Lorentz boosts in x^0x^1-planes,

$$\Lambda(\psi)x = (x^0 \cosh\psi + x^1 \sinh\psi, x^0 \sinh\psi + x^1 \cosh\psi, x^2, x^3), \quad \psi \in \mathbb{R}.$$

To find the action of $\Lambda(\psi)$ on $\mathbb{R}^6$ we need to determine how in acts on the representative $(\boldsymbol{r}, \boldsymbol{v})$ of a straight line l. We have

$$\begin{aligned}
\Lambda(\psi)(0, \boldsymbol{r}) &= (x^1 \sinh\psi, x^1 \cosh\psi, x^2, x^3),\\
\Lambda(\psi)(c, \boldsymbol{v}) &= (c\cosh\psi + v^1 \sinh\psi, c \sinh\psi + v^1 \cosh\psi, v^2, v^3),
\end{aligned}$$

so that

$$\Lambda(\psi)(\boldsymbol{v}) = \left(\frac{cv^1 \cosh\psi + c^2 \sinh\psi}{v^1 \sinh\psi + c\cosh\psi}, \frac{cv^2}{v^1 \sinh\psi + c\cosh\psi}, \frac{cv^3}{v^1 \sinh\psi + c\cosh\psi}\right)$$

and from this we obtain

$$\begin{aligned}
\Lambda(\psi)(\boldsymbol{r}) = \Bigg(&x^1 \cosh\psi - x^1 \sinh\psi \frac{v^1 \cosh\psi + c \sinh\psi}{v^1 \sinh\psi + c\cosh\psi},\\
&x^2 - \frac{x^1v^2 \sinh\psi}{v^1 \sinh\psi + c\cosh\psi}, x^3 - \frac{x^1v^3 \sinh\psi}{v^1 \sinh\psi + c\cosh\psi}\Bigg)\\
= \Bigg(&\frac{cx^1}{v^1 \sinh\psi + c\cosh\psi}, x^2 - \frac{x^1v^2 \sinh\psi}{v^1 \sinh\psi + c\cosh\psi}, x^3 - \frac{x^1v^3 \sinh\psi}{v^1 \sinh\psi + c\cosh\psi}\Bigg).
\end{aligned}$$

Using (16.3), we get

$$\Lambda(\psi)(\boldsymbol{r}) =$$
$$\left(\frac{x^1p_0}{p^1 \sinh\psi + p_0 \cosh\psi}, x^2 - \frac{x^1p^2 \sinh\psi}{p^1 \sinh\psi + p_0 \cosh\psi}, x^3 - \frac{x^1p^3 \sinh\psi}{p^1 \sinh\psi + p_0 \cosh\psi}\right).$$

To obtain the action of the Lorentz boost on the momentum vector $\boldsymbol{p}$ we need to use equation (16.2). Namely, $\Lambda(\psi)(\boldsymbol{p}) = \tilde{\boldsymbol{p}}$ is relativistic momentum for the velocity vector $\tilde{\boldsymbol{v}} = \Lambda(\psi)(\boldsymbol{v})$. Denoting $\tilde{v} = |\tilde{\boldsymbol{v}}|$ we get

$$1 - \frac{\tilde{v}^2}{c^2} = \frac{c^2}{(v^1 \sinh\psi + c\cosh\psi)^2}\left(1 - \frac{v^2}{c^2}\right).$$

Using

$$p^0 = \frac{mc}{\sqrt{1 - \dfrac{v^2}{c^2}}},$$

we obtain

$$\tilde{\boldsymbol{p}} = \frac{\tilde{\boldsymbol{v}}}{\sqrt{1 - \dfrac{\tilde{v}^2}{c^2}}} = (p^1 \cosh\psi + p^0 \sinh\psi, p^2, p^3),$$

so that

$$\Lambda(\psi)(\boldsymbol{p}) = (p^1 \cosh\psi + p^0 \sinh\psi, p^2, p^3).$$

The vector field corresponding to the $\mathscr{K}_1$ action on $\mathbb{R}^6$ is given by

$$\begin{aligned} X_1(f)(\boldsymbol{p}, \boldsymbol{r}) &= \frac{d}{d\psi}\bigg|_{\psi=0} f(\Lambda(-\psi)\boldsymbol{p}, \Lambda(-\psi)\boldsymbol{r}) \\ &= \frac{x^1}{p^0}\left(p^1\frac{\partial}{\partial x^1} + p^2\frac{\partial}{\partial x^2} + p^3\frac{\partial}{\partial x^3}\right) - p^0\frac{\partial}{\partial p^1}. \end{aligned}$$

Thus we obtained that X is a Hamiltonian vector field with the Hamiltonian function $K_1(\boldsymbol{p}, \boldsymbol{r}) = x^1\sqrt{p^2 + m^2c^2}$, i.e.,

$$X = J(dK_1).$$

Similarly, we see that vector fields X_2 and X_3 for one-parameter subgroups $\mathscr{K}_2$ and $\mathscr{K}_3$ are Hamiltonian vector field with the Hamiltonian function $K_2(\boldsymbol{p}, \boldsymbol{r}) = x^2\sqrt{p^2 + m^2c^2}$ and $K_3(\boldsymbol{p}, \boldsymbol{r}) = x^3\sqrt{p^2 + m^2c^2}$.

Since Hamiltonian vector fields preserves symplectic form, the Poincaré group $\mathfrak{P}$ acts on $\mathbb{R}^6$ by canonical transformations (symplectomorphisms). The following theorem summarizes obtained results.

THEOREM 16.1. *The defined above action of the Poincaré group $\mathfrak{P}$ on the phase space $\mathbb{R}^6$ of free relativistic particle with mass m is Hamiltonian. The Hamiltonian functions corresponding to space-time translations, space rotations and Lorentz boosts are*

$$P_0 = \sqrt{p^2 + m^2c^2}, \quad P_i = -p^i, \quad J_i = \varepsilon_{ijk}x^j p^k, \quad K_i = x^i\sqrt{p^2 + m^2c^2},$$

$i = 1, 2, 3$. *They satisfy the following Poisson brackets*

$$\{P_i, P_j\} = \{P_i, P_0\} = \{J_i, P_0\} = 0, \quad \{J_i, J_j\} = -\varepsilon_{ijk}J_k, \tag{16.4}$$

$$\{K_i, K_j\} = \varepsilon_{ijk}J_l, \quad \{J_i, K_j\} = -\varepsilon_{ijk}K_k, \tag{16.5}$$

$$\{K_i, P_0\} = P_i, \quad \{K_i, P_i\} = -\delta_{ij}P_0, \quad \{J_i, P_j\} = -\varepsilon_{ijk}P_k. \tag{16.6}$$

PROOF. Straightforward computation using the Poisson bracket

$$\{f, g\}(\boldsymbol{p}, \boldsymbol{r}) = \frac{\partial f}{\partial \boldsymbol{p}}\frac{\partial g}{\partial \boldsymbol{r}} - \frac{\partial f}{\partial \boldsymbol{r}}\frac{\partial g}{\partial \boldsymbol{p}}.$$

□

REMARK. As in Example 6.1 in Lecture 6, Poisson brackets between Hamiltonian functions have the same form as Lie brackets of the corresponding generators of Poincaré Lie algebra, taken with the negative sign.

Using that $cp^0 = \mathscr{H}$, the Hamiltonian of a free particle, we obtain from (16.4)–(16.6),

$$\{J_i, x^j\} = -\varepsilon_{ijk}x^k, \tag{16.7}$$

$$c\{K_i, x^j\} = x^i\{\mathscr{H}, x^j\}, \tag{16.8}$$

$$\{P_i, x^j\} = -\delta_{ij}, \quad i, j = 1, 2, 3. \tag{16.9}$$

These Poisson brackets exemplify that $\mathbb{R}^6$ is a phase space of a relativistic particle.

16.3. No-interaction theorem

The relativity principle imposes very strong restriction on Hamiltonian systems: it implies that the interaction of finitely many relativistic particles is not possible! The precise statement is the following.

THEOREM 16.2. *Consider the Hamiltonian system of of n particles with the phase space $\mathbb{R}^{6n}$, the symplectic form*

$$\omega = \sum_{a=1}^{n} d\boldsymbol{p}_a \wedge d\boldsymbol{r}_a,$$

where $\boldsymbol{r}_a$ and $\boldsymbol{p}_a$ are coordinates and momenta of the a-th particle, and with the Hamiltonian function $\mathscr{H}$. Suppose that $(\mathbb{R}^{6n}, \omega, \mathscr{H})$ is a system of n relativistic particles, that is, the principle of relativity holds in the following form:

a) *There exists a set of ten generators of the Poincaré Lie algebra — ten functions $P_0 = \mathscr{H}/c$, P_i, J_i and K_i on $\mathbb{R}^{6n}$ with Poisson brackets (16.4)–(16.6).*

b) *The coordinates of the particles transform correctly under the Poincaré group — coordinates $\boldsymbol{r}_a$, $a = 1, \dots, n$, and the generators of the Poincaré Lie algebra have Poisson brackets (16.7)–(16.9).*

In addition, suppose that the system is non-degenerate,

$$\det \left\{ \frac{\partial^2 \mathscr{H}}{\partial p_a^i \partial p_b^j} \right\} \neq 0.$$

Then the acceleration of each particle vanishes,

$$\{\mathscr{H}, \{\mathscr{H}, \boldsymbol{r}_a\}\} = 0, \quad a = 1, \dots, n.$$

Equivalently, there are Darboux coordinates $\tilde{\boldsymbol{p}}_a$ and $\boldsymbol{r}_a$ (the coordinates of the particles are unchanged) and the constants $m_a > 0$ such that

$$\boldsymbol{P} = -\sum_{a=1}^{n} \tilde{\boldsymbol{p}}_a,$$

$$\mathscr{H} = \sum_{a=1}^{n} c\sqrt{\tilde{\boldsymbol{p}}_a^2 + m_a^2 c^2},$$

$$J_i = \sum_{a=1}^{n} \varepsilon_{ijk} x_a^j \tilde{p}_a^k,$$

$$K_i = \sum_{a=1}^{n} x_a^i \sqrt{\tilde{\boldsymbol{p}}_a^2 + m_a^2 c^2}.$$

The theorem is a manifestation of the fundamental fact that relativistic invariant Hamiltonian systems of interacting particles in Minkowski spacetime should have infinitely many of degrees of freedom, and the interaction is described by by a field theory. The examples we have seen so far are the theory of electromagnetism and a charged relativistic particle interacting with the external electromagnetic field. Another fundamental example in classical physics is the theory of gravity and a massive relativistic particle interacting with the external gravitational field.

PROBLEM 16.1. Prove the no-interaction theorem for $n = 1$.

LECTURE 17

General relativity

Newton's law of universal gravitation states that a particle with mass m_1 at point $\boldsymbol{r}_1$ attracts a particle with mass m_2 at point $\boldsymbol{r}_2$ with the force

$$\boldsymbol{F}_2 = -Gm_1m_2\frac{\boldsymbol{r}_2 - \boldsymbol{r}_1}{|\boldsymbol{r}_2 - \boldsymbol{r}_1|^3}$$

and $\boldsymbol{F}_1 = -\boldsymbol{F}_2$. Obviously the Newton's law is not a Lorentz invariant and one needs to find a Lorentz invariant description of gravity.

The first attempt[1] was to include the theory of gravity into the special relativity by assuming that gravitation field is determined by the four potential A^G_μ. The interaction of a relativistic particle of charge e and mass m would be described by the action

$$S = -mc\int ds - \frac{e}{c}\int A_\mu dx^\mu - m\int A^G_\mu dx^\mu.$$

Considering the case $e = 0$ and using $A^G_\mu = (\varphi, 0, 0, 0)$, one gets a Lorentz invariant modification of Newton's law of universal gravitation,

$$\frac{d\boldsymbol{p}}{dt} = -m\frac{\partial\varphi}{\partial\boldsymbol{r}}, \quad \boldsymbol{p} = \frac{m\boldsymbol{v}}{\sqrt{1 - \dfrac{v^2}{c^2}}}.$$

However, this approach does not give a correct answer for the precession of the perihelion of Mercury.

17.1. Spacetime in general relativity

A smooth connected four-manifold M is called a *Lorentzian manifold* if it carries a pseudo-Riemannian metric

$$ds^2 = g_{\mu\nu}(x)dx^\mu dx^\nu$$

with the signature $(+, -, -, -)$ at every $x \in M$. The Minkowski space is a non-compact Lorentzian manifold, and it is easy to see that every non-compact manifold admits a Lorentzian metric. However, a compact manifold M admits a Lorentzian metric if and only if its Euler characteristic vanishes. In other

[1] A. Poincaré in 1905.

words, a manifold M admits Lorentzian metric if and only if is has nowhere vanishing vector filed[2].

As for the case of Minkowski metric, a tangent vector $v \in T_x M$ is timelike, null, or spacelike if, respectively, its length is positive, zero, or negative. A curve $\gamma : [u_1, u_2] \to M$ is timelike if $\gamma'(u)$ is timelike for all $u \in [u_1, u_2]$ and is *causal* if if $\gamma'(u)$ is timelike or null for all $u \in [u_1, u_2]$. A Lorentzian manifold M is *time-orientable* if admits a timelike vector field $X \in \mathrm{Vec}(M)$ which defines a *time orientation* of M. The opposite time orientation is given by the vector field $-X$. Specifically, a timelike or null vector $u \in T_x M$ is *future-directed* (or *past-directed*), if $u \cdot X_x > 0$ (or $u \cdot X_x < 0$). A timelike curve $\gamma : [u_1, u_2] \to M$ is future-directed (or past-directed), if $\gamma'(u)$ is future-directed (or past-directed) for all $u \in [u_1, u_2]$.

Definition. A *spacetime* is time-oriented Lorentzian four-manifold M.

Definition. The *chronological future* $I_+^M(x)$ of $x \in M$ is the set of points that can be reached from x by future-directed timelike curves. The *causal future* $J_+^M(x)$ of $x \in M$ is the set of points that can be reached from x by future-directed causal curves and of x itself. Similarly, the *chronological past* $I_-^M(x)$ and *causal past* $J_-^M(x)$ of $x \in M$ are defined by using past-directed timelike and causal curves.

Proposition 17.1. *If the spacetime M is compact, there exists a closed timelike curve in M.*

Proof. The familiy $\{I_+^M(x)\}_{x \in M}$ is an open covering of M. By compactness, $M = I_+^M(x_1) \cup \cdots \cup I_+^M(x_m)$. If $x_1 \in I_+^M(x_2) \cup \cdots \cup I_+^M(x_m)$, then $x_1 \in I_+^M(x_k)$ for some $2 \le k \le m$. Then $I_+^M(x_1) \subseteq I_+^M(x_k)$ and we can omit $I_+^M(x_1)$ from the covering. Thus $x_1 \in I_+^M(x_1)$, so that there is a timelike future-directed curve starting and ending in x_1. $\square$

Since this allows for the time travel, we will consider only non-compact spacetimes. Recall that a piecewise C^1-curve in M is called *inextendible*, if no piecewise C^1-reparametrization of the curve can be continuously extended beyond any of the end points of the parameter interval. A set S is called *achronal* if there is no timelike curve which intersects S twice.

Definition. An achronal hypersurface Σ in M is a *Cauchy hypersurface* if every inextendible causal curve intersects Σ exactly once.

Proposition 17.2. *If a spacetime M admits two Cauchy hypersurfaces Σ_1 and Σ_2, then Σ_1 is diffeomorphic to Σ_2.*

Definition. A spacetime M satisfies the *causality condition* if it does not contain any closed causal curve. A spacetime M satisfies the *strong causality condition* if there are no almost closed causal curves. That is, for each $x \in M$

[2]Indeed, according to a theorem by Steenrod, a compact manifold admits everywhere defined, continuous quadratic form of signature k if and only if it admits a continuous field of tangent k planes.

and for each open neighborhood U of x there exists an open neighborhood $V \subseteq U$ of x such that no causal curve in M intersects V more then once.

Clearly the strong causality condition implies the causality condition.

DEFINITION. A space-time M is *globally hyperbolic* if it satisfies the strong causality condition and for all $x, y \in M$ the intersection $J_+^M(x) \cap J_-^M(y)$ is compact.

The following fundamental result holds[3]. It describes the structure of globally hyperbolic spacetimes explicitly: they are foliated by smooth spacelike Cauchy hypersurfaces.

THEOREM 17.1. *Let M be a spacetime M. The following are equivalent.*

(1) *M is globally hyperbolic.*

(2) *There exists a Cauchy hypersurface in M.*

(3) *M is isometric to $\mathbb{R} \times \Sigma$ with the Lorentzian metric $\beta dt^2 - \gamma_t$, where β is a smooth positive function on M, γ_t is is a Riemannian metric on Σ depending smoothly on $t \in \mathbb{R}$ and each $\{t\} \times \Sigma$ is a smooth spacelike Cauchy hypersurface in M.*

COROLLARY 17.2. *On every globally hyperbolic spacetime M there exists a smooth function $h : M \to \mathbb{R}$ whose gradient $\boldsymbol{\nabla} h \in \mathrm{Vect}(M)$ is timelike and future-directed and all level sets of h are spacelike Cauchy hypersurfaces.*

Such function h is called a *Cauchy time function* and its gradient $\boldsymbol{\nabla} h$ is defined by

$$\boldsymbol{\nabla} h = g^{\mu\nu} \frac{\partial h}{\partial x^\mu} \frac{\partial}{\partial x^\nu},$$

where $g^{\mu\nu}$ is the inverse matrix. In fact[4], for every Cauchy hypersurface Σ in M there is a Cauchy time function h such that $\Sigma = h^{-1}(0)$.

From physics point of view, a *proper time* τ along a timelike curve γ is defined by

$$\tau(u) = \frac{1}{c} \int_{u_1}^{u} ds,$$

where the integration goes over γ. It is natural to consider only those coordinates x^μ for which x^0 plays a role of a time variable, and x^1, x^2, x^3 are space coordinates. Specifically, two events occurring at a same point (x^1, x^2, x^3) in space should be connected by a timelike curve $\gamma(u) = (x^0(u), x^1, x^2, x^3)$. This implies that $g_{00} > 0$ and the proper time between these two events is

$$\tau = \frac{1}{c} \int \sqrt{g_{00}}\, dx^0.$$

[3]Bernal, A.N., Sánchez, M.: *Smoothness of time functions and the metric splitting of globally hyperbolic spacetimes*, Commun. Math. Phys. **257** (2005), 43.

[4]Bernal, A.N., Sánchez, M.: *Further results on the smoothability of Cauchy hypersurfaces and Cauchy time functions*, Lett. Math. Phys. **77** (2006), 183.

To determine the metric $dl^2 = \gamma_{ij}dx^i dx^j$ in space induced by ds^2 we cannot simply put $dx^0 = 0$ since proper time at different points in space depend differently on the coordinate x^0. However,

$$ds^2 = g_{00}(dx^0)^2 + 2g_{0i}dx^0 dx^i + g_{ij}dx^i dx^j = g_{00}\left(dx^0 + \frac{g_{0i}}{g_{00}}dx^i\right)^2 - \gamma_{ij}dx^i dx^j,$$

where

$$\gamma_{ij} = -g_{ij} + \frac{g_{0i}g_{0j}}{g_{00}}, \quad i,j = 1,2,3 \tag{17.1}$$

is a three-dimensional metric tensor. Since $g_{00} > 0$ it is a Riemannian metric tensor. It depends on x^0 so that the distance in real space depends on time. The relation

$$dx^0 + \frac{g_{0i}}{g_{00}}dx^i = 0$$

can be integrated over any curve in space to define x^0 along the curve. This allows to synchronize the clocks in general relativity along any curve in space. However, this synchronization depends on a curve connecting two points in space. Proposition 17.1 asserts that for a globally hyperbolic spacetime one can choose coordinates such that g_{0i} vanish and one can synchronize clocks over all space. The corresponding coordinates (*reference system* in physics terminology) are called *syncrhonous*.

It is easy to see from (17.1) that

$$-\gamma_{ij}g^{jk} = \delta_{ik}.$$

The relations $g_{00} > 0$ and γ_{ij} is positive-definite 3×3 matrix are equivalent to the

$$g_{00} > 0, \ \det\begin{pmatrix} g_{00} & g_{01} \\ g_{10} & g_{11}\end{pmatrix} < 0, \ \det\begin{pmatrix} g_{00} & g_{01} & g_{02} \\ g_{10} & g_{11} & g_{12} \\ g_{20} & g_{21} & g_{22}\end{pmatrix} > 0$$

and

$$g = \det\begin{pmatrix} g_{00} & g_{01} & g_{02} & g_{03} \\ g_{10} & g_{11} & g_{12} & g_{13} \\ g_{20} & g_{21} & g_{22} & g_{23} \\ g_{30} & g_{31} & g_{32} & g_{33}\end{pmatrix} < 0.$$

Physically these conditions should hold for any choice of coordinates on M which can be realized with the aid of "physical bodies".

17.2. Particle in a gravitation field

A gravitational field is a change of a metric of a space-time and is described by the metric tensor $g_{\mu\nu}(x)$. The action of a relativistic particle of mass m in a gravitational field has the same form as in Lecture 15,

$$S(\gamma) = -mc\int ds = -mc\int \sqrt{g_{\mu\nu}u^\mu u^\nu}ds, \quad u^\mu = \frac{dx^\mu}{ds}.$$

In other words, the action functional is $-mc$ times the length functional in the pseudo-Riemannian geometry. Correspondingly, the Euler-Lagrange equations are the geodesic equations with respect to the natural parameter,

$$\frac{d^2x^\lambda}{ds^2} + \Gamma^\lambda_{\mu\nu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} = 0,$$

where

$$\Gamma^\lambda_{\mu\nu} = \frac{1}{2}g^{\lambda\sigma}\left(\frac{\partial g_{\mu\sigma}}{\partial x^\nu} + \frac{\partial g_{\nu\sigma}}{\partial x^\mu} - \frac{\partial g_{\mu\nu}}{\partial x^\sigma}\right) \tag{17.2}$$

are Christoffel's symbols. The free particle in a gravitational field moves along the geodesics.

17.3. The Riemann tensor

Recall that the metric $g_{\mu\nu}(x)$ on the spacetime M determines a Levi-Civita connection[5] in the tangent bundle TM. Explicitly it is given by

$$\boldsymbol{\nabla} = d + A, \quad \text{where} \quad A = A_\mu dx^\mu.$$

Here $A_\mu(x)$ are linear operators in T_xM which in the basis $\dfrac{\partial}{\partial x^\mu}$ are given by the matrices

$$(A_\mu)^\lambda_\nu = \Gamma^\lambda_{\nu\mu}. \tag{17.3}$$

Thus a derivative of a $(1,0)$-tensor, a vector field $V = v^\mu\dfrac{\partial}{\partial x^\mu}$, in the direction $\dfrac{\partial}{\partial x^\mu}$ is given by

$$(\boldsymbol{\nabla}_\mu V)^\lambda = \frac{\partial v^\lambda}{\partial x^\mu} + \Gamma^\lambda_{\nu\mu}v^\nu,$$

while a derivative of a $(0,1)$-tensor, a 1-form $\theta = a_\mu dx^\mu$, is

$$(\boldsymbol{\nabla}_\mu\theta)_\lambda = \frac{\partial a_\lambda}{\partial x^\mu} - \Gamma^\nu_{\lambda\mu}a_\nu.$$

Directional derivative $\boldsymbol{\nabla}_\mu$ of an arbitrary (p,q)-tensor is defined similarly. We have

$$\boldsymbol{\nabla}_\lambda g_{\mu\nu} = 0 \quad \text{and} \quad \boldsymbol{\nabla}_\lambda g^{\mu\nu} = 0. \tag{17.4}$$

The curvature of the connection $\boldsymbol{\nabla}$ is $F = dA + A\wedge A$, a 2-form with values in $\operatorname{End} TM$ (see Sect. 9.1 in Lecture 9). We have

$$F = \sum_{\mu<\nu} F_{\mu\nu}dx^\mu\wedge dx^\nu,$$

[5] A metric connection with no torsion.

where
$$F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu} + [A_\mu, A_\nu].$$
On 2-forms B with values in $\operatorname{End} TM$ the connection $\boldsymbol{\nabla}$ acts by
$$\boldsymbol{\nabla} B = dB + A \wedge B - B \wedge A,$$
which gives the Bianci identity
$$\boldsymbol{\nabla} F = 0$$
for a curvature 2-form. Equivalently,
$$\boldsymbol{\nabla}_\lambda F_{\mu\nu} + \boldsymbol{\nabla}_\mu F_{\nu\lambda} + \boldsymbol{\nabla}_\nu F_{\lambda\mu} = 0.$$

Using (17.3), we obtain the following formula for the Riemann curvature tensor $R^\lambda{}_{\rho\mu\nu} = (F_{\mu\nu})^\lambda_\rho$,

$$R^\lambda{}_{\rho\mu\nu} = \frac{\partial \Gamma^\lambda_{\rho\nu}}{\partial x^\mu} - \frac{\partial \Gamma^\lambda_{\rho\mu}}{\partial x^\nu} + \Gamma^\lambda_{\sigma\mu}\Gamma^\sigma_{\rho\nu} - \Gamma^\lambda_{\sigma\nu}\Gamma^\sigma_{\rho\mu}. \tag{17.5}$$

The Bianci identity for the Riemann tensor has the form

$$\boldsymbol{\nabla}_\sigma R^\lambda{}_{\rho\mu\nu} + \boldsymbol{\nabla}_\nu R^\lambda{}_{\rho\sigma\mu} + \boldsymbol{\nabla}_\mu R^\lambda{}_{\rho\nu\sigma} = 0. \tag{17.6}$$

The Ricci curvature
$$R_{\mu\nu} = R^\lambda{}_{\mu\lambda\nu}$$
is the trace of the Riemann tensor and is given explicitly by

$$R_{\mu\nu} = \frac{\partial \Gamma^\lambda_{\mu\nu}}{\partial x^\lambda} - \frac{\partial \Gamma^\lambda_{\mu\lambda}}{\partial x^\nu} + \Gamma^\lambda_{\mu\nu}\Gamma^\sigma_{\lambda\sigma} - \Gamma^\sigma_{\mu\lambda}\Gamma^\lambda_{\sigma\nu}. \tag{17.7}$$

It follows from (17.2) that
$$\begin{aligned}\Gamma^\lambda_{\mu\lambda} &= \frac{1}{2} g^{\lambda\sigma}\left(\frac{\partial g_{\mu\sigma}}{\partial x^\lambda} + \frac{\partial g_{\lambda\sigma}}{\partial x^\mu} - \frac{\partial g_{\mu\lambda}}{\partial x^\sigma}\right)\\ &= \frac{1}{2} g^{\lambda\sigma}\frac{\partial g_{\sigma\lambda}}{\partial x^\mu}\\ &= \frac{1}{2g}\frac{\partial g}{\partial x^\mu} = \frac{\partial \log\sqrt{-g}}{\partial x^\mu}.\end{aligned}$$
Thus the Ricci tensor is symmetric, $R_{\mu\nu} = R_{\nu\mu}$, and determines a symmetric bilinear form $R_{\mu\nu}dx^\mu dx^\nu$ on the tangent space.

Finally, the scalar curvature R is the trace of Ricci curvature tensor,
$$R = g^{\mu\nu} R_{\mu\nu}.$$
Contracting λ and ν in (17.6), we get
$$2\boldsymbol{\nabla}_\mu R_{\rho\sigma} - \boldsymbol{\nabla}_\sigma R_{\rho\mu} = 0$$

and using (17.4) we obtain

$$2\nabla_\mu R^\rho_\sigma - \nabla_\sigma R^\rho_\mu = 0.$$

Finally contracting μ and ρ we get

$$2\nabla_\mu R^\mu_\sigma - \nabla_\sigma R = 0,$$

or

$$\nabla_\mu \left(R^\mu_\nu - \frac{1}{2}\delta^\mu_\nu R \right) = 0. \tag{17.8}$$

LECTURE 18

Einstein equations – I

18.1. Einstein field equations

In general relativity the Lorentzian metric $g_{\mu\nu}$ of the space-time M satisfies *Einstein equations*

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{8\pi G}{c^4} T_{\mu\nu},$$

where $R_{\mu\nu}$ is the Ricci curvature, R is the scalar curvature and $T_{\mu\nu}$ is the stress-energy tensor of matter. It is defined as

$$T_{\mu\nu} = \frac{\delta S_{\text{matter}}}{\delta g^{\mu\nu}}.$$

It follows from Bianci identity (17.8) that Einstein equations imply that necessarily

$$\boldsymbol{\nabla}_\mu T^\mu_\nu = 0, \quad \nu = 0, 1, 2, 3.$$

These are conservation laws in general relativity.

Rewriting Einstein equations in the form

$$R^\mu_\nu - \frac{1}{2} \delta^\mu_\nu \, R = \frac{8\pi G}{c^4} \, T^\mu_\nu$$

and taking traces we obtain

$$R = -\frac{8\pi G}{c^4} T,$$

where $T = T^\mu_\mu$. Thus Einstein equations can be also written as

$$R^\mu_\nu = \frac{8\pi G}{c^4} \left(T^\mu_\nu - \frac{1}{2} \delta^\mu_\nu \, T \right). \tag{18.1}$$

In particular, the empty space Einstein equations reduces to

$$R_{\mu\nu} = 0.$$

18.2. Particle in a weak gravitational field

Here we solve the geodesic equation and Einstein equations in case of a *weak gravitational field*. Namely, suppose that $M = \mathbb{R}^4$ and

$$g_{\mu\nu}(x) = \eta_{\mu\nu} + \frac{1}{c^2} g^{(2)}_{\mu\nu}(x) + O\left(\frac{1}{c^3}\right), \tag{18.2}$$

where $\eta_{\mu\nu}$ is Minkowski metric. It is also assumed that these asymptotics can be differentiated with respect to x^μ.

Timelike geodesic is *slow* if $\dot{x}^i(t) \ll c$, where $i = 1, 2, 3$ and $t = x^0/c$. Since

$$d\tau = \frac{1}{c}\sqrt{g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu}dt = \left(1 + O\left(\frac{1}{c^2}\right)\right)dt,$$

the equation for slow geodesic takes the form

$$\frac{d^2x^\lambda}{dt^2} + \Gamma^\lambda_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt} = O\left(\frac{1}{c}\right).$$

It follows from (18.2) that

$$\Gamma^0_{00} = O\left(\frac{1}{c^3}\right), \quad \Gamma^i_{00} = -\frac{1}{2}\frac{\partial g^2_{00}}{\partial x^i} + O\left(\frac{1}{c}\right),$$

and all other Christoffel's symbols are of order $O(1/c^2)$. Putting

$$g^2_{00}(x) = 2\varphi(x^0, \boldsymbol{r})$$

we see that up to the order $O(1/c)$ the geodesic equation becomes Newton's equation

$$\ddot{\boldsymbol{r}} = -\frac{\partial\varphi}{\partial\boldsymbol{r}},$$

and the force acting on a particle is $\boldsymbol{F} = -m\dfrac{\partial\varphi}{\partial\boldsymbol{r}}$.

To find the potential φ we need to use Einstein equations. The energy-momentum tensor of a macroscopic body which consists of slow moving particles is given by

$$T^{\mu\nu} = M(x)c^2u^\mu u^\nu,$$

where $M(x)$ is the mass density of the body and u^μ is a four-velocity vector. If the macroscopic motion of the body is slow, we can put $u^0 = 1$ and $u^i = 0$, $i = 1, 2, 3$. Thus the energy-momentum tensor takes the form

$$T^\mu_\nu = Mc^2\delta^\mu_0\delta^0_\nu.$$

It follows from formula (17.7) in Lecture 17 that in the weak gravitational field $R^\mu_\nu = O(1/c^2)$ and the only nontrivial contribution to Einstein equation (18.1) is

$$R^0_0 = \frac{4\pi G}{c^4}T = \frac{4\pi M}{c^2}.$$

Since

$$R^0_0 = \frac{\partial\Gamma^i_{00}}{\partial x^i} + O\left(\frac{1}{c^3}\right) = \frac{1}{c^2}\boldsymbol{\nabla}^2\varphi + O\left(\frac{1}{c^3}\right),$$

Einstein equations for the weak gravitational field reduce to the Poisson equation

$$\boldsymbol{\nabla}^2\varphi = 4\pi M$$

for the gravitational potential. Namely,

$$\varphi(\boldsymbol{r}) = -G\int \frac{M(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}d^3\boldsymbol{r}'$$

and in case $M(\boldsymbol{r}') = M\delta(\boldsymbol{r}-\boldsymbol{r}')$ we obtain Newtonian potential

$$\varphi(\boldsymbol{r}) = -\frac{GM}{\boldsymbol{r}}.$$

So that the force acting on a slow particle of mass m in a weak gravitational force generated by a particle of a mass M is the Newtonian force!

18.3. Hilbert action

On the space $\mathscr{M}$ of smooth Lorentzian metrics on the spacetime M consider the celebrated Hilbert (or Hilbert-Einstein) functional

$$S(g_{\mu\nu}) = \int R\sqrt{-g}\, d^4x,$$

where R is the scalar curvature of the metric $ds^2 = g_{\mu\nu}dx^\mu dx^\nu \in \mathscr{M}$, and $\sqrt{-g}\, d^4x$ is the corresponding volume form on M. Here integration goes over a domain D in M (usually bounded by two spacelike Cauchy hypersurfaces) and it is assumed that all metrics in $\mathscr{M}$ have the same boundary value on ∂D. In addition, normal derivatives of $g_{\mu\nu}$ on ∂D are fixed.

Proposition 18.1. *Let $u_{\mu\nu} = \delta g_{\mu\nu}$ be a tangent vector to $\mathscr{M}$ at a point $g_{\mu\nu} \in \mathscr{M}$ and $u^{\mu\nu} = g^{\mu\alpha}g^{\nu\beta}u_{\alpha\beta}$. Then the Gato derivative of the Hilbert functional S in the direction u is given by*

$$\delta_u S = \int_D \left(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R\right)u^{\mu\nu}\sqrt{-g}\, d^4x.$$

Proof. Putting

$$\delta S = \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} S_{\mathrm{EH}}(g_{\mu\nu}+\varepsilon\delta g_{\mu\nu})$$

we have

$$\delta S = \int_D (\delta g^{\mu\nu}R_{\mu\nu} + g^{\mu\nu}\delta R_{\mu\nu})\sqrt{-g}\, d^4x + \int_D R\delta(\sqrt{-g})d^4x.$$

To compute $\delta R_{\mu\nu}(x)$ we use geodesic normal coordinates at $x \in M$ to obtain

$$\delta R_{\mu\nu} = \frac{\delta\Gamma^\sigma_{\mu\nu}}{\partial x^\sigma} - \frac{\delta\Gamma^\sigma_{\mu\sigma}}{\partial x^\nu}.$$

Since $\delta\Gamma^\lambda_{\mu\nu}$ is a $(1,2)$ tensor, we get the formula

$$\delta R_{\mu\nu} = \boldsymbol{\nabla}_\sigma\delta\Gamma^\sigma_{\mu\nu} - \boldsymbol{\nabla}_\nu\delta\Gamma^\sigma_{\mu\sigma},$$

called *Palatini identity*. Since $\boldsymbol{\nabla}_\sigma g^{\mu\nu} = 0$, we obtain from the Palatini identity

$$g^{\mu\nu}\delta R_{\mu\nu} = \boldsymbol{\nabla}_\sigma(g^{\mu\nu}\delta\Gamma^\sigma_{\mu\nu}) - \boldsymbol{\nabla}_\nu(g^{\mu\nu}\delta\Gamma^\sigma_{\mu\sigma}),$$

so that

$$g^{\mu\nu}\delta R_{\mu\nu} = \boldsymbol{\nabla}_\sigma W^\sigma, \quad \text{where} \quad W^\sigma = g^{\mu\nu}\delta\Gamma^\sigma_{\mu\nu} - g^{\mu\sigma}\delta\Gamma^\rho_{\mu\rho}.$$

Since

$$\Gamma^\nu_{\mu\nu} = \frac{\partial}{\partial x^\mu}\log(\sqrt{-g}),$$

we obtain

$$\begin{aligned}\boldsymbol{\nabla}_\mu W^\mu &= \frac{\partial W^\mu}{\partial x^\mu} + \Gamma^\mu_{\nu\mu}W^\nu \\ &= \frac{1}{\sqrt{-g}}\frac{\partial}{\partial x^\mu}(\sqrt{-g}\,W^\mu).\end{aligned}$$

Thus we have

$$g^{\mu\nu}\delta R_{\mu\nu} = \frac{1}{\sqrt{-g}}\frac{\partial}{\partial x^\mu}(\sqrt{-g}\,W^\mu). \tag{18.3}$$

To find $\delta(\sqrt{-g})$, we use

$$\frac{\partial g}{\partial g_{\mu\nu}} = G^{\mu\nu} = g\,g^{\mu\nu},$$

so that

$$\delta g = \frac{\partial g}{\partial g_{\mu\nu}}\delta g_{\mu\nu} = g\,g^{\mu\nu}\delta g_{\mu\nu} = -g\,g_{\mu\nu}\delta g^{\mu\nu}$$

and we obtain

$$\delta(\sqrt{-g}) = -\frac{1}{2}\sqrt{-g}\,g_{\mu\nu}\delta g^{\mu\nu}. \tag{18.4}$$

Substituting (18.3)–(18.4) into the formula for δS we obtain

$$\begin{aligned}\delta S &= \int_D \left(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R\right)u^{\mu\nu}\sqrt{-g}\,d^4x + \int_D \frac{\partial}{\partial x^\mu}(\sqrt{-g}\,W^\mu)d^4x \\ &= \int_D \left(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R\right)u^{\mu\nu}\sqrt{-g}\,d^4x.\end{aligned}$$

Here we used the Stokes theorem and the condition that $\delta\Gamma^\lambda_{\mu\nu} = 0$ on ∂D, which follows from our assumptions on the space $\mathscr{M}$ of Lorentzian metrics on M. $\square$

Remark. 'Tautologically' computing variation of the Hilbert-Einstein action we obtain the relation

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{1}{\sqrt{-g}}\left\{\frac{\partial(\sqrt{-g}\,R)}{\partial g^{\mu\nu}} - \frac{\partial}{\partial x^\lambda}\frac{\partial(\sqrt{-g}\,R)}{\partial \frac{\partial g^{\mu\nu}}{\partial x^\lambda}}\right\}.$$

Remark. If one fixes only the values of metric tensor $g_{\mu\nu}$ on ∂D then δS will contain the boundary term. It is possible to add to the Hilbert-Einstein functional S the so-called *Gibbons-Hawking-York boundary term* so that the δS is still given by Hilbert's formula. This boundary term is the integral over ∂D of trace of the second fundamental form over the volume form of the induced metric on ∂D.

Denote

$$S_{\text{gravity}} = -\frac{c^3}{16\pi G} S(g).$$

The total action of the gravitational field in the presence of a matter with the density function $\Lambda(x)$, depending only on $g_{\mu\nu}$ and its first derivatives, is given by

$$S = S_{\text{gravity}} + S_{\text{matter}},$$

where

$$S_{\text{matter}} = \frac{1}{c}\int \Lambda\sqrt{-g}\, d^4x.$$

Defining symmetric stress-energy tensor by

$$T_{\mu\nu} = \frac{2c}{\sqrt{-g}}\frac{\delta S_{\text{matter}}}{\delta g^{\mu\nu}} = \frac{2}{\sqrt{-g}}\left\{\frac{\partial(\sqrt{-g}\,\Lambda)}{\partial g^{\mu\nu}} - \frac{\partial}{\partial x^\lambda}\frac{\partial(\sqrt{-g}\,\Lambda)}{\frac{\partial g^{\mu\nu}}{\partial x^\lambda}}\right\}$$

from $\delta S = 0$ we obtain Einstein equations

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G}{c^4}T_{\mu\nu}.$$

When Λ depends only on $g_{\mu\nu}$, the formula for the stress-energy tensor simplifies

$$T_{\mu\nu} = 2\frac{\partial\Lambda}{\partial g^{\mu\nu}} - g_{\mu\nu}\Lambda.$$

Thus for the electromagnetic field

$$\Lambda = -\frac{1}{16\pi}F_{\alpha\beta}F^{\alpha\beta} = -\frac{1}{16\pi}F_{\alpha\beta}F_{\gamma\delta}g^{\alpha\gamma}g^{\beta\delta}$$

and we obtain

$$T_{\mu\nu} = \frac{1}{4\pi}\left(-F_{\mu\lambda}F_{\nu\sigma}g^{\lambda\sigma} + \frac{1}{4}g_{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}\right).$$

Up to the factor $1/4\pi$ this is formula (11.2) in Lecture 11. For a macroscopic body the energy-momentum tensor is

$$T_{\mu\nu} = (p+\varepsilon)u_\mu u_\nu - pg_{\mu\nu},$$

where p is the pressure and ε is the energy density of the body.

For a complete determination of the distribution and motion of the matter one must add to Einstein equations equation of the state of the matter, that is, equation relating the pressure density and temperature. This equation must be given along with the Einstein equations.

LECTURE 19

Einstein equations – II

19.1. Palatini formalism

In this approach to general relativity we consider the metric tensor $g_{\mu\nu}$ on the space-time M and affine torsion-free connection $\Gamma^\lambda_{\mu\nu}$ on TM as independent fields (due to the condition $\Gamma^\lambda_{\mu\nu} = \Gamma^\lambda_{\nu\mu}$ there are $50 = 10 + 40$ independent functions). Consider the action

$$S_{\mathrm{P}} = \int_M g^{\mu\nu} R_{\mu\nu} \sqrt{-g}\, d^4x,$$

where $R_{\mu\nu}$ is given by formula (17.7) in Lecture 21,

$$R_{\mu\nu} = \frac{\partial \Gamma^\lambda_{\mu\nu}}{\partial x^\lambda} - \frac{\partial \Gamma^\lambda_{\mu\lambda}}{\partial x^\nu} + \Gamma^\lambda_{\mu\nu}\Gamma^\sigma_{\lambda\sigma} - \Gamma^\sigma_{\mu\lambda}\Gamma^\lambda_{\sigma\nu}.$$

Its variation with respect to $\Gamma^\lambda_{\mu\nu}$ is still given by the Palatini identity

$$\delta R_{\mu\nu} = \boldsymbol{\nabla}_\lambda(\delta\Gamma^\lambda_{\mu\nu}) - \boldsymbol{\nabla}_\nu(\delta\Gamma^\lambda_{\mu\lambda}),$$

whereas variation of $\sqrt{-g}$ is given by formula (18.4), in Lecture 22,

$$\delta(\sqrt{-g}) = -\frac{1}{2}\sqrt{-g}\, g_{\mu\nu}\delta g^{\mu\nu}.$$

Indeed,

$$\begin{aligned}
\delta R_{\mu\nu} &= \frac{\partial \delta\Gamma^\lambda_{\mu\nu}}{\partial x^\lambda} - \frac{\partial \delta\Gamma^\lambda_{\mu\lambda}}{\partial x^\nu} + \delta\Gamma^\lambda_{\mu\nu}\Gamma^\sigma_{\lambda\sigma} + \Gamma^\lambda_{\mu\nu}\delta\Gamma^\sigma_{\lambda\sigma} - \delta\Gamma^\sigma_{\mu\lambda}\Gamma^\lambda_{\sigma\nu} - \Gamma^\sigma_{\lambda\mu}\delta\Gamma^\lambda_{\sigma\nu} \\
&= \frac{\partial \delta\Gamma^\lambda_{\mu\nu}}{\partial x^\lambda} + \Gamma^\sigma_{\lambda\sigma}\delta\Gamma^\lambda_{\mu\nu} - \Gamma^\sigma_{\lambda\mu}\delta\Gamma^\lambda_{\sigma\nu} - \Gamma^\lambda_{\sigma\nu}\delta\Gamma^\sigma_{\mu\lambda} - \frac{\partial \delta\Gamma^\lambda_{\mu\lambda}}{\partial x^\nu} + \Gamma^\lambda_{\mu\nu}\delta\Gamma^\sigma_{\lambda\sigma} \\
&= \boldsymbol{\nabla}_\lambda(\delta\Gamma^\lambda_{\mu\nu}) - \boldsymbol{\nabla}_\nu(\delta\Gamma^\lambda_{\mu\lambda}).
\end{aligned}$$

Denoting $R = g^{\mu\nu}R_{\mu\nu}$ and using Stokes' theorem we obtain

$$\begin{aligned}
\delta S_{\mathrm{P}} &= \int_M \left(R_{\mu\nu}\delta g^{\mu\nu} + g^{\mu\nu}\delta R_{\mu\nu} + R\frac{\delta(\sqrt{-g})}{\sqrt{-g}} \right)\sqrt{-g}\, d^4x \\
&= \int_M \left(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R \right)\delta g^{\mu\nu}\sqrt{-g}\, d^4x + \int_M g^{\mu\nu}\delta R_{\mu\nu}\sqrt{-g}\, d^4x \\
&= \int_M \left(\left(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R \right)\delta g^{\mu\nu} + Q^{\mu\nu}_\lambda \delta\Gamma^\lambda_{\mu\nu} \right)\sqrt{-g}\, d^4x,
\end{aligned}$$

where

$$Q_\lambda^{\mu\nu} = -\frac{1}{\sqrt{-g}}\frac{\partial(\sqrt{-g}g^{\mu\nu})}{\partial x^\lambda} + g^{\mu\nu}\Gamma^\sigma_{\lambda\sigma} - g^{\mu\sigma}\Gamma^\nu_{\lambda\sigma} - g^{\nu\sigma}\Gamma^\mu_{\lambda\sigma}$$
$$+\delta^\nu_\lambda\left(\frac{1}{\sqrt{-g}}\frac{\partial(\sqrt{-g}g^{\mu\sigma})}{\partial x^\sigma} + g^{\rho\sigma}\Gamma^\mu_{\rho\sigma}\right).$$

Thus equation $\delta S_{\rm P} = 0$ yileds

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 0 \quad \text{and} \quad Q_\lambda^{\mu\nu} = 0.$$

Using

$$\frac{\partial\sqrt{-g}}{\partial x^\lambda} = -\frac{1}{2}\sqrt{-g}\,g_{\mu\nu}\frac{\partial g^{\mu\nu}}{\partial x^\lambda}$$

and definition of the covariant derivative,

$$\boldsymbol{\nabla}_\lambda g^{\mu\nu} = \frac{\partial g^{\mu\nu}}{\partial x^\lambda} + \Gamma^\mu_{\lambda\sigma}g^{\sigma\nu} + \Gamma^\nu_{\lambda\sigma}g^{\mu\sigma},$$

we can rewrite equation $Q_\lambda^{\mu\nu} = 0$ as

$$(19.1) \qquad -\boldsymbol{\nabla}_\lambda g^{\mu\nu} + \frac{1}{2}g^{\mu\nu}g_{\sigma\rho}\boldsymbol{\nabla}_\lambda g^{\sigma\rho} + \delta^\nu_\lambda\left(\boldsymbol{\nabla}_\sigma g^{\mu\sigma} - \frac{1}{2}g^{\mu\alpha}g_{\sigma\rho}\boldsymbol{\nabla}_\alpha g^{\sigma\rho}\right) = 0.$$

Equation (19.1) has free indices λ, μ and ν. Putting $\lambda = \nu$ and summing over ν gives

$$-\boldsymbol{\nabla}_\nu g^{\mu\nu} + \frac{1}{2}g^{\mu\nu}g_{\sigma\rho}\boldsymbol{\nabla}_\nu g^{\sigma\rho} + 4\left(\boldsymbol{\nabla}_\sigma g^{\mu\sigma} - \frac{1}{2}g^{\mu\alpha}g_{\sigma\rho}\boldsymbol{\nabla}_\alpha g^{\sigma\rho}\right) = 0,$$

whence

$$\boldsymbol{\nabla}_\nu g^{\mu\nu} = \frac{1}{2}g^{\mu\nu}g_{\sigma\rho}\boldsymbol{\nabla}_\nu g^{\sigma\rho}.$$

Substituting this formula to (19.1) gives,

$$(19.2) \qquad \boldsymbol{\nabla}_\lambda g^{\mu\nu} = \frac{1}{2}g^{\mu\nu}g_{\sigma\rho}\boldsymbol{\nabla}_\lambda g^{\sigma\rho}.$$

Contracting (19.2) $g_{\mu\nu}$ using $g_{\mu\nu}g^{\mu\nu} = 4$ yields

$$g_{\sigma\rho}\boldsymbol{\nabla}_\lambda g^{\sigma\rho} = 0,$$

and putting it back to (19.2) we finally obtain

$$\boldsymbol{\nabla}_\lambda g^{\mu\nu} = 0.$$

This shows that $\boldsymbol{\nabla}$ is the Levi-Civita connection. Thus in the Palatini formalism equations (17.2) for the Christoffel's symbols appear from the principle of the least action.

19.2. The Schwarzschild solution

For the case of static spherically symmetric metric in the empty space we consider the following ansatz

$$ds^2 = g_{00}(r)c^2dt^2 - g_{11}(r)dr^2 - r^2(d\theta^2 + \sin^2\theta\, d\varphi^2),$$

where we are using spherical coordinates

$$x = r\cos\theta\cos\varphi, \quad y = y\cos\theta\sin\varphi, \quad z = r\cos\theta.$$

It describes the gravitational field outside a spherical mass, on the assumption that the electric charge of the mass and angular momentum of the mass are all zero. Computing $\Gamma^\lambda_{\mu\nu}$, where $x^0 = ct, x^1 = r, x^2 = \theta, x^3 = \varphi$, and solving $R_{\mu\nu} = 0$ we obtain

$$g_{00}(r) = 1 - \frac{a}{r}, \quad g_{11} = \frac{1}{1 - \frac{a}{r}},$$

where a is a constant. Thus

$$ds^2 = \left(1 - \frac{a}{r}\right)c^2dt^2 - \frac{dr^2}{1 - \frac{a}{r}} - r^2d\Omega^2,$$

where $d\Omega^2$ is the induced metric on $S^2 \subset \mathbb{R}^3$. In the limit $r \to \infty$ we should have

$$g_\mu = \eta_{\mu\nu} + \frac{1}{c^2}\overset{2}{g}_{\mu\nu} + O\left(\frac{1}{c^3}\right),$$

so

$$\overset{2}{g}_{00} = -\frac{ac^2}{r} = -\frac{2MG}{r},$$

where M is the mass of a body creating gravitational field. By definition, the quantity

$$a = \frac{2MG}{c^2}$$

is called *Schwarzschild radius* and is denoted by r_s[1].

Thus the Schwarzschild metric is

$$ds^2 = \left(1 - \frac{r_s}{r}\right)c^2dt^2 - \frac{dr^2}{1 - \frac{r_s}{r}} - r^2d\Omega^2$$

and it is applicable for $r > R$, the radius of the body. At $r = r_s$ we have *event horizon* and $r < r_s$ describes the *black hole*, where the time coordinate t becomes spacelike and the radial coordinate r becomes timelike. The singularity at $r = r_s$ is apparent and can be eliminated by the change of coordinates, called *Gullstrand-Painlevé coordinates.*

[1]For the Earth $r_s = 0.8.9$ mm, while for the Sun $r_s = 3$ km.

LECTURE 20

Kaluza-Klein theory

In the 1920s the only knows fundamental forces were electromagnetism and the force of gravity, and the only known elementary particles were electron and proton. Einstein's idea of the unified field theory was to obtain electromagnetism and general relativity from a single fundamental field. Toward this goal, T. Kaluza (1921) and O. Klein (1926) proposed to consider the five-dimensional space-time $\mathcal{M} = M \times S^1_r$, where the fifth dimension in the circle of small radius

$$r = \sqrt{\frac{\hbar G}{c^3}} \sim 10^{-35}\, m$$

— the Planck's length ℓ_{P}. The coordinates on $\mathcal{M}$ will be denoted by $\tilde{x}^a$, $a = 0,1,2,3,4$, where $\tilde{x}^4 = \theta$, so that using x^μ, $\mu = 0,1,2,3$, for coordinates on M we have $\tilde{x}^\mu = x^\mu$. Consider the following pseudo-Riemannian metric on $\mathcal{M}$ of signature $(+,-,-,-,-)$,

$$\tilde{g}_{ab} = \begin{pmatrix} g_{00} - A_0A_0 & g_{01} - A_0A_1 & g_{02} - A_0A_2 & g_{03} - A_0A_3 & A_0 \\ g_{10} - A_1A_0 & g_{11} - A_1A_1 & g_{12} - A_1A_2 & g_{13} - A_1A_3 & A_1 \\ g_{20} - A_2A_0 & g_{21} - A_2A_1 & g_{22} - A_2A_2 & g_{23} - A_2A_3 & A_2 \\ g_{30} - A_3A_0 & g_{31} - A_3A_1 & g_{32} - A_3A_2 & g_{33} - A_3A_3 & A_3 \\ A_0 & A_1 & A_2 & A_3 & -1 \end{pmatrix}$$

so that

$$d\tilde{s}^2 = \tilde{g}_{ab} d\tilde{x}^a \tilde{x}^b = g_{\mu\nu} dx^\mu dx^\nu - (A_\mu dx^\mu - d\theta)^2.$$

Also assume that the metric $g_{\mu\nu}dx^\mu dx^\nu$ and the 1-form $A_\mu dx^\mu$ on M do not depend on θ.

We have the following basic facts.

1) For $\tilde{g} = \det \tilde{g}_{ab}$ one has $\tilde{g} = -g$, where $g = \det g_{\mu\nu}$.

2) The inverse matrix $\tilde{g}^{ab}$ is given by

$$\begin{pmatrix} g^{00} & g^{01} & g^{02} & g^{03} & A^0 \\ g^{10} & g^{11} & g^{12} & g^{13} & A^1 \\ g^{20} & g^{21} & g^{22} & g^{23} & A^2 \\ g^{30} & g^{31} & g^{32} & g^{33} & A^3 \\ A^0 & A^1 & A^2 & A^3 & -1 + A_\mu A^\mu \end{pmatrix}$$

3) Under the change of coordinates $x \mapsto x' = F(x)$, $\theta \mapsto \theta + \lambda(x)$ we have $A_\mu \mapsto A'_\mu + \partial_\mu \lambda$, so that U(1)-gauge invariance is a relativity in the fifth dimension!

20.1. Geodesic equation on $\mathcal{M}$

From formulas for Christoffel's symbols we get for metric $\tilde{g}_{ab}$:

$$\begin{aligned}
\tilde{\Gamma}^{\mu}_{\alpha\beta} &= \Gamma^{\mu}_{\alpha\beta} + \frac{1}{2}g^{\mu\sigma}(A_\alpha F_{\sigma\beta} + A_\beta F_{\sigma\alpha}),\\
\tilde{\Gamma}^{\mu}_{\alpha 4} &= \frac{1}{2}g^{\mu\sigma}F_{\alpha\sigma},\\
\tilde{\Gamma}^{4}_{\alpha\beta} &= A_\mu\Gamma^{\mu}_{\alpha\beta} - \frac{1}{2}\left(A^\mu(A_\alpha F_{\beta\mu} + A_\beta F_{\alpha\mu}) - \frac{\partial A_\alpha}{\partial x^\beta} - \frac{\partial A_\beta}{\partial x^\alpha}\right),\\
\tilde{\Gamma}^{4}_{\alpha 4} &= \frac{1}{2}A^\mu F_{\alpha\mu},\\
\tilde{\Gamma}^{a}_{44} &= 0.
\end{aligned}$$

As usual, here

$$F_{\alpha\beta} = \frac{\partial A_\beta}{\partial x^\alpha} - \frac{\partial A_\alpha}{\partial x^\beta}.$$

For the free particle of mass m on the five-dimensional space-time $\mathcal{M}$ we have the action

$$S = -mc\int d\tilde{s} = -mc\int\sqrt{\tilde{g}_{ab}\frac{d\tilde{x}^a}{d\tilde{s}}\frac{d\tilde{x}^b}{d\tilde{s}}}\,d\tilde{s}.$$

Using the formulas for Christoffel's symbols $\tilde{\Gamma}^a_{bc}$ and putting $u^a = \dfrac{d\tilde{x}^a}{d\tilde{s}}$, we get the following equations

$$\frac{du^\mu}{d\tilde{s}} + \Gamma^{\mu}_{\alpha\beta}u^\alpha u^\beta = -g^{\mu\sigma}A_\alpha F_{\sigma\beta}u^\alpha u^\beta - g^{\mu\sigma}F_{\alpha\sigma}u^\alpha u^4, \quad \mu = 0,1,2,3,$$

and

$$\frac{du^4}{d\tilde{s}} + A_\mu\Gamma^{\mu}_{\alpha\beta}u^\alpha u^\beta = -A^\sigma F_{\alpha\sigma}u^\alpha u^4 + A^\sigma A_\alpha F_{\beta\sigma}u^\alpha u^\beta + \frac{\partial A_\alpha}{\partial x^\beta}u^\alpha u^\beta.$$

Multiplying first equations by A_μ and adding them to the second equation yields

$$\frac{du^4}{d\tilde{s}} - A_\mu\frac{du^\mu}{d\tilde{s}} - \frac{\partial A_\alpha}{\partial x^\beta}u^\alpha u^\beta = 0$$

so that

$$\frac{d}{d\tilde{s}}(u^4 - A_\mu u^\mu) = 0.$$

Thus $u^4 - A_\mu u^\mu = \xi$ is constant and the first equation takes the form

$$\frac{du^\mu}{d\tilde{s}} + \Gamma^{\mu}_{\alpha\beta}u^\alpha u^\beta = -\xi g^{\mu\nu}F_{\alpha\nu}u^\alpha.$$

Since $1 = g_{\mu\nu}u^\mu u^\nu + (u^4 - A_\mu u^\nu)^2$ we have $g_{\mu\nu}u^\mu u^\nu = 1 - \xi^2$, i.e.,

$$\frac{ds}{d\tilde{s}} = \sqrt{1-\xi^2}.$$

Whence

$$\frac{dx^\mu}{ds} = u^\mu \frac{d\tilde{s}}{ds} = \frac{u^\mu}{\sqrt{1-\xi^2}}$$

and we obtain

$$\frac{d^2x^\mu}{ds^2} + \Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{ds}\frac{dx^\beta}{ds} = -\frac{\xi}{\sqrt{1-\xi^2}}\, g^{\mu\sigma}F_{\alpha\sigma}\frac{dx^\alpha}{ds}.$$

Putting

$$\xi = \frac{e}{\sqrt{m^2c^4+e^2}}$$

we see that the right hand side becomes

$$\frac{e}{mc^2}g^{\mu\sigma}F_{\alpha\sigma}\frac{dx^\alpha}{ds}.$$

Thus we get the equation of a free charged particle moving in external gravitational and magnetic fields, obtained from the action

$$-mc\int ds - \frac{e}{c}\int A_\mu dx^\mu.$$

This is the so-called *first Kaluza miracle*.

20.2. Hilbert action on $\mathcal{M}$

By a direct and lengthy computation on gets

$$\tilde{R} = R + \frac{1}{4}F_{\mu\nu}F^{\mu\nu},$$

which is *Kaluza's second miracle*. The pure gravity action on $\mathcal{M}$ is proportional to the Hilbert action,

$$S_{\mathcal{M}} = -\frac{c^3}{16\pi\tilde{G}}\int_{\mathcal{M}}\tilde{R}\sqrt{\tilde{g}}\,d^5\tilde{x},$$

where $\tilde{G}$ is the gravitational constant $\mathcal{M}$. Putting $\tilde{G} = 2\pi rG$, replacing A_μ by κA_μ, where $\kappa = 2\sqrt{G}/c^2$, and trivially integrating over S^1_r we finally obtain

$$S_{\mathcal{M}} = -\frac{c^3}{16\pi G}\int_M\left(R + \frac{1}{16\pi c}F_{\mu\nu}F^{\mu\nu}\right)\sqrt{-g}\,d^4x.$$

This is the desired unification of general relativity and electromagnetism. It yields Einstein equations

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G}{c^4}T_{\mu\nu}$$

with the energy-momentum tensor of the electromagnetic field on M,

$$T_{\mu\nu} = \frac{1}{4\pi}\left(-F_{\mu\lambda}F_{\nu\sigma}g^{\lambda\sigma} + \frac{1}{4}g_{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}\right),$$

and Maxwell's equations

$$\boldsymbol{\nabla}_\nu F^{\mu\nu} = 0$$

on M in the presence of the gravitation field $g_{\mu\nu}$. Thus the Kaluze-Klein pure gravity action in the five-dimensional space $\mathcal{M}$ naturally produces Einstein-Hilbert-Maxwell action on the space-time M.

20.3. Criticism of the Kaluza-Klein theory

Though mathematically elegant, Kaluza-Klein theory gives unrealistic predictions for the masses of particles. Namely, consider the massless scalar field $\Phi(x,\theta)$ on $\mathcal{M}$ satisfying the five-dimensional wave equation

$$\left(\Box_4 - \frac{\partial^2}{\partial\theta^2}\right)\Phi = 0,$$

where $g_{\mu\nu}$ is the Minkowski metric. Corresponding Fourier coefficients

$$\Phi(x,\theta) = \sum_{n=-\infty}^{\infty} \varphi_n(x)e^{\frac{in\theta}{r}}$$

satisfy Klein-Gordon equations

$$(\Box_4 + m_n^2)\varphi_n = 0$$

with masses

$$m_n^2 = \frac{n^2}{r^2}.$$

However, these masses are very large! Thus assuming that $n = 1$ gives electron, the obtained mass would $m_e \sim 3\cdot 10^{30}\,\mathrm{MeV}$, while the actual electron mass is only $0.5\,\mathrm{MeV}$.

Geometrically one can consider general Kaluza-Klein metrics

$$\tilde{g}_{ab}(x,\theta) = \begin{pmatrix} g_{\mu\nu} - \Phi A_\mu A_\nu & \Phi A_\mu \\ \Phi A_\nu & -\Phi \end{pmatrix},$$

where $\Phi(x,\theta)$ is a function on $\mathcal{M}$, and consider the corresponding pure gravity Hilbert action. However, even assuming that the metric $\tilde{g}_{ab}$ does not depend on θ, setting $\Phi = 1$ in the field equations is not the same as setting first $\Phi = 1$ and consider the resulting field equations, which unify general relativity and electromagnetism. In other words, this unification is obtained considered a special subvariety of metrics on $\mathcal{M}$ which have $\Phi = 1$.