Classical Field Theory For Mathematicians

Alexander Kirillov, Jr.

Leon A. Takhtajan

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

Email address: kirillov@math.stonybrook.edu

Email address: leontak@math.stonybrook.edu

Contents

Preface xi

Part 1. Foundations of Classical Mechanics and Special Relativity

Chapter 1. Lagrangian Mechanics 3
§1.1. Generalized coordinates 3
§1.2. Principle of least action 4
§1.3. Euler–Lagrange equations 4
§1.4. Newtonian spacetime 8
§1.5. Lie algebra of the Galilean group 9
§1.6. Examples of Lagrangian systems 11
§1.7. Exercises 13

Chapter 2. Integrals of Motion and Noether's Theorem 15
§2.1. Conservation of energy 15
§2.2. Noether's theorem 16
§2.3. Examples of conservation laws 19
§2.4. Exercises 21

Chapter 3. Integration of Equations of Motion 23
§3.1. One-dimensional motion 23
§3.2. Harmonic oscillator and small oscillations 24
§3.3. Coupled harmonic oscillators 25
§3.4. Two-body problem 27
§3.5. Exercises 29

Chapter 4. Hamiltonian Formalism 31
§4.1. Legendre transform 31

§4.2. Hamilton's equations 32
§4.3. Examples of Hamiltonian systems 33
§4.4. Symplectic manifolds 35
§4.5. Lagrangian submanifolds and polarizations 37
§4.6. Classical observables and Poisson bracket 37
§4.7. Completely integrable systems 39
§4.8. Exercises 41

Chapter 5. Hamiltonian Action and Moment Map 43
§5.1. Hamiltonian actions 43
§5.2. Noether's theorem 46
§5.3. Galilean group action in $T^*\mathbb{R}^3$ 47
§5.4. Hamiltonian reduction 50
§5.5. Exercises 51

Chapter 6. Hamiltonian Systems with Constraints 53
§6.1. First order Lagrangian formalism 53
§6.2. Singular Lagrangians 55
§6.3. First class constraints and reduced phase space 57
§6.4. Exercises 62

Chapter 7. Special Relativity 63
§7.1. Minkowski spacetime and the Lorentz group 63
§7.2. Structure of the Lorentz group 65
§7.3. Lie algebra of the Lorentz and Poincare groups 67
§7.4. Lorentz group as the deformation of Galilean group 69
§7.5. Exercises 72

Chapter 8. Relativistic Particle 75
§8.1. World line of a particle 75
§8.2. The principle of the least action 76
§8.3. Noether integrals 79
§8.4. The Hamiltonian formulation 80
§8.5. Hamiltonian action of Poincaré group on $T^*\mathbb{R}^3$ 81
§8.6. No-interaction theorem 83
§8.7. Exercises 84

Chapter 9. Spinors and Dirac Operator 87
§9.1. Spin group and definition of spinors 87
§9.2. Low-dimensional examples 88
§9.3. Clifford algebra construction 89

§9.4. Lorentz group and Weyl spinors 92
§9.5. Dirac equation 94
§9.6. Exercises 95

Chapter 10. Notes and References 97

Part 2. Basics of Classical Field Theory

Chapter 11. Lagrangian Formulation of Field Theory 103
§11.1. From particles to fields 103
§11.2. Classical fields 104
§11.3. Calculus of variations 105
§11.4. Lagrangian density 107
§11.5. Euler–Lagrange equations 110
§11.6. Coordinate-free form of Euler-Lagrange equations 112
§11.7. Relativistic fields 115
§11.8. Examples of relativistic field theories 116
§11.9. Exercises 119

Chapter 12. Conservation Laws 121
§12.1. Symmetries and Noether's theorem 121
§12.2. Stress-energy tensor 125
§12.3. Examples 126
§12.4. Exercises 130

Chapter 13. Hamiltonian Formulation of Classical Field Theory 131
§13.1. The phase space and algebra of classical observables 131
§13.2. Symplectic form and Poisson bracket 133
§13.3. Legendre transform and Hamilton's equations 134
§13.4. Representation of the Poincaré group 135
§13.5. Free scalar field as a completely integrable Hamiltonian system 139
§13.6. Exercises 142

Chapter 14. Notes and References 145

Part 3. Classical Gauge Theories

Chapter 15. Maxwell's Equations 149
§15.1. Physics formulation 149
§15.2. Maxwell's equations in the language of differential forms 151
§15.3. Maxwell's equations with sources 155
§15.4. Vector potential and Lagrangian 156

§15.5. The stress-energy tensor 158
§15.6. Exercises 160

Chapter 16. Gauge Fixing and Hamiltonian Formalism in Electromagnetism 165
§16.1. Gauge fixing 165
§16.2. Plane waves 167
§16.3. The general solution 169
§16.4. Hamiltonian formalism in classical electrodynamics 171
§16.5. Coupling electromagnetiic field with matter 175
§16.6. Exercises 177

Chapter 17. Connections and Curvature 179
§17.1. Vector bundles and G-bundles 179
§17.2. Connections 182
§17.3. Connections in principal G-bundles 183
§17.4. One-form of a connection 184
§17.5. Gauge transformations 187
§17.6. Curvature of a connection 188
§17.7. Chern–Weil theory 190
§17.8. Exercises 192

Chapter 18. Yang–Mills Theory 195
§18.1. Yang–Mills theory 195
§18.2. Line bundles and Maxwell's equations 198
§18.3. Stress-energy tensor for Yang–Mills theory 198
§18.4. Coupling Yang-Mills field with matter 199
§18.5. Hamiltonian formalism in Yang–Mills theory 200
§18.6. Self-duality equations 205
§18.7. Hitchin's equations 207
§18.8. Exercises 208

Chapter 19. Chern–Simons Theory 209
§19.1. Topology of simple Lie groups 209
§19.2. Chern–Simons action 211
§19.3. Chern–Simons form as a secondary characteristic class 212
§19.4. Equations of motion in the Chern–Simons theory 214
§19.5. Chern–Simons theory for manifolds with boundary 215
§19.6. The phase space of Chern–Simons theory 218
§19.7. Exercises 221

Chapter 20. Notes and References 223

Part 4. Theory of Gravity

Chapter 21. General Relativity 229
§21.1. Spacetime in general relativity 229
§21.2. Particle in a gravitation field and Levi-Civita connection 233
§21.3. The Riemann curvature tensor 235
§21.4. Exercises 237

Chapter 22. Einstein Equations 239
§22.1. Einstein field equations 239
§22.2. Hilbert action 241
§22.3. Einstein equations with matter 243
§22.4. Asymptotically flat spacetime 246
§22.5. Exercises 247

Chapter 23. Hamiltonian Formulation and Exact Solutions 249
§23.1. Palatini first order formalism 249
§23.2. Hamiltonian formalism in general relativity 251
§23.3. The Schwarzschild solution 253
§23.4. Cosmological constant and AdS and dS spacetimes 254
§23.5. Kaluza-Klein theory 256
§23.6. Exercises 259

Chapter 24. Notes and References 261

Bibliography 263

Preface

This book is based on graduate courses taught by the authors over the last fourteen years in the mathematics department of Stony Brook University. The goal of these courses — the one year sequence Mathematical Physics I and II — was to introduce second year mathematics graduate students with no prior knowledge of physics to the basic concepts of classical mechanics and classical field theory, and then to quantum mechanics and (very introductory) quantum field theory.

Classical physics, especially classical mechanics and theory of electromagnetism, was an integral part of mathematical education up to the early twentieth century, with lecture courses given by Hilbert and Poincaré. To revive this tradition, the sequence Mathematical Physics I and II was created, with the goal to make classical and quantum physics accessible to graduate students and research mathematicians. The monograph [**Tak2008**] partially filled the gap by exposing mathematicians to the apparatus of quantum mechanics.

There are many excellent physics textbooks on theory of electromagnetism, special relativity and theory of gravity, starting from the classic text [**LL1971**], and every textbook on quantum field theory starts with succinct review of the first two topics. Development in the twentieth century, especially the advent of the theory of gauge fields — the Yang-Mills theory — revealed the geometric origin of classical field theories and put geometric methods at the forefront. This led to the great advances in fundamental mathematics, especially in topology and geometry of 4-manifolds, with several monograph dedicated to this subject.

However, a graduate student or a working mathematician can find few sources on classical field theory. The most comprehensive are lectures [**DF1999**] by P. Deligne and D. Freed at the special year 1996-1997 on Quantum Field Theory at the Institute for Advanced Study, Princeton.

The present book gives a comprehensive pedagogical treatment of classical field theory from a mathematics perspective. It is intended as an introduction rather than an in-depth monograph; yet we believe it touches all the important topics in classicl filed theory — with a notable expection of supersymmetry.

As the warm-up, the book starts with Part 1 devoted to classical mechanics and special relativity. This material is quite standard; our exposition is based on Hamiltonian approach and emphasizes the Hamiltonian action of the relevant Lie groups. Thus in Chapter 6 we show how the Galilean group acts naturally on the phase space of a free particle, and show that this action is Hamiltonian, while in Chapter 8 we exhibit the Hamiltonian action of the Poincaré group on the phase space of a free relativistic particle. In the same chapter we discuss the no interaction theorem in special relativity, which explains that relativistic invariance requires that physics phenomena is described by fields. These topics are not usually covered in mathematics textbooks. Also in Chapter 7 we discuss Hamiltonian systems with constraints. This formalism goes back to Dirac and is fundamental for the quantization of gauge theories. Surprisingly, a popular mathematical notion of Hamiltonian reduction is the special case of Dirac formalism.

Part 2 bridges classical mechanics and classical field theory. Here we, as in [**DF1999**] develop all necessary tools: Lagrangian formulation of the classical field theory, conservation laws and Noether theorem and Hamiltonian formulation. Compared with more abstract [**DF1999**], our approach can be considered as "neoclassical". We give all necessary facts about jet bundles, multivariable calculus of variations, etc; however, we mostly skip the analytic subtleties necessary to develop rigorous geometric methods for infinite-dimensional manifolds, leaving it to a more sophisticated and interested reader. Instead, we do things more at a formal level (like physicists). Continuing the line of group actions on phase spaces, in Chapter 13 we discuss the Hamiltonian action of the Poincaré group on the infinite-dimensional phase space of a free scalar field.

In Part 3 we discuss the gauge field theory: Maxwell's theory with the abelian structure group U(1), and the Yang-Mills theory with the structure group being semi-simple compact Lie group. For the convenience of the reader, in Chapter 17 we collect all necessary facts about connections and curvature in vector and principal bundles. Using Dirac formalism developed in Chapter 7, we carefully discuss Hamiltonian formulation of the Maxwell's and Yang-Mills theories. These theories play a fundamental role in the Standard Model of elementary particles, and we briefly discuss their couplings with matter fields. From mathematics perspective, we also discuss self-duality equations and Hitchin's equations, which play important role in fundamental mathematics. Finally, in Chapter 19 we introduce the Chern-Simons theory, the basic example of topological field theory, which plays a very important role in low-dimensional topology.

In Part 4 we briefly discuss theory of gravity, the Einsten's general relativity. Since there are plenty of physics and mathematics sources, our goal here is a coherent mathematical exposition of the basic notions. Thus in Chapter 21 we carefully discuss basic properties of the spacetime in general relativity and emphasize the notion of globally hyperbolic spacetime. In Chapter 22 we give a standard derivation of the Einstein's field equations with matter. In Chapter 23 we discuss the so-called Palatini formalism, an approach to Hilbert-Einstein action when 10 matrix elements of the metric tensor and 40 components of the symmetric Christoffel symbols are independent variables. We also briefly discuss Hamiltonian formalism for Einstein equations and their special solutions, with and without the cosmological constant.

Though we freely use all the necessary tools of modern mathematics, for convenience of the reader we motivate and remind them. Each chapter in the book concludes with a special *Notes and references* section, which provides references to the necessary mathematics background and physics sources. A courageous reader can actually learn the relevant mathematics by studying the main text and consulting these references, and with enough sophistication, could "translate" corresponding parts in physics textbooks into the mathematics language. For the physics students, the book presents an opportunity to become familiar with the geometric foundation of the classical field theory.

There are several ways to study the material in this book. A casual reader can study the main text in a cursory manner, and ignore numerous remarks and problems, located at the end of the sections. This would be sufficient to obtain basic minimal knowledge. A determined reader is supposed to fill in the details of the computations in the main text, which is the only way to master the material, and to attempt to solve the basic problems. Finally, a truly devoted reader should try to solve all the problems (probably consulting the corresponding references at the end of each part) and to follow up on the remarks, which may often be linked to other topics not covered in the main text.

Part 1

Foundations of Classical Mechanics and Special Relativity

Chapter 1

Lagrangian Mechanics

We assume that the reader is familiar with the basic notions from the theory of smooth (i.e., 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. For a smooth mapping of manifolds $f\colon M \to N$ we denote by $f_*\colon TM \to TN$ and $f^*\colon T^*N \to T^*M$, respectively, the induced mappings on tangent and cotangent bundles.

We will frequently use local coordinates: a choice of a coordinate chart (i.e. an open subset $U \subset M$ and an embedding $\varphi\colon U \to \mathbb{R}^n$) defines local coordinates $\boldsymbol{q} = (q^1, \ldots, q^n)$ in U. For a real-valued function f defined on U, we will use the shorthand notation

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

for the gradient of f.

For a vector field X on M, we will denote by ∂_X the corresponding differential operator; in particular, given a choice of local coordinate system q^i, we will denote $\partial_i = \dfrac{\partial}{\partial q^i}$.

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 the generalized coordinates q^i and the generalized velocities $\dot{q}^i$, so that classical trajectories satisfy a system of second order ordinary differential equations, called *equations of motion*.

1.2. Principle of least action

Classical mechanical systems are most commonly described using *Lagrangian formalism*. In this approach, a mechanical system with a configuration space M is determined by a smooth, real-valued function L on $TM \times \mathbb{R}$, called the *Lagrangian function* (or simply, *Lagrangian*).

The motion of a Lagrangian system (M, L) is described by the *principle of least action* formulated as follows.

Definition 1.1. For a path $\gamma\colon [t_0, t_1] \to M$, its action $S(\gamma) \in \mathbb{R}$ is defined by

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

Principle of Least Action. Classical trajectories of a Lagrangian system are critical points of the action functional on the space of smooth paths with fixed endpoints

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

Note that the path space is infinite-dimensional. One can rigorously define the notion of a critical point on such space using the language of Fréchet manifolds and Gateaux derivatives. However, it is also possible to define the notion of a critical point in a simpler way: a path $\gamma \in P(M) = P(M)_{q_0,t_0}^{q_1,t_1}$ is a critical point of the action functional, if for any one-parameter family of paths $\gamma_\varepsilon \in P(M)$ with $\gamma_0 = \gamma$, one has

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} S(\gamma_\varepsilon) = 0. \tag{1.1}$$

The critical points of the action functional are called *extremals*; thus, the principle of least action states that a Lagrangian system (M, L) moves along the extremals.

Remark 1.2. 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.

1.3. Euler–Lagrange equations

The extremality condition (1.1) can be rewritten as a system of differential equations, called equations of motion, or the *Euler–Lagrange equations*. In this section, we will write these equations in local coordinates on TM.

Specifically, the equations of motion have the most elegant form for the following choice of local coordinates on TM.

Definition 1.3. 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 $T_q M$, are called *standard coordinates.*

For a path $\gamma(t)$ in M, let $\gamma'(t)$ be the canonical lifting of γ to TM:

$$\gamma'(t) = (\gamma(t), \dot{\gamma}(t)) \in TM, \tag{1.2}$$

where the dot stands for the time derivative. We will call $\gamma'(t)$ *tangential lift* of γ. In this case, it is immediate that if γ in local coordinates on U is described by $\boldsymbol{q}(t) = (q^1(t), \dots, q^n(t))$, then $\gamma'(t)$ in standard coordinates is given by

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

Motivated by this and following a centuries long tradition[1], 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[2].

Any change of local coordinates in M induces a change of standard local coordinates in TM. We leave it to the reader to check that under such changes, for any function f on TM, components of the vector

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

transform in the same way as components of a one-form on M; in particular, this implies that

$$\theta_f = \sum_{i=1}^{n} \frac{\partial f}{\partial \dot{q}^i} dq^i \tag{1.3}$$

is a well-defined 1-form on TM which does not depend on the choice of local coordinates.

Theorem 1.4. *The equations of motion of a Lagrangian system (M, L) in standard coordinates on TM are given by the Euler-Lagrange equations*

$$\frac{\partial L}{\partial \boldsymbol{q}}(\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t), t) - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t), t) \right) = 0. \tag{1.4}$$

---

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

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

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 parts, gives

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

where, as usual in the calculus of variations, we denote

$$\delta q^i(t) = \left.\frac{\partial}{\partial\varepsilon}\right|_{\varepsilon=0} q^i(t,\varepsilon).$$

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

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

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 $\boldsymbol{v}(t) = \sum_{i=1}^{n} v^i(t)\frac{\partial}{\partial q^i}$ is a vector field along the path γ in M with $\boldsymbol{v}(t_0) = \boldsymbol{v}(t_1) = 0$, representing a tangent vector $V \in T_\gamma P(M)$. 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 1.6. It is also convenient to consider a space $\widehat{P(M)} = \{\gamma\colon [t_0, t_1] \to M\}$ 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 1.4 yields the following formula for the *first variation of the action with free ends*:

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

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

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

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 1.7. 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 the Lagrangian system is called *singular*.

It is easy to see 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

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

Remark 1.8. Let (L, M) be a time independent non-degenerate Lagrangian system and let $\gamma(t; q_0, v_0)$ be the solution of differential equation (1.7) with the initial conditions

$$\gamma(t_0) = q_0 \in M \quad \text{and} \quad \gamma'(t_0) = v_0 \in T_{q_0}M.$$

It follows from basic theorems of the theory of ordinary differential equations that there exist some $t_1 > 0$, a neighborhood V_0 of v_0 in $T_{q_0}M$, and neighborhoods U_t of $\gamma(t; q_0, v_0)$ in M for times $t_0 < t < t_1$, such that the mapping

$$V_0 \ni v \mapsto q(t) = \gamma(t; q_0, v) \in U_t \tag{1.8}$$

is a diffeomorphism. It is said that the extremal $\gamma_0(t) = \gamma(t; q_0, v_0)$ is included into the *central field of extremals* $\gamma(t) = \gamma(t, q_0, v)$ for times $t_0 < t < t_1$, where $v \in V_0$. In standard coordinates the mapping (1.8) is given by $\dot{\boldsymbol{q}} \to \boldsymbol{q}(t) = \gamma(t; \boldsymbol{q}_0, \dot{\boldsymbol{q}})$. We define the action as a function of coordinates (also called *the 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 connecting $\boldsymbol{q}_0$ and $\boldsymbol{q} \in \bigcup_{t_0<\tau\le t} U_\tau$.

1.4. Newtonian spacetime

The Newtonian space E^3 is a 3-dimensional Riemannian manifold which is isometric to the standard Euclidean space $\mathbb{R}^3$ with the usual metric. Note, however, that we do not fix a choice of an isometry between E^3 and $\mathbb{R}^3$; such an isometry is called a *frame* in E^3. Since it is known that any isometry of $\mathbb{R}^3$ with itself is an affine transformation, it follows that any two choices of a frame in E^3 are related by an affine transformation

$$\boldsymbol{r} \mapsto g \cdot \boldsymbol{r} + \boldsymbol{r}_0, \qquad g \in \mathrm{O}(3), \quad \boldsymbol{r}_0 \in \mathbb{R}^3.$$

The Newtonian *spacetime* is defined as 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 in the Newtonian spacetime $E^3 \times \mathbb{R}$ is an isomorphism $E^3 \times \mathbb{R} \to \mathbb{R}^3 \times \mathbb{R}$ which has the following properties.

- It preserves the notion of simultaneous events; for such events, it also preserves the distance between events.
- It preserves time intervals and time direction.

An obvious choice of an inertial frame is choosing a frame in E^3, i.e. an identification $\varphi\colon E^3 \to \mathbb{R}^3$, and extending it trivially to $E^3 \times \mathbb{R}$. However, this is not the most general choice. It is easy to show that composing any inertial frame with a transformation of $\mathbb{R}^3 \times \mathbb{R}$ given below gives again an inertial frame:

$$\begin{pmatrix} \boldsymbol{r} \\ t \end{pmatrix} \mapsto \begin{pmatrix} g \cdot \boldsymbol{r} + \boldsymbol{v}t + \boldsymbol{r}_0 \\ t + t_0 \end{pmatrix} \tag{1.9}$$

for some $g \in \mathrm{O}(3)$, $\boldsymbol{r}_0 \in \mathbb{R}^3$, $\boldsymbol{v} \in \mathbb{R}^3$, $t_0 \in \mathbb{R}$. Conversely, any two inertial frames can be obtained from each other by applying a transformation of this form.

The group G of all transformations of $\mathbb{R}^3 \times \mathbb{R}$ of the form (1.9) is called the *Galilean group*. This group is a semidirect product $G = G_0 \ltimes \mathbb{R}^4$, where G_0 is the homogeneous Galilean group, consisting of transformations

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

In particular, transformations with $g = e$, the identity in $\mathrm{O}(3)$, describe a *uniform motion* of spacetime. Again, it is easy to see that G_0 is just the Euclidean group $\mathrm{E}(3)$ — a semidirect product $\mathrm{E}(3) = \mathrm{O}(3) \ltimes \mathbb{R}^3$; it can be explicitly described as a subgroup in $\mathrm{GL}(4, \mathbb{R})$:

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

acting on the spacetime $\mathbb{R}^3 \times \mathbb{R}$ by

$$\begin{pmatrix} \boldsymbol{r} \\ t \end{pmatrix} \mapsto \begin{pmatrix} g & \boldsymbol{v} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{r} \\ t \end{pmatrix} = \begin{pmatrix} g \cdot \boldsymbol{r} + \boldsymbol{v}t \\ t \end{pmatrix}. \tag{1.10}$$

Thus, the final description of the full Galilean group is

$$G = \mathrm{E}(3) \ltimes \mathbb{R}^4. \tag{1.11}$$

This is a 10-dimensional Lie group. Its action on the spacetime can be described by the following subgroup in $\mathrm{GL}(5, \mathbb{R})$:

$$\begin{pmatrix} \boldsymbol{r} \\ t \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} g & \boldsymbol{v} & \boldsymbol{r}_0 \\ 0 & 1 & t_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{r} \\ t \\ 1 \end{pmatrix} = \begin{pmatrix} g \cdot \boldsymbol{r} + \boldsymbol{v}t + \boldsymbol{r}_0 \\ t + t_0 \\ 1 \end{pmatrix}. \tag{1.12}$$

One of the fundamental principles of Newtonian mechanics is the relativity principle, formulated by Galileo.

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

In particular, it shows that in Newtonian mechanics, the space is *homogeneous* (laws of nature are invariant under translations) and *isotropic* (invariance under rotations) and the time is *homogeneous*. Note, however, that the time in Newtonian mechanics is absolute (up to translations).

The Galilean invariance imposes restrictions on Lagrangians of mechanical systems. In particular, the Lagrangian of a *closed system*[3] in Newtonian mechanics does not explicitly depend on time.

1.5. Lie algebra of the Galilean group

For future use, it will also be convenient to have an explicit description of the Lie algebra of the Galilean group. It is easy to see that it is given by

$$\mathfrak{g} = (\mathfrak{so}(3) \ltimes \mathbb{R}^3) \ltimes \mathbb{R}^4 \tag{1.13}$$

where

$$\mathfrak{so}(3) = \{a \in \mathfrak{gl}(3) \mid a^t + a = 0\} \tag{1.14}$$

is the Lie algebra of skew-symmetric matrices. This Lie algebra has a standard basis

$$J_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad J_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \qquad J_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{1.15}$$

corresponding to one-parameter subgroups of rotations around x, y, and z axes. More concisely, they can be defined by the following formula, where as usual we assume summation over repeating indices:

$$J_i e_j = \epsilon_{ijk} e_k \tag{1.16}$$

[3] A system is called closed if its motion is not influenced by the outside material bodies.

where e_i is the standard basis in $\mathbb{R}^3$ and ϵ_{ijk} is the fully antisymmetric tensor:

(1.17)
$$\epsilon_{ijk} = \begin{cases} \operatorname{sgn}(\sigma) & \text{if } (i,j,k) \text{ is obtained from } (1,2,3) \text{ by applying a permutation } \sigma, \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to check that these generators satisfy the commutation relations

$$[J_1, J_2] = J_3, \qquad [J_3, J_1] = J_2, \qquad [J_2, J_3] = J_1,$$

which is commonly written as

$$[J_i, J_j] = \epsilon_{ijk} J_k. \tag{1.18}$$

It follows from (1.18) that the Lie algebra $\mathfrak{so}(3)$ is isomorphic to the vector space $\mathbb{R}^3$, considered as a Lie algebra with respect to the cross-product (see Exercise 1.6 for details); the latter also gives the action of $\mathfrak{so}(3)$ on $\mathbb{R}^3$. This fact will be very useful later, when we discuss angular momenta (see Example 2.2). However, the isomorphism $\mathfrak{so}(3) \simeq \mathbb{R}^3$ is a pure coincidence, a special property of the three-dimensional space. In other dimensions, there is no isomorphism between $\mathfrak{so}(n)$ and $\mathbb{R}^n$, and $\mathbb{R}^n$ does not have a natural Lie algebra structure (other than the zero one). The proper generalization of commutation relations (1.18) to $\mathfrak{so}(n)$ will be given in Exercise 2.6.

We can also give an explicit basis in the Lie algebra $\mathfrak{se}(3)$ of the homogeneous Galilean group $E(3) = \mathrm{O}(3) \ltimes \mathbb{R}^3$, by combining the basis $J_i \in \mathfrak{so}(3)$ defined above with the standard basis in $\mathbb{R}^3$, which we denote by K_i, $i = 1, 2, 3$. We leave it as an exercise to the reader to check, using (1.10), that the commutation relations of $\mathfrak{se}(3)$ in this basis are given by

$$\begin{aligned} [J_i, J_j] &= \epsilon_{ijk} J_k, \\ [J_i, K_j] &= \epsilon_{ijk} K_k, \\ [K_i, K_j] &= 0. \end{aligned} \tag{1.19}$$

The Lie algebra $\mathfrak{g}$ of the Galilean group is obtained by adding to the basis J_i, K_i elements P_i and P_0, corresponding to the standard basis in $\mathbb{R}^3 \times \mathbb{R}$. As it follows from (1.11), they satisfy commutation relations

$$\begin{aligned} [P_i, P_j] &= 0, \quad [P_i, P_0] = 0, \\ [J_i, P_0] &= 0, \quad [K_i, P_0] = P_i, \\ [K_i, P_j] &= 0. \end{aligned} \tag{1.20}$$

Commutation relations (1.19)–(1.20) describe the Lie algebra $\mathfrak{g}$ of the Galilean group.

Remark 1.9. The Lie algebra $\mathfrak{g}$ of the Galilean group is not semisimple and admits a one-dimensional central extension

$$\tilde{\mathfrak{g}} = \mathfrak{g} \oplus \mathbb{R}z.$$

Here z is the central element and the commutation relations in $\tilde{\mathfrak{g}}$ are given by (1.19)–(1.20) except relations $[K_i, P_j] = 0$, which are replaced by

$$[K_i, P_j] = \delta_{ij} z \tag{1.21}$$

(in fact, it can be shown that this is the only non-trivial central extension of $\mathfrak{g}$, see Section 7.4).

We leave it to the reader to show that this central extension can be lifted to define a a central extension $\tilde{G}$ of the Galilean group G:

$$0 \to \mathbb{R} \to \tilde{G} \to G \to 0.$$

1.6. Examples of Lagrangian systems

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

Example 1.1 (Free particle in Euclidean space)**.** The configuration space for a free particle is $M = \mathbb{R}^3$, and the Lagrangian is given by

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

Here $m > 0$ is the mass of a particle[4] and $\dot{\boldsymbol{r}}^2 = |\dot{\boldsymbol{r}}|^2$ is the length square of the velocity vector $\dot{\boldsymbol{r}} \in T_{\boldsymbol{r}}\mathbb{R}^3 \simeq \mathbb{R}^3$.

This system is invariant under Galilean transformations. Indeed, under the Galilean transformation $\boldsymbol{r} \mapsto \boldsymbol{r} + \boldsymbol{v}t$ we have

$$L = \tfrac{1}{2}m\dot{\boldsymbol{r}}^2 \mapsto L' = \tfrac{1}{2}m(\dot{\boldsymbol{r}} + \boldsymbol{v})^2 = L + \frac{d}{dt}(m\boldsymbol{r}\boldsymbol{v} + \tfrac{1}{2}\boldsymbol{v}^2 t), \tag{1.22}$$

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

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

More generally, it is not hard to show that $\ddot{\boldsymbol{r}} = 0$ is the only possible equation of motion in $\mathbb{R}^3$ which is invariant under the Galilean group.

Example 1.2 (Interacting particles)**.** A 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 the position of the a-th particle, $a = 1, \dots, N$. 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 the *kinetic energy* of a system and $V(\boldsymbol{r})$ is the *potential energy.* The Euler-Lagrange equations give *Newton's equations*

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

where

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

[4] Condition $m > 0$ is necessary for the action functional to be bounded from below.

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*: change in positions of particles produces instantanous change in the forces exerted by these particles on other parts of the system.

It follows from the homogeneity of the space that potential energy $V(\boldsymbol{r})$ of a 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 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 pairwise 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 the space that $V(\boldsymbol{r})$ depends only on relative distances between the particles, so that the Lagrangian of a system of N particles with pairwise 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

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

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 (Particle in an external potential field)**.** Here $M = \mathbb{R}^3$ and

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

where potential energy can explicitly depend on time. Equations of motion are Newton's equations

$$m \ddot{\boldsymbol{r}} = \boldsymbol{F} = -\frac{\partial V}{\partial \boldsymbol{r}}.$$

If $V = V(|\boldsymbol{r}|)$ is a function only of the distance $|\boldsymbol{r}|$, the potential field is called *central.*

Example 1.5 (Harmonic oscillator)**.** Consider a particle of mass m on a line $M = \mathbb{R}$ in a potential field $V(x) = \frac{1}{2} k x^2$, i.e. with the Lagrangian

$$L(x) = \tfrac{1}{2} m \dot{x}^2 - \tfrac{1}{2} k x^2. \tag{1.23}$$

In this case, the Euler–Lagrange equation of motion takes the form

$$m \ddot{x} = -kx.$$

Denoting $\omega = \sqrt{k/m}$, we can rewrite this equation in the familiar form

$$\ddot{x} + \omega^2 x = 0.$$

Solutions of this equation are

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

and describe periodic motion with frequency ω and with period $T = 2\pi/\omega$.

This system is called the *harmonic oscillator* and is probably the simplest mechanical system after the free particle.

We will consider generalizations of this system in the next chapter.

Example 1.6 (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

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

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

$$S(\gamma) = \tfrac{1}{2}\int_{t_0}^{t_1} \|\dot{\gamma}(t)\|^2 dt = \tfrac{1}{2}\int_{t_0}^{t_1} g_{\mu\nu}(x)\dot{x}^\mu \dot{x}^\nu dt$$

is called the action functional in Riemannian geometry. In this case, it can be shown that the variation of the action is given by

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

where ∇ is the Levi-Civita connection — the unique torsion-free metric connection in the tangent bundle TM — and ∇_ξ is the covariant derivative with respect to the vector field $\xi \in \mathrm{Vect}(M)$. Extremals of this action are the geodesics on M, traversed with constant speed: $\|\dot{\gamma}\| = \mathrm{const}$.

1.7. Exercises

Exercise 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 $TM \times \mathbb{R}$,

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

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

Exercise 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, $L'(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) = L(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) + \dfrac{d}{dt} f(\boldsymbol{q}, t)$ have the same equations of motion. Thus the Lagrangian is defined up to an addition of a total time derivative of a function of coordinates and time.

Exercise 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.

Exercise 1.4. Show that every extremal can be included into the central field, introduced in Remark 1.8.

Exercise 1.5 (Action as a function of coordinates)**.** Let $S(\boldsymbol{q}, t; \boldsymbol{q}_0, t_0)$ be the classical action, introduced in Remark 1.8. Using formula (1.6) show that

$$\frac{\partial S}{\partial \boldsymbol{q}}(\boldsymbol{q}, t) = \frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\boldsymbol{q}, \dot{\boldsymbol{q}}) \quad \text{and} \quad \frac{\partial S}{\partial t} = L(\boldsymbol{q}, \dot{\boldsymbol{q}}) - \dot{\boldsymbol{q}} \frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\boldsymbol{q}, \dot{\boldsymbol{q}}).$$

Exercise 1.6. Consider the following vector space isomorphism

$$\begin{aligned} \mathbb{R}^3 &\to \mathfrak{so}(3) \\ \boldsymbol{r} &\mapsto \hat{\boldsymbol{r}} = r_1 J_1 + r_2 J_2 + r_3 J_3 \end{aligned}$$

where $J_i \in \mathfrak{so}(3)$ are defined by (1.15).

Prove that then

(1) $\hat{\boldsymbol{u}} \cdot \boldsymbol{v} = \boldsymbol{u} \times \boldsymbol{v}$

(2) $[\hat{\boldsymbol{u}}, \hat{\boldsymbol{v}}] = \widehat{\boldsymbol{u} \times \boldsymbol{v}}$

where $\boldsymbol{u} \times \boldsymbol{v}$ is the usual cross-product in $\mathbb{R}^3$, given by $e_i \times e_j = \epsilon_{ijk} e_k$, or equivalently,

$$\boldsymbol{u} \times \boldsymbol{v} = \begin{pmatrix} u_2 v_3 - v_2 u_3 \\ u_3 v_1 - v_3 u_1 \\ u_1 v_2 - v_1 u_2 \end{pmatrix}.$$

Exercise 1.7. Consider the Lie algebra $\mathfrak{u}(2)$ of skew-hermitian matrices. This algebra has a standard basis $i\sigma_0, i\sigma_1, i\sigma_2, i\sigma_3$, where σ_k are so-called *Pauli matrices*:

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.24}$$

In particular, matrices $i\sigma_1$, $i\sigma_2$, $i\sigma_3$ form a basis in the algebra $\mathfrak{su}(2)$ of traceless skew-symmetric matrices.

Define a linear map $\mathfrak{so}(3) \to \mathfrak{su}(2)$ by

$$J_k \mapsto \frac{1}{2i} \sigma_k.$$

Show that this map is an isomorphism of Lie algebras. (It can be shown that this isomorphism lifts to an isomorphism of Lie groups $\mathrm{SO}(3) \simeq \mathrm{SU}(2)/\mathbb{Z}_2$.)

As an immediate corollary, we see that the same formulas also give an isomorphism of complex Lie algebras

$$\mathfrak{so}(3, \mathbb{C}) \cong \mathfrak{sl}(2, \mathbb{C}).$$

Chapter 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. However, in many cases this problem can be simplified by using conservation laws.

Definition 2.1. A smooth function $I\colon TM\times\mathbb{R}\to\mathbb{R}$ is called an *integral of motion*, or *first integral*, or *conservation law* for a Lagrangian system (M,L), if

$$\frac{d}{dt}I(\gamma(t),\dot{\gamma}(t),t)=0$$

for all extremals γ of the action functional.

In this chapter we discuss how one can construct such integrals of motion.

2.1. Conservation of energy

The best known example of a conserved quantity of a mechanical system is the *energy*.

Definition 2.2. The *energy* of a Lagrangian system (M,L) is the function E on $TM\times\mathbb{R}$, defined in standard coordinates on TM by

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

It follows from the fact that components of $\dfrac{\partial L}{\partial\dot{q}^i}$ transform in the same way as components of a one-form on M (see Section 1.3) that the expression $\dot{\boldsymbol{q}}\dfrac{\partial L}{\partial\dot{\boldsymbol{q}}}$ is independent of the choice of local coordinates, and thus the energy is well-defined on $TM\times\mathbb{R}$.

Proposition 2.3 (Conservation of energy)**.** *The energy of a Lagrangian system in which the Lagrangian has no explicit dependence on time is an integral of motion.*

Proof. For an extremal γ put $E(t) = E(\gamma(t), \dot{\gamma}(t))$. We have, according to the Euler-Lagrange equations,

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

Thus $\dfrac{\partial L}{\partial t} = 0$ implies that the energy is conserved. □

This proof is not very illuminating; to understand the true meaning of the energy conservation law, we will need to understand the relation between integrals of motion and the symmetries of a system, which will be done in the next section.

2.2. Noether's theorem

In Section 1.1, we discussed that one of the postulates of Newtonian mechanics is invariance of the equations of motion under the group of Galilean transformations of the spacetime. This a common phenomenon: in many cases, mechanical systems we consider have some symmetries. The definition below makes it precise. For simplicity, we begin by considering only Lagrangians which have no explicit dependence on time; more general case will be considered later (see Remark 2.8).

Definition 2.4. A Lagrangian $L\colon TM \to \mathbb{R}$ is invariant with respect to a diffeomorphism $g\colon M \to M$ if $L(g_*(v)) = L(v)$ for all $v \in TM$. In this case, the diffeomorphism g is called a *symmetry* of the Lagrangian system (M, L).

A Lie group G is a *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 $M \to M\colon x \mapsto g \cdot x$ is a symmetry.

In particular, it is immediate from the definition that if $g \in \mathrm{Diff}(M)$ is a symmetry of the system, then the set of extremals is invariant under g. Note that converse is false: a transformation can preserve the extremals and yet not preserve the Lagrangian. For example, Galilean transformations of the spacetime do not preserve the Lagrangian of the free particle, see Example 1.1.

The fundamental theorem of Emmy Noether establishes a relation between continuous symmetries and conservation laws.

Theorem 2.5 (Noether)**.** *Suppose that Lagrangian $L\colon TM \to \mathbb{R}$ 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 — the Noether integral — given in standard coordinates on TM by*

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

where $\xi = \sum_{i=1}^{n} \xi^i(\boldsymbol{q}) \frac{\partial}{\partial q^i}$ *is the vector field on* M *associated with the flow* g_s*:*

$$\xi = \left.\frac{dg_s(\boldsymbol{q})}{ds}\right|_{s=0}.$$

Proof. In Section 1.2 we have shown (see Remark 1.5) that for any one-parameter family of trajectories γ_ε (not necessarily with fixed endpoints), we have

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} S(\gamma_\varepsilon) = \int_{t_0}^{t_1} \left(\frac{\partial L}{\partial \boldsymbol{q}} - \frac{d}{dt}\frac{\partial L}{\partial \dot{\boldsymbol{q}}}\right) \xi \, dt + \left.\frac{\partial L}{\partial \dot{\boldsymbol{q}}} \xi \right|_{t_0}^{t_1},$$

where $\xi = \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} \gamma_\varepsilon$ (for simplicity, we drop the indices).

Let us apply it to the one-parameter family $\gamma_\varepsilon = g_\varepsilon(\gamma)$. Since g_ε is a symmetry of the Lagrangian, $S(\gamma_\varepsilon)$ is constant and thus the left-hand side of the previous formula is zero. If, in addition, γ is an extremal, then by Euler–Lagrange equations the integrand is identically zero, so we get

$$\frac{\partial L}{\partial \dot{\boldsymbol{q}}} \xi(t_0) = \frac{\partial L}{\partial \dot{\boldsymbol{q}}} \xi(t_1),$$

which is another way of writing that $\frac{\partial L}{\partial \dot{\boldsymbol{q}}} \xi$ is a first integral. □

Corollary 2.6. *Let* G *be a Lie group of symmetries of a system* (M, L)*, and let* $\mathfrak{g}$ *be the Lie algebra of* G*. Then every element* $a \in \mathfrak{g}$ *defines a one-parameter group of symmetries* $g_s = \exp(sa)$*,* $s \in \mathbb{R}$*, and thus an integral of motion* I_a *of the system.*

We will later show that in fact one can define a Lie algebra structure on the space of integrals of motion, so that $a \mapsto I_a$ is a morphism of Lie algebras.

Remark 2.7. Theorem 2.5 has an infinitesimal analog, in which diffeomorphisms are replaced by vector fields. Namely, a vector field X on M is called an *infinitesimal symmetry*, if the corresponding "time s" local flow g_s of X 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.

It is easy to verify that X is an infinitesimal symmetry if and only if $dL(X') = 0$ on TM. We leave it to the reader to modify the proof of Theorem 2.5 to show that in this case,

$$I(\boldsymbol{q}, \dot{\boldsymbol{q}}) = \frac{\partial L}{\partial \dot{\boldsymbol{q}}} X = \theta_L(X'),$$

where $\theta_L \in \Omega^1(TM)$ is defined by (1.3), is an integral of motion.

Remark 2.8. Noether's theorem generalizes to time-dependent Lagrangians $L\colon TM \times \mathbb{R} \to \mathbb{R}$. 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 $I = -E$. This explains the energy conservation law discussed in the previous section: energy conservation is a corollary of time-independence; energy is the first integral corresponding to the time translations.

Noether's theorem can be generalized further: instead of requiring $dL(X')$ to be zero, we can require it to be a complete derivative.

Proposition 2.9. *Suppose that for a given Lagrangian $L\colon TM \to \mathbb{R}$ there exist a vector field $X = \sum_{i=1}^n X^i \dfrac{\partial}{\partial q^i}$ on M and a function K on TM such that for every path γ in M,*

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

where, as before, γ' is the tangential lift of γ. Then

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

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

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

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

For a closed, non-degenerate Lagrangian system (M, L) this result can be generalized further by allowing coefficients $X^i(\boldsymbol{q})$ 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

$$\tilde{X} = \sum_{i=1}^n X^i(\boldsymbol{q}, \dot{\boldsymbol{q}}) \frac{\partial}{\partial q^i} + \sum_{i,j=1}^n \left(\dot{q}^j \frac{\partial X^i}{\partial q^j}(\boldsymbol{q}, \dot{\boldsymbol{q}}) + F^j(\boldsymbol{q}, \dot{\boldsymbol{q}}) \frac{\partial X^i}{\partial \dot{q}^j}(\boldsymbol{q}, \dot{\boldsymbol{q}})\right) \frac{\partial}{\partial \dot{q}^i}. \tag{2.1}$$

Proposition 2.10. *Suppose that for a closed, non-degenerate Lagrangian L there exist a vector field $\tilde{X}$ on TM of the form (2.1), and a function K on TM such that for every path γ in M,*

$$dL(\tilde{X})(\gamma'(t)) = \frac{d}{dt} K(\gamma'(t)). \tag{2.2}$$

Then

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

is an integral of motion.

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

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

2.3. Examples of conservation laws

Example 2.1 (Conservation of momentum)**.** Let M be an affine space associated with the vector space V, and suppose that a Lagrangian L is invariant under translations $g_s(q) = q + sv$ for some vector $v \in V$. According to Noether's theorem,

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

is an integral of motion. In particular, if L is invariant under translations by any vector v, this implies that each of the quantities

$$P_i = \frac{\partial L}{\partial \dot{q}^i} \tag{2.3}$$

is an integral of motion.

As a special case, consider the Lagrangian system of N interacting particles in $\mathbb{R}^3$ discussed in Example 1.2. We have $M = V = \mathbb{R}^{3N}$, and the Lagrangian L is invariant under a simultaneous translation of coordinates $\boldsymbol{r}_a = (x_a^1, x_a^2, x_a^3)$ of all particles by the same vector $\boldsymbol{c} \in \mathbb{R}^3$. Thus for every $\boldsymbol{c} = (c^1, c^2, c^3) \in \mathbb{R}^3$, the Lagrangian L is invariant under translations by a multiple of $v = (\boldsymbol{c}, \dots, \boldsymbol{c}) \in \mathbb{R}^{3N}$, which implies that

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

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 the momenta of individual particles. Conservation of total momentum is a fundamental physical law which reflects the homogeneity of the space.

Traditionally, for any Lagrangian system with local coordinates q^i, expressions $p_i = \dfrac{\partial L}{\partial \dot{q}^i}$ 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. Assume that the Lagrangian L is invariant under the action of some subgroup $G \subset \mathrm{SO}(V)$. Denote by $\mathfrak{g} \subset \mathfrak{so}(V)$ the Lie algebra of G. As discused in Corollary 2.6, in this case every $u \in \mathfrak{g}$ defines an integral of motion

$$I_u = \sum_{i=1}^{n} (u \cdot q)^i \, \frac{\partial L}{\partial \dot{q}^i}.$$

In particular, let (M, L) be a Lagrangian system of N interacting particles in $\mathbb{R}^3$, 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, in this case $G = \mathrm{SO}(3)$ is a group of symmetries of this system, and for every $u \in \mathfrak{so}(3)$,

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

is an integral of motion.

Let $u = u^1 J_1 + u^2 J_2 + u^3 J_3$, where J_1, J_2, J_3 is the standard basis in $\mathfrak{so}(3) \simeq \mathbb{R}^3$, introduced in Section 1.1. Then $u \cdot \boldsymbol{r}_a = \boldsymbol{u} \times \boldsymbol{r}_a$, where $\boldsymbol{u} = (u^1, u^2, u^3)$ (see Exercise 1.6) and 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 the angular momenta of individual particles. Conservation of angular momentum is a fundamental physical law which reflects the isotropy of the space.

In particular, for a single particle in $\mathbb{R}^3$, the generator J_i of $\mathfrak{so}(3)$ corresponds to the following integral of motion

$$M_i = (\boldsymbol{r} \times \boldsymbol{p})_i = \epsilon_{ijk} x^j p_k, \qquad \boldsymbol{p} = \frac{\partial L}{\partial \dot{\boldsymbol{r}}}. \tag{2.4}$$

Generalizations of these formulas from $\mathbb{R}^3$ to $\mathbb{R}^n$ is duscussed in Exercise 2.6.

Example 2.3 (The center of mass)**.** Let (M, L) be a Lagrangian system of N interacting particles, considered in Example 1.2. Under a one-parameter group of simultaneous

Galilean transformations of all coordinates $\boldsymbol{r}_a \mapsto g_s(\boldsymbol{r}_a) = \boldsymbol{r}_a + s\boldsymbol{v}t$, and corresponding transformations of velocities $\dot{\boldsymbol{r}}_a \mapsto \dot{\boldsymbol{r}}_a + s\boldsymbol{v}$, using (1.22) we have

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

Therefore for an 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 (2.2) holds, where the function K is given by

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

According to Proposition 2.10, 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

$$\boldsymbol{R} = \frac{1}{m}\sum_{a=1}^{N} m_a \boldsymbol{r}_a,$$

where $m = \sum_{a=1}^{N} m_a$ is the total mass, moves with the constant velocity $\boldsymbol{V} = \boldsymbol{P}/m$.

Of course, in case of a free particle we have rather obvious integrals motion, components of the vector $t\boldsymbol{P} - m\boldsymbol{r}$.

2.4. Exercises

Exercise 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.

Exercise 2.2 (Second tangent bundle)**.** Let $\pi\colon TM \to M$ 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 $i\colon \pi^*(TM) \simeq T_{\mathrm{V}}(TM)$, where $\pi^*(TM) \to TM$ is the pullback of the tangent bundle TM of M under the map π.

Exercise 2.3 (Invariant definition of the 1-form θ_L)**.** Let θ_L be the one-form on TM associated with a Lagrangian L (see (1.3)). Show that $\theta_L(v) = dL(\pi_* v)$, where $v \in T(TM)$ and π is the projection map $TM \to M$.

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

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

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

where $\dot{\gamma}'(t)$ is the velocity vector of the path $\gamma'(t)$ in TM and $\imath_{\dot{\gamma}'(t)}$ stands for the interior product with $\dot{\gamma}'(t)$.

Exercise 2.6. Let E_{ij} be the standard basis of matrix units in the algebra of $n \times n$ real matrices. Define

$$M_{ij} = E_{ji} - E_{ij}. \tag{2.5}$$

Obviously, $M_{ij} = -M_{ji}$ and M_{ij}, $i < j$, form a basis in the Lie algebra $\mathfrak{so}(n)$ of skew-symmetric matrices.

(1) Prove that so defined M_{ij} satisfy commutation relations

$$[M_{ij}, M_{kl}] = \delta_{jl} M_{ik} - \delta_{jk} M_{il} - \delta_{il} M_{jk} + \delta_{ik} M_{jl}. \tag{2.6}$$

(For $n = 3$, the basis of M_{ij} coincides with the basis J_i introduced in (1.15): $M_{ij} = \epsilon_{ijk} J_k$, e.g. $M_{12} = J_3$. Thus, formula above generalizes the commutation relations (1.18).)

We will generalize this further, replacing $\mathbb{R}^n$ by an arbitrary vector space with a non-degenerate bilinear form, in Exercise 7.1.

(2) Consider a mechanical system with configuration space $\mathbb{R}^n$ and a Lagrangian L which is invariant under the group $\mathrm{SO}(n)$. Show that in this case, the *angular momenta*

$$M_{ij} = q^i p_j - q^j p_i, \qquad p_i = \frac{\partial L}{\partial q^i} \tag{2.7}$$

are integrals of motion of this system. (Compare with Example 2.2).

Chapter 3

Integration of Equations of Motion

In this chapter, we give several examples in which one can not only write down the equations of motion, but can also solve them explicitly.

3.1. One-dimensional motion

The motion of systems with one degree of freedom (i.e., when the configuration space is one-dimensional) is called one-dimensional motion. In the simplest case $M = \mathbb{R}$, with global coordinate x, the Lagrangian takes the form

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

The conservation of energy

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

(see Section 2.1) allows one to solve the equations 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 $\frac{1}{2}m\dot{x}^2$ 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. 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.

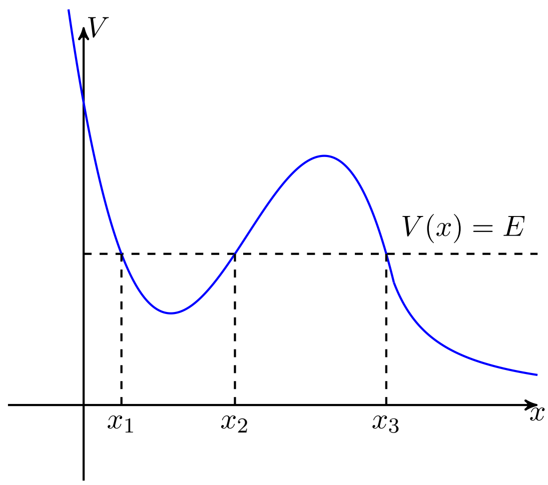

Figure 3.1

On the phase plane with coordinates (x, y), where $y = m\dot{x}$ is the momentum, 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.

3.2. Harmonic oscillator and small oscillations

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 1.5. The general solution of the equation of motion is

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

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

More generally, consider a particle of mass m in $\mathbb{R}^n$ 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 the 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

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

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

$$L = \tfrac{1}{2}m\dot{\boldsymbol{q}}^2 - \tfrac{1}{2}\sum_{i=1}^{n} \omega_i^2 (q^i)^2, \tag{3.1}$$

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 (non-interacting) harmonic oscillators with frequencies $\omega_1, \dots, \omega_n$.

3.3. Coupled harmonic oscillators

Consider a system of n particles on real line; for simplicity assume that each has mass 1 and the Lagrangian of the system has the form

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

where $\boldsymbol{q} = (q^1, \dots, q^n)$ are coordinates of the particles.

If $V(\boldsymbol{q}) = \sum V(q^i)$, then Euler–Lagrange equations become $\ddot{q}^i = -\dfrac{\partial V}{\partial q^i}$; this is a system of n independent equations, each involving only one of the variables q^i. In this case, physicists say that the particles are non-interacting. Otherwise, the particles are interacting, or *coupled.*

The simplest case of a system of interacting particles is given by the Lagrangian

$$L(\boldsymbol{q}) = \tfrac{1}{2}\dot{\boldsymbol{q}}^2 - V(\boldsymbol{q}), \tag{3.2}$$

where $V(\boldsymbol{q}) = \frac{1}{2}\sum a_{ij} q^i q^j$ is a positive definite quadratic form. In this case the same argument as in the previous section shows that one can find an orthogonal change of coordinates $Q^i = c_{ij} q^j$ such that in new coordinates, the Lagrangian becomes

$$L = \tfrac{1}{2}\sum_{i=1}^{n} \left((\dot{Q}^i)^2 - \omega_i^2 (Q^i)^2 \right)$$

where frequencies ω_i can be computed from the fact that $\lambda_i = \omega_i^2$ are eigenvalues of matrix (a_{ij}).

Thus, after the change of coordinates, the system becomes equivalent to a system of n decoupled harmonic oscillators, and

$$Q^i(t) = A_i \cos(\omega_i t + \alpha_i), \quad i = 1, \dots, n.$$

In physics literature, such solutions are called *normal modes* of the system.

Remark 3.1. If quadratic form $V(\boldsymbol{q})$ is only positive semi-definite, the normal modes Q^i with $\omega_i = 0$ will represent free particles.

Example 3.1 (One-dimensional crystal)**.** Consider the Lagrangian

$$L(\boldsymbol{q}, \dot{\boldsymbol{q}}) = \tfrac{1}{2}\sum_{j=1}^{N} (\dot{q}^j)^2 - \tfrac{1}{2}\sum_{j=1}^{N} \omega^2 (q^{j+1} - q^j)^2, \quad q^{N+1} = q^1. \tag{3.3}$$

Here periodic boundary condition $q^{N+1} = q^1$ mean that we have a system of N harmonic oscillators on the circle with the nearest-neighbor interaction. This system describes a one-dimensional "crystal" — a large polyatomic molecule. At the equilibrium its atoms are evenly spaced on a circle, and q_j is the displacement of a j-th atom from its equilibrium.

Lagrangian (3.3) has the form (3.2) with symmetric positive semi-definite $N \times N$ matrix $A = \{a_{ij}\}$:

$$A = \omega^2 \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & -1 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & -1 \\ -1 & 0 & 0 & \cdots & -1 & 2 \end{pmatrix}.$$

To diagonalize A, we use the $\mathbb{Z}_N$ symmetry of the problem. Namely, let $B = \{\delta_{i+1,j}\}$ be the $N \times N$ cyclic shift matrix, where the indices are considered modulo N.

It is easy to see that matrices A and B commute, so that A preserves the eigenspaces of B. Diagonalizing the matrix B is easy: it satisfies $B^N = I$, the $N \times N$ identity matrix, so its eigenvalues are the N-th roots of unity. Denoting by ζ a primitive N-th root of unity, it easy to check that the eigenvalues of B are exactly ζ^k, $k = 0 \dots N-1$, each with multiplicity 1; the corresponding eigenvectors are

$$v_k = \frac{1}{\sqrt{N}} \begin{pmatrix} \zeta^k \\ \vdots \\ \zeta^{Nk} \end{pmatrix}, \quad k = 0 \dots, N-1,$$

(they have been normalized to form an orthonormal basis). Since A commutes with B, each of these vectors is an eigenvector for A, and explicit computation shows that $Av_k = \lambda_k v_k$, where

$$\lambda_k = \omega^2 (2 - \zeta^k - \zeta^{-k}), \quad k = 0, \dots, N-1.$$

(This computation is just a simplest case of a general approach to diagonalizing operators using symmetries. Readers familiar with representation theory will observe that the same approach works for any system with a symmetry described by a finite or compact group G. In this case, eigenvectors of B should be replaced by irreducible representations of G.)

The eigenvector v_0 corresponds to the eigenvalue $\lambda_0 = 0$ and reflects the motion of the center of mass (see Example 2.3). Using the symmetry $\bar{v}_k = v_{N-k}$, it is easy to write down the corresponding orthonormal basis in $\mathbb{R}^N$ and to introduce the normal modes $Q^1, \dots, Q^N$. Correspondingly, Lagrangian (3.3) takes the form

$$L = \tfrac{1}{2}(\dot{Q}^1)^2 + \tfrac{1}{2}\sum_{i=2}^{N}\left((\dot{Q}^i)^2 - \omega_i^2 (Q^i)^2\right), \tag{3.4}$$

where for even N

$$\omega_{2l} = \omega_{2l+1} = 2\omega \sin\frac{\pi l}{N}, \quad l = 1, \dots, \frac{N}{2} - 1 \quad \text{and} \quad \omega_N = 2\omega,$$

and for odd N

$$\omega_{2l} = \omega_{2l+1} = 2\omega \sin\frac{\pi l}{N}, \quad l = 1, \dots, \frac{N-1}{2}.$$

3.4. 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 with constant velocity. 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

$$L = \tfrac{1}{2}\mu(\dot{r}^2 + r^2\dot{\varphi}^2) - V(r), \tag{3.5}$$

which describes the motion of a particle of mass μ in a central force field with potential $V(r)$.

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

$$\mu r^2\dot{\varphi} = M, \tag{3.6}$$

which is equivalent to *Kepler's second law*[1]. Using (3.6) we get for the total energy

$$E = \tfrac{1}{2}\mu(\dot{r}^2 + r^2\dot{\varphi}^2) + V(r) = \tfrac{1}{2}\mu\dot{r}^2 + V(r) + \frac{M^2}{2\mu r^2}. \tag{3.7}$$

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

$$t = \sqrt{\frac{\mu}{2}}\int \frac{dr}{\sqrt{E - V_{\text{eff}}(r)}}. \tag{3.8}$$

It follows from (3.6) that the angle φ is a monotonic function of t, given by another quadrature

$$\varphi = \frac{M}{\sqrt{2\mu}}\int \frac{dr}{r^2\sqrt{E - V_{\text{eff}}(r)}}, \tag{3.9}$$

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., $\Delta\varphi = 2\pi\dfrac{m}{n}$ 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 $\sqrt{\frac{2}{\mu}(E - V)}$.

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

3.5. Exercises

All problems below discuss the motion of a particle in a central force field described by (3.5).

Exercise 3.1. 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.

Exercise 3.2. Prove that all finite trajectories in the central field are closed only when

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

Exercise 3.3. Prove that trajectories for the Kepler problem $V(r) = -\alpha/r$ are conic sections.

Exercise 3.4. Show that the Kepler problem has three additional integrals of motion — the *Laplace-Runge-Lenz vector* $\boldsymbol{W}$, given by

$$\boldsymbol{W} = \dot{\boldsymbol{r}} \times \boldsymbol{M} - \frac{\alpha \boldsymbol{r}}{r}, \quad \boldsymbol{r} = (x^1, x^2, x^3).$$

Exercise 3.5. For the Kepler problem, consider vector fields $\boldsymbol{Y} = (Y^1, Y^2, Y^3)$ on $\mathbb{R}^6$, defined by (2.1) with $a^{ij}(\boldsymbol{r}, \dot{\boldsymbol{r}}) = 2\dot{x}^i x^j - x^i \dot{x}^j - \delta^{ij} \boldsymbol{r} \cdot \dot{\boldsymbol{r}}$. Prove that they satisfy (2.2) 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.

Exercise 3.6. 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 exercise.)

Chapter 4

Hamiltonian Formalism

In the previous chapters, we have discussed the Lagrangian formalism, where a state of the system is described by a point in the tangent bundle TM to the configuration space M, and equations of motion are derived from the principle of least action. In this chapter we will discuss an alternative approach, called the *Hamiltonian formalism*, in which a state of the system is described by a point in the cotangent bundle T^*M.

4.1. Legendre transform

Let T^*M be the cotangent bundle of an n-dimensional manifold M. As in case of the tangent bundle, a choice of local coordinates $\boldsymbol{q} = (q^1, \ldots, q^n)$ on open chart $U \subset M$ defines, for any point $x \in U$, a basis dq^i in the cotangent space $T^*_x M$ and thus a choice of coordinates

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

on T^*U. Such coordinates will be called *standard coordinates.*[1]

Definition 4.1. The 1-form θ on T^*M, defined in standard coordinates by

$$\theta = \sum_{i=1}^{n} p_i \, dq^i = \boldsymbol{p} \, d\boldsymbol{q}, \tag{4.1}$$

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

It is easy to show that θ doesn't depend on the choice of local coordinates and thus is well-defined on T^*M. It also admits an invariant definition:

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

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

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

Recall (see Section 1.3) that given a function L on TM, the partial derivatives $\dfrac{\partial L}{\partial \dot{\boldsymbol{q}}} = \left(\dfrac{\partial L}{\partial \dot{q}^1}, \dots, \dfrac{\partial L}{\partial \dot{q}^n}\right)$ transform as a one-form and thus can be considered as a point in the cotangent bundle T^*M.

Definition 4.2. Legendre transform associated with the Lagrangian L is the fiberwise mapping $\tau_L \colon TM \to T^*M$ given in the standard coordinates 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 element $\boldsymbol{p} \in T^*_{\boldsymbol{q}}M$ is called the *momentum*, and the space T^*M is called the *phase space*.

The Legendre transform also admits an invariant definition: it is uniquely defined by the condition

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

where $\theta_L = \sum_{i=1}^n \dfrac{\partial L}{\partial \dot{q}^i}\, dq^i$ (see (1.3)) and θ is the Liouville's canonical 1-form on T^*M.

Lemma 4.3. *The mapping τ_L is a local diffeomorphism if and only if the Lagrangian L is non-degenerate (see Definition 1.7).*

Example 4.1. Let $M = \mathbb{R}^n$ and let the Lagrangian be given by

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

In this case, it is immediate that the Legendre transform is given by $p_i = m\dot{q}^i$, or $\boldsymbol{p} = m\dot{\boldsymbol{q}}$.

More generally, if M is a Riemannian manifold, and $L = \frac{1}{2} m(\dot{\boldsymbol{q}}, \dot{\boldsymbol{q}}) - V(\boldsymbol{q})$, where $(\ ,\ )$ is the inner product defined by the Riemannian metric (see Example 1.6), then $\boldsymbol{p} \in T^*_{\boldsymbol{q}}M$ is defined by $\langle \boldsymbol{p}, \boldsymbol{v} \rangle = m(\dot{\boldsymbol{q}}, \boldsymbol{v})$ for any $\boldsymbol{v} \in T^*_{\boldsymbol{q}}M$, where $\langle\ ,\ \rangle$ is the canonical pairing between the tangent and cotangent spaces. Thus, up to a factor of m, in this case the Legendre transform τ_L is just the fiberwise isomorphism between a vector space and its dual, defined by the bilinear form $(\ ,\ )$.

4.2. Hamilton's equations

Let (M, L) be a Lagrangian mechanical system, and let $\tau_L \colon TM \to T^*M$ be the Legendre transform defined in the previous section. Assume that τ_L is a diffeomorphism. Then we show that in this case the equations of motion of the system (M, L), when rewritten in terms of the phase space T^*M, take an especially simple form.

Definition 4.4. Suppose that the Legendre transform $\tau_L \colon TM \to T^*M$ is a diffeomorphism. The *Hamiltonian* function $H \colon T^*M \to \mathbb{R}$, associated with the Lagrangian L is defined by

$$H(\boldsymbol{p}, \boldsymbol{q}) = E_L = \dot{\boldsymbol{q}}\, \frac{\partial L}{\partial \dot{\boldsymbol{q}}} - L, \qquad \text{where } (\boldsymbol{p}, \boldsymbol{q}) = \tau_L(\boldsymbol{q}, \dot{\boldsymbol{q}}).$$

In other words, the Hamiltonian is the energy function of the mechanical system, as defined in Section 2.1. In standard coordinates, the Hamiltonian is given by

$$H(\boldsymbol{p},\boldsymbol{q}) = (\boldsymbol{p}\dot{\boldsymbol{q}} - L(\boldsymbol{q},\dot{\boldsymbol{q}}))\big|_{\boldsymbol{p}=\frac{\partial L}{\partial \dot{\boldsymbol{q}}}}, \tag{4.2}$$

where $\dot{\boldsymbol{q}}$ is a function of $\boldsymbol{p}$ and $\boldsymbol{q}$ defined by the equation $\boldsymbol{p} = \dfrac{\partial L}{\partial \dot{\boldsymbol{q}}}(\boldsymbol{q},\dot{\boldsymbol{q}})$ (recall that we are assuming that τ_L is a diffeomorphism).

Theorem 4.5. *Suppose that the Legendre transform $\tau_L\colon TM \to T^*M$ is a diffeomorphism. Then the Euler-Lagrange equations (1.4) on TM are equivalent to the following system of first order differential equations in standard coordinates on T^*M:*

$$\dot{p}_i = -\frac{\partial H}{\partial q^i}, \quad \dot{q}^i = \frac{\partial H}{\partial p_i}, \quad i = 1,\dots,n. \tag{4.3}$$

These equations are called Hamilton's equations (or canonical equations).

Proof. We have

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

Thus under the Legendre transform,

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

It follows from the energy conservation law (Proposition 2.3) that the Hamiltonian is constant on the trajectories of the Hamilton's equations; we will give an independent proof of this result later (see Corollary 4.18).

Remark 4.6. Let $S(\boldsymbol{q}, t; \boldsymbol{q}_0, t_0)$ be the classical action, introduced in Remark 1.8. It follows from Exercise 1.5 that

$$\boldsymbol{p} = \frac{\partial S}{\partial \boldsymbol{q}} \quad \text{and} \quad H = -\frac{\partial S}{\partial t}, \tag{4.4}$$

where $\boldsymbol{p}$ is determined by the velocity $\dot{\boldsymbol{q}}$ of the central field extremal at time t.

4.3. Examples of Hamiltonian systems

For the Lagrangian

$$L = \tfrac{1}{2}m\dot{\boldsymbol{r}}^2 - V(\boldsymbol{r}) = T - V, \quad \boldsymbol{r} \in \mathbb{R}^3,$$

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

$$\boldsymbol{p} = \frac{\partial L}{\partial \dot{\boldsymbol{r}}} = m\dot{\boldsymbol{r}}. \tag{4.5}$$

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

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

In this case Hamilton's equations

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

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

For the Lagrangian system describing small oscillations, considered in Section 3.2, we have $\boldsymbol{p} = m\dot{\boldsymbol{q}}$, and using normal coordinates we get

$$H(\boldsymbol{p},\boldsymbol{q}) = (\boldsymbol{p}\dot{\boldsymbol{q}} - L(\boldsymbol{q},\dot{\boldsymbol{q}}))\big|_{\dot{\boldsymbol{q}}=\frac{\boldsymbol{p}}{m}} = \frac{\boldsymbol{p}^2}{2m} + V_0(\boldsymbol{q}) = \frac{1}{2m}\Big(\boldsymbol{p}^2 + m^2\sum_{i=1}^n \omega_i^2 (q^i)^2\Big). \tag{4.6}$$

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

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

In particular, for a system with pairwise interactions,

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

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

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

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

Remark 4.7. Writing equation (4.5) in coordinates gives $p_i = m\dot{x}^i$, which looks like a rather disheartening mixing of vector and covector indices. Of course, in this case we have a Lagrangian system with the matrix $a_{ij}(\boldsymbol{r}) = m\delta_{ij}$, so raising and lowering the indices is done by using the Euclidean metric on $\mathbb{R}^3$.

4.4. Symplectic manifolds

Hamilton's equations on the phase space T^*M can be rewritten in an invariant form, without using a choice of local coordinates. To do so, observe that T^*M has a structure of a *symplectic manifold*.

Definition 4.8. A symplectic manifold is a manifold $\mathscr{M}$ together with a 2-form $\omega \in \Omega^2(\mathscr{M})$ which is closed ($d\omega = 0$) and non-degenerate: for every $x \in \mathscr{M}$, the pairing $T_x\mathscr{M} \otimes T_x\mathscr{M} \to \mathbb{R}$, given by ω, is non-degenerate.

Example 4.2. Let $\mathscr{M} = T^*M$, and let $\omega = d\theta$, where $\theta \in \Omega^1(T^*M)$ is the Liouville form (4.1). Then $\mathscr{M}$ is a symplectic manifold, with the symplectic form $\omega = d\theta$.

Indeed, in standard local coordinates we have $\theta = \boldsymbol{p}\, d\boldsymbol{q} = \sum_{i=1}^n p_i\, dq^i$, and

$$\omega = d\boldsymbol{p} \wedge d\boldsymbol{q} = \sum_{i=1}^n dp_i \wedge dq^i, \tag{4.7}$$

which is easily seen to be non-degenerate.

Not every symplectic manifold has the form T^*M (for example, the two-dimensional sphere S^2 with the volume form is an example of a compact symplectic manifold); however, it is easy to see that any symplectic manifold is even-dimensional, and Darboux theorem states that locally any symplectic manifold admits coordinates p_i, q^i such that the symplectic form takes the form (4.7). Such coordinates are called *canonical or Darboux coordinates*.

Example 4.3. A non-degenerate Lagrangian function $L\colon TM \to \mathbb{R}$ provides a tangent bundle TM with a structure of a symplectic manifold. Indeed, the 2-form $\omega_L = d\theta_L$, where θ_L is defined by (1.3), is obviously closed and is non-degenerate (see Exercise 2.1). In this case, if the Legendre transform $\tau_L\colon TM \to T^*M$ is a global diffeomorphism, it is also an isomorphism of symplectic manifolds.

On a symplectic manifold $\mathscr{M}$, the symplectic form ω defines an isomorphism

$$J\colon T^*\mathscr{M} \to T\mathscr{M} \tag{4.8}$$

between tangent and cotangent bundles, given by

$$\omega(u, J\vartheta) = \langle u, \vartheta\rangle, \qquad u \in T_x\mathscr{M},\ \vartheta \in T_x^*\mathscr{M},$$

or equivalently

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

The mapping J induces the isomorphism

$$\Omega^1(\mathscr{M}) \simeq \mathrm{Vect}(\mathscr{M})$$

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

$$\omega(Y, X) = \langle Y, \vartheta\rangle \quad \text{for all} \quad Y \in \mathrm{Vect}(\mathscr{M}), \tag{4.9}$$

and, correspondingly,

$$\vartheta = J^{-1}(X) = -\imath_X\omega. \tag{4.10}$$

As a corollary, we see that on a symplectic manifold $\mathscr{M}$ any function $H \in C^\infty(\mathscr{M})$ defines a vector field $X_H = J(dH)$, called the *Hamiltonian vector field*. This vector field satisfies

$$dH = -\imath_{X_H}\omega. \tag{4.11}$$

In canonical coordinates, this vector field is given by

$$X_H = \sum_{i=1}^n \left(\frac{\partial H}{\partial p_i}\frac{\partial}{\partial q^i} - \frac{\partial H}{\partial q^i}\frac{\partial}{\partial p_i}\right) = \frac{\partial H}{\partial \boldsymbol{p}}\frac{\partial}{\partial \boldsymbol{q}} - \frac{\partial H}{\partial \boldsymbol{q}}\frac{\partial}{\partial \boldsymbol{p}}. \tag{4.12}$$

In particular,

$$X_{\boldsymbol{p}} = \frac{\partial}{\partial \boldsymbol{q}} \quad \text{and} \quad X_{\boldsymbol{q}} = -\frac{\partial}{\partial \boldsymbol{p}}.$$

Assume now that $\mathscr{M} = T^*M$ and $H = \boldsymbol{p}\dot{\boldsymbol{q}} - L$ is the Hamiltonian function defined in Definition 4.4. Comparing the formula for X_H given above with the Hamilton's equations (4.3), we see that these equations can be rewritten in the simple form

$$\dot{x} = X_H, \tag{4.13}$$

where $x = (\boldsymbol{p}, \boldsymbol{q}) \in T^*M$ and H is the Hamiltonian.

Motivated by this, we will call any pair $(\mathscr{M}, H)$, where $\mathscr{M}$ is a symplectic manifold and $H \in C^\infty(\mathscr{M})$ a *Hamiltonian system*, and will call (4.13) equations of motion, or Hamilton's equations, for this system.

If the vector field X_H on $\mathscr{M}$ is complete, i.e., its integral curves exist for all times, then the corresponding one-parameter group $\{g_t\}_{t\in\mathbb{R}}$ of diffeomorphisms of $\mathscr{M}$ generated by X_H is called the *Hamiltonian phase flow* generated by H. It is defined by $g_t(x) = x(t)$, where $x(t)$ is a solution of Hamilton's equations satisfying $x(0) = x$.

Theorem 4.9. *Hamiltonian phase flow on a symplectic manifold $\mathscr{M}$ preserves the symplectic form.*

Proof. It suffices to prove that $L_{X_H}\omega = 0$, where L_{X_H} is the Lie derivative along the Hamiltonian vector field X_H. Using the Cartan identity

$$L_X(\omega) = d(\imath_X\omega) + \imath_X(d\omega) \tag{4.14}$$

and the fact that the symplectic form is by definition closed, we get $L_X\omega = d(\imath_X\omega)$. On the other hand, for a Hamiltonian vector field $X = X_H$, we have by (4.10) that $i_X\omega = -dH$, so $L_{X_H}\omega = -d(dH) = 0$. □

The canonical symplectic form ω on a symplectic manifold $\mathscr{M}$ defines the volume form

$$\frac{\omega^n}{n!} = \frac{1}{n!}\underbrace{\omega \wedge \cdots \wedge \omega}_{n}, \quad n = \dim \mathscr{M}/2, \tag{4.15}$$

on $\mathscr{M}$, called the *Liouville's volume form.*

Corollary 4.10 (Liouville's theorem)**.** *The Hamiltonian phase flow on T^*M preserves Liouville's volume form.*

4.5. Lagrangian submanifolds and polarizations

The basic example of a symplectic manifold is the cotangent bundle: $\mathscr{M} = T^*M$. It is easy to see, however, that not every symplectic manifold has such a form (e.g., a compact symplectic manifold can never be a cotangent bundle). Instead, we can introduce a weaker notion.

Definition 4.11. A submanifold $\mathscr{L}$ of a symplectic manifold $\mathscr{M}$ is called a *Lagrangian submanifold* if $\dim \mathscr{L} = \dim \mathscr{M}/2$ and $\omega|_{\mathscr{L}} = 0$.

It follows from Theorem 4.9 that the image of a Lagrangian submanifold under the Hamiltonian phase flow is a Lagrangian submanifold.

For example, for $\mathscr{M} = T^*M$, each fiber T_xM is a Lagrangian submanifold. This motivates the follwoing definition.

Definition 4.12. Let $\mathscr{M}$ be a symplectic manifold of dimension $2n$. A *polarization* of $\mathscr{M}$ is a foliation of $\mathscr{M}$ whose leaves are (locally) Lagrangian submanifolds.

Equivalently, the polarization is described by a subbundle $V \subset T\mathscr{M}$ such that

(1) For every point $x \in \mathscr{M}$, $V_x \subset T_x\mathscr{M}$ is a Lagrangian subspace.
(2) V is an integrable distribution.

In the case of cotangent bundle $\mathscr{M} = T^*M$, we have a canonical polarization, whose leaves are fibers of the natural projection $T^*M \to M$. In general, polarization can be thought of as representing $\mathscr{M}$ locally as the fiber bundle with Lagrangian fibers. Note that not every symplectic manifold admits a polarization.

4.6. Classical observables and Poisson bracket

Let $\mathscr{M}$ be a symplectic manifold, with symplectic form ω (for example, $\mathscr{M} = T^*M$). Any function $f \in C^\infty(\mathscr{M})$ is called an *observable*; from physicists' point of view, such functions represent measurable characteristics of a system, such as coordinates, momenta, or energy.

Functions on any manifold form a commutative algebra with respect to pointwise multiplication. However, it turns out that on a symplectic manifold, the algebra of functions has an additional structure, called the *Poisson bracket.*

Recall that on a symplectic manifold any function f gives rise to a Hamiltonian vector field $X_f = J(df)$, defined by

$$\omega(v, X_f) = \langle v, df \rangle = \partial_v f$$

for any vector field v on $\mathscr{M}$.

Definition 4.13. Let $f, g \in C^\infty(\mathscr{M})$. Their Poisson bracket $\{f, g\} \in C^\infty(\mathscr{M})$ is defined by

$$\{f, g\} = \partial_{X_f} g = \omega(X_f, X_g). \tag{4.16}$$

This definition immediately implies that for any observable H, we have

$$\partial_{X_H} f = \{H, f\}. \tag{4.17}$$

It follows from (4.12) that in canonical coordinates the Poisson bracket takes the form

$$\{f, g\} = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q^i} - \frac{\partial f}{\partial q^i} \frac{\partial g}{\partial p_i} \right) = \frac{\partial f}{\partial \boldsymbol{p}} \frac{\partial g}{\partial \boldsymbol{q}} - \frac{\partial f}{\partial \boldsymbol{q}} \frac{\partial g}{\partial \boldsymbol{p}}. \tag{4.18}$$

In particular,

$$\begin{aligned} \{p_i, p_j\} &= \{q^i, q^j\} = 0, \\ \{p_i, q^j\} &= \delta_i^j. \end{aligned} \tag{4.19}$$

These relations are commonly called the *canonical relations.*

Theorem 4.14. *The Poisson bracket has the following properties.*

(1) *It satisfies the Leibniz rule:*

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

(2) *It is skew-symmetric:* $\{f, g\} = -\{g, f\}$.

(3) *It satisfies Jacobi identity:*

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

(4)

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

where the $[\,,\,]$ *in the right hand side is the commutator of vector fields, which is defined by*

$$\partial_{[\xi,\eta]} = \partial_\xi \partial_\eta - \partial_\eta \partial_\xi.$$

Proof. The Leibniz rule and the skew-symmetry immediately follow from the definition (4.16) of the Poisson bracket.

Jacobi identity can easily verified by a straightforward computation using formula (4.18) for the Poisson bracket in canonical coordinates. The more conceptual proof goes as follows.

Let ξ be a Hamiltonian vector field on $\mathscr{M}$, and let φ_t be the time t flow of ξ (defined for small t in a neighborhood of some point). By Theorem 4.9 we have $\varphi_t^*(\omega) = \omega$, which together with (4.16) implies that the Poisson bracket is equivariant with respect to φ_t:

$$\varphi_t^*\{f, g\} = \{\varphi_t^* f, \varphi_t^* g\}.$$

Taking derivative at $t = 0$, we get

$$\partial_\xi \{f, g\} = \{\partial_\xi f, g\} + \{f, \partial_\xi g\}.$$

In particular, if $\xi = X_h$ so that $\partial_\xi f = \{h, f\}$, we get

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

which is equivalent to Jacobi identity.

The last part follows from the Jacobi identity: indeed, rewriting it in the form

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

we see that

$$\partial_{X_{\{f,g\}}} h = \partial_{X_f}\partial_{X_g} h - \partial_{X_g}\partial_{X_f} h.$$

□

Parts (2)–(4) of the theorem can be restated as follows.

Corollary 4.15. *The Poisson bracket defines on the space of observables $C^\infty(\mathscr{M})$ a structure of a Lie algebra, and $C^\infty(\mathscr{M}) \to \mathrm{Vect}(\mathscr{M})\colon f \mapsto X_f$ is a morphism of Lie algebras.*

Remark 4.16. A manifold P with a bilinear operation $\{\ ,\ \}$ on $C^\infty(P)$, satisfying properties (1)–(3) of Theorem 4.14 is called a *Poisson manifold*; thus, Theorem 4.14 can be restated by saying that any symplectic manifold is also automatically a Poisson manifold. Converse is obviously false (e.g., one can take the Poisson bracket to be identically zero).

Using Poisson bracket, one can rewrite Hamiltonian equations of motion in terms of observables as follows.

Lemma 4.17. *Let $(\mathscr{M}, H)$ be a Hamiltonian system. Let $x(t)$ be a classical trajectory, i.e. a solution of Hamilton's equations. Then for any observable $f \in C^\infty(\mathscr{M})$, we have*

$$\frac{d}{dt} f(x(t)) = \{H, f\}(x(t)). \tag{4.20}$$

Proof. By (4.13), we have $\dot{x} = X_H$, so $\dfrac{df}{dt} = \partial_{X_H} f = \{H, f\}$. □

As before, we say that an observable f is an integral of motion (or conserved quantity) if it is constant on classical trajectories: $\dfrac{d}{dt} f(x(t)) = 0$.

Corollary 4.18. *An observable $f \in C^\infty(\mathscr{M})$ is an integral of motion of a system $(\mathscr{M}, H)$ if and only if $\{H, f\} = 0$.*

In particular, this immediately implies that H itself is an integral of motion: since the Poisson bracket is skew-symmetric, we have $\{H, H\} = 0$.

4.7. Completely integrable systems

Let $(\mathscr{M}, H)$ be a Hamiltonian system.

Lemma 4.19. *The space $\mathcal{I} \subset C^\infty(\mathscr{M})$ of all integrals of motion of a system $(\mathscr{M}, H)$ is closed under the Poisson bracket.*

Proof. By Corollary 4.18, an observable f is an integral of motion iff $\{H, f\} = 0$. Now the result follows from Jacobi identity: if $\{H, f\} = \{H, g\} = 0$, then

$$\{H, \{f, g\}\} = \{\{H, f\}, g\} + \{f, \{H, g\}\} = 0. \quad \square$$

Of special interest is the case when we have a collection of integrals of motion $I_1, \dots, I_k$ such that $\{I_a, I_b\} = 0$ for all a, b. In this case we say that these integrals of motion Poisson-commute, or that they are *in involution*. If, in addition, we have an open subset $U \subset \mathscr{M}$ such that the covectors $dI_1(x), \dots, dI_k(x) \in T^*_x\mathscr{M}$ are linearly independent for every point $x \in U$, then we say that $I_1, \dots, I_k$ are *independent* on U.

Lemma 4.20. *Let $I_1, \dots, I_k$ be integrals of motion in involution, which are independent on open $U \subset \mathscr{M}$, and let $X_1 = X_{I_1}, \dots, X_k = X_{I_k}$ be the corresponding Hamiltonian vector fields. Let $\mathcal{D} \subset T\mathscr{M}$ be the subbundle generated by $X_1, \dots, X_k$. Then $\mathcal{D}$ is an integrable distribution, so that through every point $x \in U$ there is a unique (locally) submanifold $N \ni x$ of dimension k such that X_i are tangent to N. Moreover, this integral submanifold is* isotropic*: restriction of the symplectic form ω to TN is zero.*

This lemma is a very special case of general Frobenius integrability criterion: since the functions I_a are in involution, the vector fields X_a commute. The fact that N is isotropic follows from $\omega(X_a, X_b) = \{I_a, I_b\} = 0$.

Let $\dim \mathscr{M} = 2n$. Since it is well-known that the dimension of an isotropic subspace in $2n$-dimensional symplectic vector space can not be more than n, we see that one can not have more than n independent integrals of motion in involution.

Definition 4.21. A Hamiltonian system $(\mathscr{M}, H)$, $\dim \mathscr{M} = 2n$, is called *completely integrable* if there exist n integrals of motion $I_1, \dots, I_n$ in involution which are independent everywhere on $\mathscr{M}$.

Remark 4.22. Sometimes the requirement of being everywhere independent is replaced by the requirement of being independent on an open dense subset.

A fundamental result of Liouville shows that every completely integrable system is in a certain sense exactly solvable. We present it here, without proof, in a modern form due to Arnold. Before stating it, recall that a (multivalued) complex function of one variable is *expressed by quadratures* if it can be obtained from constant functions by using arithmetic operation, logarithm, exponential function, differentiation and integration. Similarly, a differential equation is solved by quadratures if its solutions are expressed by quadratures. It should be noted that this is a rare and wonderful event: general differential equation, even in one variable, is not solvable by quadratures.

Theorem 4.23 (Liouville–Arnold)**.** *Let $(\mathscr{M}, H)$, $\dim \mathscr{M} = 2n$, be a completely integrable system, with independent integrals of motion in involution $H = I_1, I_2, \dots, I_n$. For $\boldsymbol{a} = (a_1, \dots, a_n)$, let $M_{\boldsymbol{a}} = \{x \in \mathscr{M} \mid I_i(x) = a_i\}$ be the level set of $I_1, \dots, I_n$. Then*

(1) *For each $\boldsymbol{a}$, $M_{\boldsymbol{a}}$ is a smooth Lagrangian submanifold in $\mathscr{M}$, invariant under the flow of each of vector fields $X_i = X_{I_i}$ (in particular, under the flow of X_H).*

(2) *Assume additionally that $M_{\boldsymbol{a}}$ is compact. Then each connected component M of $M_{\boldsymbol{a}}$ is a torus: $M \simeq T^n$, where $T = \mathbb{R}/2\pi\mathbb{Z}$. Moreover, there exist a neighborhood $U_{\boldsymbol{a}} \subset \mathbb{R}^n$ of $\boldsymbol{a}$, a neighborhood U of M and an isomorphism $U \simeq U_{\boldsymbol{a}} \times T^n \colon u \mapsto (I_1, \dots, I_n, \varphi_1, \dots, \varphi_n)$ such that the Hamiltonian H only*

depends on $\boldsymbol{I} = (I_1, \dots, I_n)$*, and the symplectic form* ω *is given by*

$$\omega = d\boldsymbol{I} \wedge d\boldsymbol{\varphi} = \sum_{i=1}^{n} dI_i \wedge d\varphi_i.$$

Coordinates I_i, φ_i *are called* action-angle variables.

(3) *Hamilton's equations on* M *are solved by quadratures.*

In action-angle variables Hamilton's equations take form

$$\begin{aligned} \dot{I}_i &= 0, \\ \dot{\varphi}_i &= \frac{\partial H}{\partial I_i} = \omega_i(\boldsymbol{I}), \end{aligned}$$

and the trajectories are quasiperiodic: time t flow is given by

$$\begin{aligned} I_i(t) &= const, \\ \varphi_i(t) &= \varphi_i(0) + \omega_i(\boldsymbol{I})t \mod 2\pi\mathbb{Z}. \end{aligned}$$

The Liouville-Arnold theorem explains a special role of integrable systems in classical mechanics; essentially all systems we can solve explicitly are completely integrable.

Example 4.4. Let $\dim \mathscr{M} = 2$, so that $n = 1$. In this case any such system is completely integrable (assuming that $dH \neq 0$): one can take $I_1 = H$.

Example 4.5. Consider a particle in a two-dimensional central field: $\mathscr{M} = T^*\mathbb{R}^2$, with

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

The Hamiltonian system $(\mathscr{M}, H)$ is completely integrable, as one can take energy E and the total angular momentum $M = |\boldsymbol{p} \times \boldsymbol{r}|$ as two independent integrals of motion.

4.8. Exercises

Exercise 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 $\boldsymbol{p} = \dfrac{\partial L}{\partial \dot{\boldsymbol{q}}}$.

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

Exercise 4.3. Let $(\mathscr{M}, H)$ be a Hamiltonian system. Show that if all level sets of H are compact submanifolds of $\mathscr{M}$, then the Hamiltonian vector field X_H is complete.

Exercise 4.4. Prove that $L_{X_H}(\theta) = d(-H + \theta(X_H))$, where θ is Liouville's canonical 1-form.

Exercise 4.5 (The principle of least action in Hamiltonian mechanics)**.** Consider the Hamiltonian system $\mathscr{M} = T^*M$, $H \in C^\infty(\mathscr{M})$. For a path $\gamma \colon [t_0, t_1] \to \mathscr{M}$ in the phase space, let $\tilde{\gamma}$ be its lift to the extended phase space $\mathscr{M} \times \mathbb{R}$: $\tilde{\gamma}(t) = (\gamma(t), t)$. For such a path, define the action functional by

$$S(\gamma) = \int_{\tilde{\gamma}} \boldsymbol{p}\, d\boldsymbol{q} - H\, dt.$$

Show that then for any pair of points $\boldsymbol{q}_0, \boldsymbol{q}_1 \in M$, extremals of the action on the space

$$P(M)^{\boldsymbol{q}_1,t_1}_{\boldsymbol{q}_0,t_0} = \{\gamma\colon [t_0,t_1] \to \mathscr{M} \mid \boldsymbol{q}(t_0) = \boldsymbol{q}_0,\ \boldsymbol{q}(t_1) = \boldsymbol{q}_1\}.$$

are exactly the solutions of the Hamiltons' equations (4.3). Note that we only fix the generalized coordinates $\boldsymbol{q}(t_0), \boldsymbol{q}(t_1)$; there are no restrictions on the momenta $\boldsymbol{p}(t_0), \boldsymbol{p}(t_1)$.

Exercise 4.6. Let $\mathscr{M} = T^*\mathbb{R}^3$, and let M_1, M_2, M_3 be the components of the angular momentum $\boldsymbol{M} = \boldsymbol{r} \times \boldsymbol{p}$ (see (2.4)):

$$M_i = \epsilon_{ijk} x^j p_k.$$

(1) Show by explicit computation that the Poisson brackets of M_i are given by

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

(these could also be derived from general theory, see Example 5.3).

(2) Prove that $\boldsymbol{M}^2 = M_1^2 + M_2^2 + M_3^2$ Poisson commutes with each of M_i, i.e. $\{\boldsymbol{M}^2, M_i\} = 0$.

Exercise 4.7. Let $\mathscr{M}$ be a symplectic manifold, and let V be a polarization of $\mathscr{M}$. For an open subset $U \subset \mathscr{M}$, denote by $\mathcal{A}_V(U) \subset C^\infty(U)$ the space of functions which are locally constant along the leaves of the foliation:

$$\mathcal{A}_V(U) = \{f \in C^\infty(U) \mid \partial_\xi f = 0\, \forall \xi \in V\}$$

Prove that then the Poisson bracket vanishes on $\mathcal{A}_V(U)$: for any $f, g \in \mathcal{A}_V(U)$, we have $\{f, g\} = 0$.

In particular, in the case $\mathscr{M} = T^*M$, with canonical polarization, we get $\mathcal{A}_V(\mathscr{M}) \simeq C^\infty(M)$.

Exercise 4.8. Let $\mathfrak{g}$ be a finite-dimensional real Lie algebra, and let $\mathfrak{g}^*$ be its dual space. Define a Poisson bracket on $\mathfrak{g}^*$ (considered as a smooth manifold) by letting, for $f, g \in C^\infty(\mathfrak{g}^*)$

$$\{f, g\}(u) = \langle u, [df(u), dg(u)]\rangle, \qquad u \in \mathfrak{g}^*,$$

where $\langle\ ,\ \rangle$ is the pairing between $\mathfrak{g}^*$ and $\mathfrak{g}$, and we use a natural identification $T^*_u\mathfrak{g}^* = (\mathfrak{g}^*)^* = \mathfrak{g}$ to consider $df(u)$, $dg(u)$ as vectors in $\mathfrak{g}$. Prove that $\{\ ,\ \}$ defines on $\mathfrak{g}^*$ a structure of a Poisson manifold. (This Poisson bracket was introduced by Sophus Lie and is called *linear*, or *Lie-Poisson* bracket.)

Show that $\{f, g\}(u) = 0$ for all g if and only if $\operatorname{ad}^* a \cdot u = 0$, where $a = df(u) \in \mathfrak{g}$ and ad^* is the coadjoint action: $\langle \operatorname{ad}^* a \cdot u, b\rangle = -\langle u, [a, b]\rangle$.

Exercise 4.9. In the notation of the previous problem, let G be a connected Lie group with Lie algebra $\mathfrak{g}$. Let Ad^* be the coadjoint action of G on $\mathfrak{g}^*$, defined by $\langle \operatorname{Ad}^* g \cdot u, x\rangle = \langle u, \operatorname{Ad}(g^{-1}) \cdot x\rangle$, where $u \in \mathfrak{g}^*$, $x \in \mathfrak{g}$.

For an element $u \in \mathfrak{g}^*$ let $\mathcal{O}_u = G \cdot u$ be the orbit of the coadjoint action containing u.

(1) Let $f \in C^\infty(\mathfrak{g}^*)$; as before, let a vector field X_f on $\mathfrak{g}^*$ be defined by $\partial_{X_f}(g) = \{f, g\}$, where $\{\ ,\ \}$ is the Lie–Poisson bracket. Show that then, for any $u \in \mathfrak{g}^*$, $X_f(u)$ is tangent to the coadjoint orbit $\mathcal{O}_u$.

(2) Show that the Lie–Poisson bracket descends to a Poisson bracket on $\mathcal{O}_u$.

(3) Show that the Poisson structure on $\mathcal{O}_u$ defined above is non-degenerate and thus, $\mathcal{O}_u$ has a canonical structure of a symplectic manifold. The corresponding symplectic form is called the *Kirillov-Kostant-Souriau form.*

Chapter 5

Hamiltonian Action and Moment Map

5.1. Hamiltonian actions

Let G be a Lie group with Lie algebra $\mathfrak{g}$, and let $\rho\colon G \to \mathrm{Diff}(M)$ be an action of G on a smooth manifold M, which we will simply denote by $g \cdot x$. In this case, every $a \in \mathfrak{g}$ defines a vector field ξ_a on M, a generator of the one-parameter group $g_s = e^{sa}$ of diffeomorphisms of M.

Lemma 5.1. *The map*

$$\begin{aligned} \mathfrak{g} &\to \mathrm{Vect}(M) \\ a &\mapsto -\xi_a \end{aligned} \tag{5.1}$$

(note the minus sign!) is a morphism of Lie algebras:

$$-\xi_{[a,b]} = [-\xi_a, -\xi_b], \quad a, b \in \mathfrak{g}$$

where, as before, commutator of vector fields is defined by $\partial_{[\xi,\eta]} = \partial_\xi \partial_\eta - \partial_\eta \partial_\xi$.

Proof. It immediately follows from the fact that we have an action of G on the space $C^\infty(M)$, defined by $(g \cdot f)(x) = f(g^{-1}x)$, where $g \in G, f \in C^\infty(M)$, so

$$-\xi_a(f)(x) = \frac{d}{ds}\bigg|_{s=0} f(e^{-sa}x)$$

is the action of Lie algebra $\mathfrak{g}$. □

Note that without the negative sign in (5.1), the lemma would fail.

Assume now that $\mathscr{M}$ is a symplectic manifold with the symplectic form ω, and G acts on $\mathscr{M}$ by symplectomorphisms: for any $g \in G$, $g^*(\omega) = \omega$. In this case, for every $a \in \mathfrak{g}$, the corresponding vector field ξ_a satisfies $L_{\xi_a}(\omega) = 0$, where L is the Lie derivative. It is easy to show (see the proof of Theorem 4.9) that this implies that locally, ξ_a is a Hamiltonian vector field: there exists (locally) a smooth function H_a such that $X_{H_a} = -\xi_a$, where $X_H = J(dH)$

is the vector field defined by (4.12). The global existence on M of such a generating function H_a is not guaranteed, so we adopt the following definition.

Definition 5.2. An action of a Lie group G by symplectomorphisms on a symplectic manifold $\mathscr{M}$ is called *Hamiltonian* if we are given, for every $a \in \mathfrak{g}$, a function $H_a \in C^\infty(\mathscr{M})$ such that

(1) For any $a \in \mathfrak{g}$, $-\xi_a = X_{H_a}$.
(2) The map $\mathfrak{g} \to C^\infty(\mathscr{M})\colon a \mapsto H_a$ is linear.
(3) $\{H_a, H_b\} = H_{[a,b]}$.

In other words, in this case we have a commutative diagram

$$\begin{array}{ccc} \mathfrak{g} & \xrightarrow{a \mapsto H_a} & C^\infty(M) \\ & {\scriptstyle a\mapsto -\xi_a} \searrow \quad \swarrow {\scriptstyle H\mapsto X_H} & \\ & \mathrm{Vect}(M) & \end{array}$$

Note that all maps in this diagram are morphisms of Lie algebras. Also note that $-\xi_a = X_{H_a}$ is equivalent to the condition

$$\xi_a(f) = -\{H_a, f\} = \{f, H_a\} \quad \text{for all} \quad f \in C^\infty(\mathscr{M}). \tag{5.2}$$

It is easy to show that if G is connected (which we will always assume below unless explicitly stated otherwise), conditions (1)–(3) imply that the map $\mathfrak{g} \to C^\infty(\mathscr{M})$, given by $a \mapsto H_a$, is G-equivariant: $H_{\mathrm{Ad}\, g \cdot a} = g^* H_a$, for any $g \in G$, where Ad stands for the adjoint action of G on $\mathfrak{g}$.

For future use, we elucidate the role of condition (3) in this definition. Namely, it follows from Theorem 4.14, Part (4), and Lemma 5.1 that conditions (1)–(2) imply that

$$X_{\{H_a, H_b\}} = [X_{H_a}, X_{H_b}] = X_{H_{[a,b]}},$$

so

$$\{H_a, H_b\} = H_{[a,b]} + c(a,b)$$

for some locally constant function $c(a,b)$. In case when $\mathscr{M}$ is connected, $c(a,b)$ is constant, and we get a skew-symmetric function

$$c\colon \mathfrak{g} \otimes \mathfrak{g} \to \mathbb{R}.$$

It follows from the Jacobi identity that c is a Lie algebra 2-cocycle:

$$c([x,y],z) + c([y,z],x) + c([z,x],y) = 0, \quad x, y, z \in \mathfrak{g}$$

(we refer to Section 7.4 for the discussion of the Lie algebra cohomology). Note that conditions (1)–(2) determine the Hamiltonians H_a up to the addition of a linear function $\langle \lambda, a \rangle$ for some $\lambda \in \mathfrak{g}^*$. It is easy to see that under such change, cocycle $c(a,b)$ is changed by a coboundary:

$$c(a,b) \mapsto c(a,b) + \delta_1 \lambda(a,b) = c(a,b) + \langle \lambda, [a,b] \rangle.$$

In other words, such a change allows one to make the cocycle c zero if and only if its cohomology class[1] $[c]$ in $H^2(\mathfrak{g})$ is zero, and this is the meaning of condition (3). In particular,

[1] In notation of Section 7.4, $H^2(\mathfrak{g}) = H^2(\mathfrak{g}, \mathbb{R})$, where $\mathbb{R}$ is considered as trivial $\mathfrak{g}$-module.

if Lie algebra $\mathfrak{g}$ is semisimple, it is known that $H^2(\mathfrak{g}) = 0$, and condition (3) is satisfied automatically.

If the cohomology class $[c] \in H^2(\mathfrak{g})$ is non-zero, we can use it to define a central extension $\tilde{\mathfrak{g}}$ of $\mathfrak{g}$, by letting $\tilde{\mathfrak{g}} = \mathfrak{g} \oplus \mathbb{R} \cdot z$ with the commutator given by

$$[a, b]_{\tilde{\mathfrak{g}}} = [a, b] + c(a, b)z, \quad [z, a] = 0, \quad a \in \mathfrak{g}.$$

In this case, extending the map $\mathfrak{g} \to C^\infty(\mathscr{M})$ by letting $H_z = 1$, we see that this map is a morphism of Lie algebras. In other words, in the situation when $[c] \neq 0$, the Poisson brackets of Hamiltonians H_a satisfy the relations of a central extension $\tilde{\mathfrak{g}}$ of $\mathfrak{g}$.

An alternative way of describing Hamiltonian actions is by using the moment map.

Definition 5.3. Let $\rho\colon G \to \mathrm{Diff}(\mathscr{M})$ be a Hamiltonian action, with the Hamiltonian functions H_a. Then the moment map $\mu\colon \mathscr{M} \to \mathfrak{g}^*$ is defined by

$$\langle \mu(x), a\rangle = H_a(x), \quad x \in \mathscr{M}, \quad a \in \mathfrak{g},$$

where $\langle\ ,\ \rangle$ is the pairing between $\mathfrak{g}^*$ and $\mathfrak{g}$.

It follows from the definition of the Hamiltonian action that for connected G, the moment map μ is G-equivariant,

$$\mathrm{Ad}^* g \cdot \mu(x) = \mu(g \cdot x),$$

where Ad^* is a coadjoint action of G on $\mathfrak{g}^*$, and that for any vector field v on $\mathscr{M}$,

$$\partial_v \langle \mu(x), a\rangle = \partial_v H_a(x) = \omega(v, -\xi_a)(x), \qquad a \in \mathfrak{g}. \tag{5.3}$$

Conversely, it is easy to see that if we have an action of G on $\mathscr{M}$ and a G-equivariant map $\mu\colon \mathscr{M} \to \mathfrak{g}^*$ satisfying (5.3), then this action is Hamiltonian, with $H_a(x) = \langle \mu(x), a\rangle$.

Example 5.1. Let M be a manifold with an action of a Lie group G. Then the corresponding action of G on $\mathscr{M} = T^*M$ is Hamiltonian, and the moment map is given by

$$H_a(\boldsymbol{p}, \boldsymbol{q}) = \langle \mu(\boldsymbol{p}, \boldsymbol{q}), a\rangle = -\langle \boldsymbol{p}, \xi_a(\boldsymbol{q})\rangle, \qquad \boldsymbol{p} \in T^*_{\boldsymbol{q}}M,\ \boldsymbol{q} \in M, \quad a \in \mathfrak{g}.$$

Indeed, in this case it is easy to see that the Liouville one-form $\theta = \boldsymbol{p}\, d\boldsymbol{q}$ on T^*M is G-invariant, so for any $a \in \mathfrak{g}$

$$L_{\xi'_a}(\theta) = 0.$$

Here ξ'_a is the vector field on on T^*M, induced by the vector field ξ_a on M, and L is the Lie derivative. It follows from the Cartan's formula (4.14) that

$$0 = L_{\xi'_a}(\theta) = i_{\xi'_a}\omega + d(\langle \theta, \xi'_a\rangle), \quad \text{where} \quad \omega = d\theta.$$

Therefore, $i_{\xi'_a}\omega = dH_a$, where $H_a \in C^\infty(T^*M)$ is given by

$$H_a(\boldsymbol{p}, \boldsymbol{q}) = -\langle \theta, \xi'_a\rangle(\boldsymbol{p}, \boldsymbol{q}) = -\langle \boldsymbol{p}, \xi_a(\boldsymbol{q})\rangle.$$

We leave it as an exercise to check that so defined Hamiltonians satisfy $\{H_a, H_b\} = H_{[a,b]}$.

In particular, if $M = V$ is a vector space, so that $T^*V = V \oplus V^*$, $G = \mathrm{GL}(V)$, then $\xi_a(\boldsymbol{q}) = a \cdot \boldsymbol{q}$ and thus

$$H_a(\boldsymbol{p}, \boldsymbol{q}) = \langle \mu(\boldsymbol{p}, \boldsymbol{q}), a\rangle = -\langle \boldsymbol{p}, a \cdot \boldsymbol{q}\rangle, \qquad \boldsymbol{p} \in V^*,\ \boldsymbol{q} \in V.$$

Choosing standard coordinates q^i, p_i on T^*V, we see that the Hamiltonian function for a matrix $a = (a_{ij}) \in \mathfrak{gl}(n)$ is

$$H_a = -\sum_{i,j=1}^{n} a_{ij} p_i q^j.$$

5.2. Noether's theorem

Recall that in Section 2.2 we have proved Noether's theorem in Lagrangian mechanics: any one-parameter group of symmetries of a Lagrangian system gives rise to an integral of motion. In this section, we give an analog of this result in Hamiltonian mechanics.

Theorem 5.4. *Let $(\mathscr{M}, H)$ be a Hamiltonian system, with a Hamiltonian action of a Lie group G which preserves H. Then for each $a \in \mathfrak{g}$, the corresponding Hamiltonian function H_a is an integral of motion of $(\mathscr{M}, H)$.*

Proof. The proof is immediate: since the action of G preserves H, it implies that $\partial_{\xi_a} H = 0$ for any $a \in \mathfrak{g}$. On the other hand, $\xi_a = -X_{H_a}$, so $\partial_{\xi_a} H = -\{H_a, H\}$; now the result follows from Corollary 4.18. □

Note that not only each $a \in \mathfrak{g}$ gives rise to an integral of motion, but by the definition of the Hamiltonian action, Poisson brackets of these integrals satisfy the same relations as commutators in $\mathfrak{g}$: $\{H_a, H_b\} = H_{[a,b]}$.

Let us consider a special case of this theorem. Consider a Lagrangian mechanical system, with the configuration space M and a non-degenerate Lagrangian $L(\boldsymbol{q}, \dot{\boldsymbol{q}})$. As discussed in the previous chapter, in this case we can use Legendre transform to define a Hamiltonian

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

on $\mathscr{M} = T^*M$ (see (4.2)).

Assume now that we have an action of a Lie group G on M which preserves the Lagrangian. In this case, every element $a \in \mathfrak{g}$ defines an integral of motion I_a of the Lagrangian system (see Corollary 2.6). On the other hand, by Example 5.1, action of G on M lifts to a Hamiltonian action of G on $\mathscr{M} = T^*M$, so every element $a \in \mathfrak{g}$ also defines $H_a \in C^\infty(\mathscr{M})$. The following proposition shows that these two constructions coincide up to a sign.

Proposition 5.5.

(1) *Given an action of G on M which preserves the Lagrangian $L(\boldsymbol{q}, \dot{\boldsymbol{q}})$, the corresponding action of G on $\mathscr{M} = T^*M$ defined in Example 5.1 is Hamiltonian and preserves H.*

(2) *For every $a \in \mathfrak{g}$, the Hamiltonian $H_a \in C^\infty(\mathscr{M})$, defined in Theorem 5.4, after Legendre transform coincides with the integral of motion I_a in the Lagrangian mechanics, defined in Corollary 2.6 up to a sign:*

$$H_a(\boldsymbol{p}, \boldsymbol{q}) = -I_a(\boldsymbol{q}, \dot{\boldsymbol{q}})\big|_{\boldsymbol{p}=\frac{\partial L}{\partial \dot{\boldsymbol{q}}}}.$$

The proof is obvious since by Theorem 2.5 $I_a = \dfrac{\partial L}{\partial \dot{\boldsymbol{q}}}\,\xi_a(\boldsymbol{q})$, and according to Example 5.1 we have $H_a = -\boldsymbol{p}\cdot\xi_a(\boldsymbol{q})$. This also explains the origin of the minus sign: it appears because of the minus sign in the map $\mathfrak{g} \to \mathrm{Vect}(M)\colon a \mapsto -\xi_a$, which in turn was necessary to make this map a morphism of Lie algebras.

Example 5.2. Consider the action of $\mathbb{R}^n$ on itself by translations: $\boldsymbol{r} \mapsto \boldsymbol{r} + a$. In this case, the corresponding vector field is given by $\xi_a = \sum_{i=1}^n a^i \dfrac{\partial}{\partial x^i}$. By Example 5.1, the corresponding action on $T^*\mathbb{R}^n$ is Hamiltonian, with the Hamiltonian functions

$$H_a(\boldsymbol{p}, \boldsymbol{r}) = -\sum_{i=1}^n a^i p_i.$$

This also can be verified directly using (5.2), since in this case

$$\xi_a(f) = \{f, H_a\}.$$

In other words, the integrals of motion corresponding to generators of translation in x^i directions are negative of the corresponding momenta p_i. Compare with Example 2.1, where we derived the same result without the minus sign in the Lagrangian picture.

Example 5.3. Consider the action of $G = \mathrm{SO}(3)$ on $M = \mathbb{R}^3$ and thus on $T^*\mathbb{R}^3$. By Example 5.1, this action is Hamiltonian: for any $a \in \mathfrak{so}(3)$, the corresponding Hamiltonian is $H_a = -\sum a_{ij} p_i q^j$. In particular, if we take $a = J_i$ to be the infinitesimal rotation around i-th coordinate axis (see (1.15)), then the matrix entries of J_I are given by $(J_i)_{jk} = -\epsilon_{ijk}$, so the corresponding vector field is $\xi_i = \epsilon_{ijk} x^j \partial_k$, and the corresponding Hamiltonian is exactly the i-th component of the angular momentum $M = \boldsymbol{r} \times \boldsymbol{p}$ with minus sign:

$$H_{J_i} = -M_i = -\epsilon_{ijk} x^j p_k$$

(compare with Exercise 4.6). Thus, the Poisson brackets of M_1, M_2, M_3 which we found in Exercise 4.6 by explicit computation, immediately follow from the commutation relations for the standard generators of $\mathfrak{so}(3)$ discussed in Section 1.1.

5.3. Galilean group action in $T^*\mathbb{R}^3$

Here we consider the Galilean group $G = \mathrm{E}(3) \ltimes \mathbb{R}^4$, introduced in Section 1.4. It naturally acts on the Newtonian spacetime $\mathbb{R}^4 = \mathbb{R}^3 \times \mathbb{R}$ by formula (1.9), which we rewrite as

$$\mathbb{R}^4 \ni x \mapsto \lambda \cdot x + a \in \mathbb{R}^4, \tag{5.4}$$

where

$$\lambda = \begin{pmatrix} g & \boldsymbol{v} \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad x = \begin{pmatrix} \boldsymbol{r} \\ t \end{pmatrix}, \quad a = \begin{pmatrix} \boldsymbol{r}_0 \\ t_0 \end{pmatrix}.$$

Recall that here $g \in \mathrm{O}(3)$, $\boldsymbol{r}, \boldsymbol{r}_0, \boldsymbol{v} \in \mathbb{R}^3$ and $t, t_0 \in \mathbb{R}$.

Note that G acts on the *spacetime* $\mathbb{R}^3 \times \mathbb{R}$ and not on the space $\mathbb{R}^3$. Yet it turns out that one can define a natural action of G on the phase space $T^*\mathbb{R}^3$. To do so, consider first the set $\mathscr{X}$ of all solutions of equations of motion for the free particle of mass m in $\mathbb{R}^3$; we consider each such solution as (unparametrized) trajectory in spacetime $\mathbb{R}^3 \times \mathbb{R}$. Since the Galilean group acts on $\mathbb{R}^4$ preserving equations of motion of a free particle, this action descends to an action on $\mathscr{X}$; thus, we have a natural action of G on $\mathscr{X}$.

On the other hand, for each $(\boldsymbol{p},\boldsymbol{r}) \in T^*\mathbb{R}^3$, there exists a unique solution satisfying the initial conditions $\boldsymbol{r}(0)=\boldsymbol{r}$, $\left.\frac{\partial L}{\partial \dot{\boldsymbol{r}}}\right|_{t=0} = \boldsymbol{p}$. Thus, we have an identification of the phase space and $\mathscr{X}$; therefore, the action of G on $\mathscr{X}$ automatically gives an action of G on $T^*\mathbb{R}^3$.

It is easy to write this action explicitly in coordinates. Solutions of classical equations of motion for a free particle are straight lines; thus, every solution has the form

$$\boldsymbol{r}(t) = \boldsymbol{r}_0 + t\boldsymbol{v}_0$$

for some $\boldsymbol{r}_0, \boldsymbol{v}_0 \in \mathbb{R}^3$. Using $\boldsymbol{p} = m\boldsymbol{v}_0$, we see that the correspondence between $T^*\mathbb{R}^3$ and $\mathscr{X}$ is given by

$$(\boldsymbol{p},\boldsymbol{r}) \mapsto l = \left\{(\boldsymbol{r} + t\frac{\boldsymbol{p}}{m}, t)\right\}_{t\in\mathbb{R}} \in \mathscr{X}.$$

This gives the following action of the Galilean group G on $T^*\mathbb{R}^3$:

$$(\lambda,a)(\boldsymbol{p},\boldsymbol{r}) = \left(g\cdot\boldsymbol{p} + m\boldsymbol{v}, g\cdot\boldsymbol{r} - t_0 g\cdot\frac{\boldsymbol{p}}{m} - t_0\boldsymbol{v} + \boldsymbol{r}_0\right), \qquad \lambda = \begin{pmatrix} g & \boldsymbol{v} \\ 0 & 1\end{pmatrix}, \quad a = \begin{pmatrix}\boldsymbol{r}_0 \\ t_0\end{pmatrix}. \tag{5.5}$$

It is easy to see that restriction of this action to $\mathrm{SO}(3) \subset G$ coincides with the obvious action induced by the action of $\mathrm{SO}(3)$ on $\mathbb{R}^3$. It is also easy to see that elements of G preserve the Hamiltonian H of a free particle, except for the pure Galilean transformations, which act on $T^*\mathbb{R}^3$ by $(\boldsymbol{p},\boldsymbol{r}) \mapsto (\boldsymbol{p} + m\boldsymbol{v}, \boldsymbol{r})$ and do not preserve H.

It is natural to ask whether this action of G is Hamiltonian. It turns out that it is not quite true; to make it Hamiltonian, we need a central extension of G. Recall that we have defined a central extension $\tilde{G}$ of G in Remark 1.9. The action of $\tilde{G}$ on $T^*\mathbb{R}^3$ is given by the same formula (5.5), where we let the center act trivially.

Theorem 5.6. *The action of the central extension $\tilde{G}$ of the Galilean group on the phase space $T^*\mathbb{R}^3$ is Hamiltonian.*

The Hamiltonian functions corresponding to generators J_i, K_i, P_i of the Lie algebra of the Galilean group (see Section 1.5) and the central element z are given by

$$\begin{aligned}
\hat{J}_i &= -M_i = -\epsilon_{ijk}x^j p_k,\\
\hat{K}_i &= mx^i,\\
\hat{P}_0 &= H = \frac{\boldsymbol{p}^2}{2m},\\
\hat{P}_i &= -p_i, \quad i = 1,2,3,\\
\hat{z} &= m.
\end{aligned}$$

The Poisson brackets of these functions are given by

$$\{\hat{P}_i, \hat{P}_j\} = \{\hat{P}_i, \hat{P}_0\} = \{\hat{J}_i, \hat{P}_0\} = 0, \quad \{\hat{J}_i, \hat{J}_j\} = \epsilon_{ijk}\hat{J}_k, \tag{5.6}$$

$$\{\hat{K}_i, \hat{K}_j\} = 0, \quad \{\hat{J}_i, \hat{K}_j\} = \epsilon_{ijk}\hat{K}_k, \tag{5.7}$$

$$\{\hat{K}_i, \hat{P}_0\} = \hat{P}_i, \quad \{\hat{K}_i, \hat{P}_j\} = m\delta_{ij}, \quad \{\hat{J}_i, \hat{P}_j\} = \epsilon_{ijk}\hat{P}_k. \tag{5.8}$$

Proof. The result follows from explicit computations, based on (5.5). The Hamiltonian functions for space translations P_i and space rotations J_i have already been computed in Examples 5.2 and 5.3.

For the time translations $\begin{pmatrix} \boldsymbol{r} \\ t \end{pmatrix} \mapsto \begin{pmatrix} \boldsymbol{r} \\ t+t_0 \end{pmatrix}$ the formula (5.5) gives

$$(\boldsymbol{p}, \boldsymbol{r}) \mapsto \left(\boldsymbol{p}, \boldsymbol{r} - t_0 \frac{\boldsymbol{p}}{m}\right),$$

so the corresponding vector field is $\xi = -\frac{\boldsymbol{p}}{m}\frac{\partial}{\partial \boldsymbol{r}}$, and $-\xi$ is generated by the Hamiltonian function $H = \frac{\boldsymbol{p}^2}{2m}$ since

$$\xi(f) = \{\xi, H\}.$$

Finally, pure Galilean transformations act on $\mathbb{R}^4$ by

$$\begin{pmatrix} \boldsymbol{r} \\ t \end{pmatrix} \mapsto \begin{pmatrix} \boldsymbol{r} + t\boldsymbol{v} \\ t \end{pmatrix},$$

and on $T^*\mathbb{R}^3$ they act by $(\boldsymbol{p}, \boldsymbol{r}) \mapsto (\boldsymbol{p} + m\boldsymbol{v}, \boldsymbol{r})$. The corresponding vector field is $\xi_{\boldsymbol{v}} = m\boldsymbol{v}\frac{\partial}{\partial \boldsymbol{p}}$, so $-\xi$ is is generated by the Hamiltonian function $H_{\boldsymbol{v}} = m\boldsymbol{v} \cdot \boldsymbol{r}$ since

$$\xi_{\boldsymbol{v}}(f) = \{f, H_{\boldsymbol{v}}\}.$$

Taking $\boldsymbol{v}$ to be one of the basis vectors in $\mathbb{R}^3$, we get the Hamiltonian functions $\hat{K}_i = mx^i$.

For the central element, the corresponding vector field is zero, so we can take any constant function as a Hamiltonian.

The Poisson brackets of the Hamiltonian functions are obtained by an elementary computation, which is left to the reader (see Exercise 5.3). We emphasize that it is the relation $\{\hat{K}_i, \hat{P}_j\} = m\delta_{ij}$ which warrants the central extension of the Lie algebra $\mathfrak{g}$ of the Galilean group. Namely, in $\mathfrak{g}$ we have $[K_i, P_j] = 0$, whereas in the central extension $\tilde{\mathfrak{g}}$ we have $[K_i, P_j] = \delta_{ij} z$. Thus, setting $\hat{z} = m$, we see that Poisson brackets (5.6)–(5.8) exactly match the Lie brackets of $\tilde{\mathfrak{g}}$. □

Remark 5.7. Note that we also identify $\mathscr{X}$ with $T^*\mathbb{R}^3$ by setting $\boldsymbol{r}(t_0) = \boldsymbol{r}$, $\left.\frac{\partial L}{\partial \dot{\boldsymbol{r}}}\right|_{t=t_0} = \boldsymbol{p}$ for some fixed t_0 (not to be confused with the arbitrary time translation t_0 in in Galilean group!). Then corresponding Hamiltonian functions will be the same, except for those associated with the pure Galilean transformations. Namely, we obtain

$$\hat{K}_i = mx^i + t_0 \hat{P}_i,$$

and the Poisson brackets (5.7)–(5.8) will be the same. Since $\{H, \hat{K}_i\} = -\hat{P}_i$, the observables $\hat{K}_i$ are not integrals of motion, while components of the vector $\boldsymbol{I} = t\boldsymbol{P} - m\boldsymbol{r}$ in Example 2.3 (note that $\hat{\boldsymbol{P}} = -\boldsymbol{P}$), considered as functions on $T^*\mathbb{R}^3 \times \mathbb{R}$, are obviously conserved.

5.4. Hamiltonian reduction

If we have a Hamiltonian system with a symmetry group G, one can ask if it is possible to take the quotient by the action of this group, considering states which can be obtained from each other by action of G to be physically equivalent. Common examples of this situation is when the group G acts by changes of a basis, or changes a trivialization of some vector bundle. The proper mathematical formalism for such constructions is known as *Hamiltonian reduction*.

We begin with some differential geometry preliminaries. Recall that an action of a Lie group G on a manifold M is called *proper* if the map

$$M \times G \to M \times M$$
$$(m, g) \mapsto (m, g \cdot m)$$

is proper. This is equivalent to the requirement that for every compact subsets $K_1, K_2 \subset M$, the set

$$\{g \in G \mid gK_1 \cap K_2 \neq \varnothing\} \subset G$$

is compact.

It is easy to show that if G is compact, then its action is automatically proper.

Properness condition guarantees that we have a good local model of the action. In particular, for a proper action one can show that any G-orbit $G \cdot x \subset M$ is a closed smooth submanifold of M of dimension $\dim G - \dim G_x$, where G_x is the stabilizer of x. Moreover, if in addition the dimensions of stabilizers $\dim G_x$ are constant on some open subset $U \subset M$, then the quotient space U/G has a canonical structure of a smooth manifold of dimension $\dim M - (\dim G - \dim G_x)$. In particular, if the action is free and proper, so that $\dim G_x = 0$, then M/G is a manifold of dimension $\dim M - \dim G$. (Note that properness condition is required: just having a free action of G is not enough).

Let us now consider group actions on symplectic manifolds. In this case, one can not hope to get a symplectic structure on M/G (for example, if G-action is free and $\dim G$ is odd, then $\dim M/G = \dim M - \dim G$ is also odd). The correct construction makes use of the moment map.

Recall that the action of a Lie group G on a symplectic manifold $\mathscr{M}$ is called Hamiltonian if we have a moment map $\mu\colon \mathscr{M} \to \mathfrak{g}^*$ which, among other things, is required to be G-equivariant. This implies that for every coadjoint orbit $\mathcal{O} \subset \mathfrak{g}^*$, its preimage $\mu^{-1}(\mathcal{O})$ is a G-invariant subset of $\mathscr{M}$ (see Exercise 4.9 for the discussion of coadjoint orbits). We will denote by $\mathcal{O}_u$ the orbit of $u \in \mathfrak{g}^*$.

Theorem 5.8. *Let $\mathscr{M}$ be a symplectic manifold with a proper Hamiltonian action of a real Lie group G, and let $\mu\colon \mathscr{M} \to \mathfrak{g}^*$ be the corresponding moment map. Let $u \in \mathfrak{g}^*$ be such that the following conditions hold:*

(1) *The point $u \in \mathfrak{g}^*$ is a regular value of μ, so $\mu^{-1}(u) \subset \mathscr{M}$ is a submanifold.*
(2) *The action of the stabilizer $G_u \subset G$ of u (under the coadjoint action of G on $\mathfrak{g}^*$) on $\mu^{-1}(u)$ is free and proper, so that $\mu^{-1}(u)/G_u$ is a smooth manifold.*

Then $\mu^{-1}(u)/G_u = \mu^{-1}(\mathcal{O}_u)/G$ has a canonical structure of a symplectic manifold, inherited from $\mathscr{M}$. This manifold is called the reduced phase space.

Example 5.4. Let M be a C^∞ manifold with a free proper action of a Lie group G; let $\mu\colon T^*M \to \mathfrak{g}^*$ be the corresponding moment map (see Example 5.1). By Theorem 5.8, $\mu^{-1}(0)/G$ has a canonical structure of a symplectic manifold, inherited from T^*M. Then we have a symplectomorphism

$$T^*(M/G) \simeq \mu^{-1}(0)/G.$$

5.5. Exercises

Exercise 5.1. Consider the two-dimensional sphere $S^2 \subset \mathbb{R}^3$. Let ω be the volume form on S^2 induced by the metric in $\mathbb{R}^3$; this defines on S^2 structure of a symplectic manifold. Let $G = S^1$ act on S^2 by rotations around z-axis. Show that this action is Hamiltonian, and the corresponding moment map is given by $\mu(x, y, z) = z$.

Exercise 5.2.

(1) Let $V = \mathbb{R}^2$ be the two-dimensional symplectic vector space with coordinates p, q and let $\omega = dp \wedge dq$. Let $G = \mathrm{SL}(2, \mathbb{R})$ acting on V in the obvious way. Show that this action is Hamiltonian and write the corresponding moment map.

(2) Consider the bilinear form on $\mathfrak{g} = \mathfrak{gl}(2, \mathbb{R})$ given by $(a, b) = \mathrm{tr}(ab)$. Show that this form is non-degenerate and thus gives an isomorphism $\mathfrak{g}^* \cong \mathfrak{g}$; therefore, this allows us to consider the moment map of the previous part as a map $V \to \mathfrak{g}$. Show that then, for any $v \in V$, $\mu(v)$ is nilpotent.

Exercise 5.3. Prove formulas (5.6)–(5.8).

Exercise 5.4. Consider the space $V = \mathrm{End}(\mathbb{C}^n)$ of complex $n \times n$ matrices. Consider the action of the group $G = \mathrm{U}(n)$ on V given by conjugation: $g \cdot A = gAg^{-1}$. Show that this action is Hamiltonian, with the moment map given by

$$\mu(A) = \frac{i}{2}[A, A^*]$$

where A^* is conjugate transpose of A, and as in Exercise 5.2, we identify $\mathfrak{g} \cong \mathfrak{g}^*$ using the bilinear form $(a, b) = \mathrm{tr}(ab)$.

Exercise 5.5.

(1) In the assumptions of the previous problem, show that $\mu^{-1}(0)$ consists exactly of those matrices A that can be diagonalized by a unitary change of basis. Deduce from it that $\mu^{-1}(0)/\mathrm{U}(n) = \mathbb{C}^n/S_n$.

(2) Let $X = \mu^{-1}(0)/\mathrm{U}(n) = \mathbb{C}^n/S_n$; it is not a smooth manifold, but it contains an open dense set $X^0 \subset X$ consisting of conjugacy classes of diagonalizable matrices with distinct eigenvalues. By Theorem 5.8, X^0 has a canonical symplectic structure. Describe this symplectic structure explicitly in terms of coordinates $x_i = \mathrm{Re}(\lambda_i)$, $y_i = \mathrm{Im}(\lambda_i)$.

Chapter 6

Hamiltonian Systems with Constraints

The construction of a reduced phase space by Hamiltonian reduction, discussed int he previous chapter, is a special case of a more general procedure, developed by Dirac for Lagrangian and Hamiltonian systems with constraints. As we will see later, such systems naturally occur in gauge theories and in the theory of gravity.

6.1. First order Lagrangian formalism

Definition 6.1. A mechanical system in first order Lagrangian formalism is a pair $(\mathcal{M}, \mathcal{L})$, where $\mathcal{M}$ is a smooth manifold and a $\mathcal{L} \in C^\infty(T\mathcal{M})$ is the Lagrangian function which is linear in generalized velocities: if ξ^α, $\alpha = 1, \dots, N$ are local coordinates in $\mathcal{M}$ and $(\xi^\alpha, \dot{\xi}^\alpha)$ are corresponding coordinates in $T\mathcal{M}$, then $\mathcal{L}$ has the form

$$\mathcal{L}(\boldsymbol{\xi}, \dot{\boldsymbol{\xi}}) = \sum_{\alpha=1}^{N} f_\alpha(\boldsymbol{\xi})\dot{\xi}^\alpha - H(\boldsymbol{\xi}), \qquad N = \dim \mathcal{M}. \tag{6.1}$$

This is a special case of Lagrangian formalism developed in Chapter 1. In particular, for a path $\gamma\colon [t_0, t_1] \to \mathcal{M}$ one can define the action

$$S(\gamma) = \int_{t_0}^{t_1} \mathcal{L}(\gamma'(t))\, dt = \int_{\hat{\gamma}} \big(\sum f_\alpha(\boldsymbol{\xi}) d\xi^\alpha - H(\boldsymbol{\xi}) dt\big) \tag{6.2}$$

where γ' is the lifting of γ to $T\mathcal{M}$, and $\hat{\gamma} = \{(\gamma(t), t)\} \subset \mathcal{M} \times \mathbb{R}$. We can then study extremals of this action, which are defined by Euler–Lagrange equations. Note, however, that Lagrangian (6.1) can not possibly be non-degenerate in the sense of Definition 1.7, so existence and uniqueness of solutions, even locally, is not guaranteed.

We will denote by $\vartheta_\mathcal{L}$ the one-form on $\mathcal{M} \times \mathbb{R}$ which appears in the formula for action:

$$\vartheta_\mathcal{L} = \sum f_\alpha(\boldsymbol{\xi}) d\xi^\alpha - H(\boldsymbol{\xi}) dt. \tag{6.3}$$

This form is obviously invariant under translations in t direction. Conversely, it is obvious that any 1-from on $\mathcal{M} \times \mathbb{R}$ which is invariant under time translations has the form above, and thus can be described as $\vartheta_{\mathcal{L}}$ for some first order Lagrangian $\mathcal{L}$. Thus, a first order Lagrangian on $\mathcal{M}$ is effectively the same as a time-independent 1-form on $\mathcal{M} \times \mathbb{R}$.

Note that it is obvious from (6.2) that adding a full derivative to $\vartheta_{\mathcal{L}}$ doesn't change equations of motion: if first order Lagrangians $\mathcal{L}$, $\mathcal{L}'$ are related by $\vartheta_{\mathcal{L}'} = \vartheta_{\mathcal{L}} + dS$ for some function $S \in C^\infty(\mathcal{M})$, then the Euler–Lagrange equations for these Lagrangians have the same solutions.

Even though technically first order Lagrangian systems are special cases of Lagrangian systems described in Chapter 1, their physical interpretation is quite different. Namely, as the examples below show, in this case the manifold $\mathcal{M}$ plays the role of the phase space (and not the configuration space!) of the system.

Example 6.1. Let (M, L) be a Lagrangian system with the Lagrangian L being non-degenerate as defined in Definition 1.7. Let $\mathcal{M} = TM$ and define the Lagrangian $\mathcal{L}$ on $\mathcal{M}$ by

$$\mathcal{L}(\boldsymbol{q}, \boldsymbol{v}, \dot{\boldsymbol{q}}, \dot{\boldsymbol{v}}) = \sum_{i=1}^{n} \frac{\partial L}{\partial v^i}(\dot{q}^i - v^i) + L(\boldsymbol{q}, \boldsymbol{v}) = \frac{\partial L}{\partial \boldsymbol{v}}(\dot{\boldsymbol{q}} - \boldsymbol{v}) + L(\boldsymbol{q}, \boldsymbol{v}), \quad n = \dim M. \tag{6.4}$$

Then Euler–Lagrange equations for system $(\mathcal{M}, \mathcal{L})$ are equivalent to Euler–Lagrange equations for (M, L).

Indeed, the Euler–Lagrange equations for $(\mathcal{M}, \mathcal{L})$ are

$$\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{q}}} - \frac{\partial \mathcal{L}}{\partial \boldsymbol{q}} = 0 \quad \text{and} \quad \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{v}}} - \frac{\partial \mathcal{L}}{\partial \boldsymbol{v}} = 0. \tag{6.5}$$

Since $\frac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{v}}} = 0$, the second equation in (6.5) gives

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

Since Lagrangian L is non-degenerate, this is equivalent to

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

Using this relation, we can rewrite the first equation in (6.5) as

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

which are exactly the Euler-Lagrange equations for the Lagrangian system (M, L).

Note that Euler–Lagrange equations for (M, L) are a system of 2nd order differential equations in n variables q^i (see (1.7)), while Euler–Lagrange equations for $(\mathcal{M}, \mathcal{L})$ is a system of first order differential equations in $2n$ variables q^i, v^i. Thus, this example is a special case of a familiar phenomenon in the theory of differential equations: a system of second order differential equations in n variables can be replaced by a system of first order differential equations in $2n$ variables.

Example 6.2. As in Chapter 4, consider a Hamitonian mechanical system, described by a symplectic manifold $(\mathcal{M}, \omega)$ and the Hamitlonian function $H \in C^\infty(\mathcal{M})$.

By Darboux theorem, we can choose local coordinates $\boldsymbol{p}, \boldsymbol{q}$ in $\mathcal{M}$ such that the symplectic form is given by

$$\omega = d\boldsymbol{p} \wedge d\boldsymbol{q}.$$

Define (locally) the first order Lagrangian on $\mathcal{M}$ by

$$\mathcal{L} = \boldsymbol{p}\dot{\boldsymbol{q}} - H(\boldsymbol{p}, \boldsymbol{q}) \tag{6.6}$$

(compare with Definition 4.4).

Then it is easy to see that the Euler–Lagrange equations for $(\mathcal{M}, \mathcal{L})$ are exactly the canonical equations of Hamiltonian mechanics (see Theorem 4.5):

$$\dot{\boldsymbol{q}} = \frac{\partial H}{\partial \boldsymbol{p}} \quad \text{and} \quad \dot{\boldsymbol{p}} = -\frac{\partial H}{\partial \boldsymbol{q}}. \tag{6.7}$$

Thus, in this case the Hamiltonian system $(\mathcal{M}, \omega, H)$ can also be described (locally) by the first order Lagrangian formalism.

Thse two examples are closely related. Recall that given a non-degenerate Lagrangian system (M, L), we can define Legendre transform $\tau_L \colon TM \to T^*M$. We leave it to the reader to check that in this case, the Legendre transform identifies the first order Lagrangian (6.4) on TM with the first order Lagrangian (6.6) on T^*M.

Motivated by the last example, we give the following definition.

Definition 6.2. Lagrangian $\mathcal{L}$ given by (6.1) is called non-degenerate if the 2-form

$$\omega_{\mathcal{L}} = d\left(\sum_{\alpha=1}^{N} f_\alpha(\boldsymbol{\xi}) d\xi^\alpha\right) = \sum_{\alpha,\beta=1}^{N} \frac{\partial f_\beta}{\partial \xi^\alpha}(\boldsymbol{\xi}) d\xi^\alpha \wedge d\xi^\beta \tag{6.8}$$

is non-degenerate on $\mathcal{M}$.

In particular, it is easy to see that the Lagrangians (6.4), (6.6) in examples above are non-degenerate.

If the Lagrangian $\mathcal{L}$ is non-degenerate, it follows from Darboux theorem that $N = 2n$ is even and there exist local canonical coordinates $(\boldsymbol{p}, \boldsymbol{q}) = (p_1, \ldots, p_n, q^1, \ldots, q^n)$ on $\mathcal{M}$ such that

$$\omega = d\boldsymbol{p} \wedge d\boldsymbol{q} \quad \text{and} \quad \vartheta_{\mathcal{L}} = \boldsymbol{p} d\boldsymbol{q} - H(\boldsymbol{p}, \boldsymbol{q}) dt + dS$$

for some function S on $\mathcal{M}$. Since adding dS doesn't change the equations of motion, without loss of generality we can assume that in this case

$$\mathcal{L} = \boldsymbol{p}\dot{\boldsymbol{q}} - H(\boldsymbol{p}, \boldsymbol{q}) \tag{6.9}$$

and therefore, we are exactly in the situation of Example 6.2.

6.2. Singular Lagrangians

Let us now consider the more general case of first order Lagrangian system where we do not assume that the 2-form $\omega_{\mathcal{L}}$ defined by (6.8) is non-degenerate. For reasons that will be clear shortly, let us change the notation and use letter $\widetilde{\mathcal{M}}$ to denote the manifold; thus, we have an N-dimensional manifold $\widetilde{\mathcal{M}}$ and a first order Lagrangian $\mathcal{L} \in C^\infty(T\widetilde{\mathcal{M}})$.

Assume for simplicity that $\omega_{\mathcal{L}}$ has constant rank $\operatorname{rk}\omega_{\mathcal{L}} = 2n$ in a neighborhood of some point p of $\widetilde{\mathcal{M}}$. Then by Darboux theorem, one can choose local coordinates $(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{\lambda}) = (p_1, \dots, p_n, q^1, \dots, q^n, \lambda_1, \dots, \lambda_m)$, $2n + m = \dim \widetilde{\mathcal{M}}$, such that

$$\begin{aligned} \omega_{\mathcal{L}} &= d\boldsymbol{p} \wedge d\boldsymbol{q} = \sum dp_i \wedge dq^i \\ \vartheta_{\mathcal{L}} &= \boldsymbol{p} \wedge d\boldsymbol{q} - H(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{\lambda})dt + dS \end{aligned} \tag{6.10}$$

for some function S on $\widetilde{\mathcal{M}}$. Since adding dS doesn't change the equations of motion, without loss of generality we can assume that

$$\begin{aligned} \vartheta_{\mathcal{L}} &= \boldsymbol{p} \wedge d\boldsymbol{q} + H(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{\lambda})dt \\ \mathcal{L} &= \boldsymbol{p}\dot{\boldsymbol{q}} - H(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{\lambda}). \end{aligned} \tag{6.11}$$

In this case, the Euler-Lagrange equations for the Lagrangian $\mathcal{L}$ have the following form

$$\begin{aligned} \dot{\boldsymbol{p}} &= -\frac{\partial H}{\partial \boldsymbol{q}} \\ \dot{\boldsymbol{q}} &= \frac{\partial H}{\partial \boldsymbol{p}} \\ \frac{\partial H}{\partial \boldsymbol{\lambda}} &= 0. \end{aligned} \tag{6.12}$$

The last equation in (6.12) shows that solutions of equations of motion must lie in the subset

$$\widetilde{\mathcal{M}}_0 = \left\{ (\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{\lambda}) \in \widetilde{\mathcal{M}} \mid \frac{\partial H}{\partial \boldsymbol{\lambda}} = 0 \right\}$$

If the $m \times m$ matrix $\left\{ \frac{\partial^2 H}{\partial \lambda_a \partial \lambda_b} \right\}_{a,b=1}^{m}$ is non-degenerate, it follows from the implicit function theorem that $\widetilde{\mathcal{M}}_0$ is a submanifold in $\widetilde{\mathcal{M}}$ of dimension $N - m = 2n$, and $\boldsymbol{p}, \boldsymbol{q}$ are local coordinates on $\widetilde{\mathcal{M}}_0$: one can express $\boldsymbol{\lambda}$ in terms of $\boldsymbol{p}, \boldsymbol{q}$. Thus, the original system is equivalent to the first order Lagrangian system on $\widetilde{\mathcal{M}}_0$ defined by

$$\tilde{H}(\boldsymbol{p}, \boldsymbol{q}) = H(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{\lambda}(\boldsymbol{p}, \boldsymbol{q})) \quad \text{and} \quad \tilde{\mathcal{L}} = \boldsymbol{p}\dot{\boldsymbol{q}} - \tilde{H}(\boldsymbol{p}, \boldsymbol{q}),$$

The latter system is a non-degenerate Lagrangian system, so in this case we are back to the situation discussed in the previous section.

Let us therefore consider the case when the matrix $\left\{ \frac{\partial^2 H}{\partial \lambda_a \partial \lambda_b} \right\}_{a,b=1}^{m}$ is degenerate. For simplicity, let us consider the extreme case, when this matrix is zero in a neighborhood of a point in $\widetilde{\mathcal{M}}_0$ (it can be shown using same ideas as above that the general case can be reduced to this one). This means that for fixed $\boldsymbol{p}, \boldsymbol{q}$, H is a linear function of $\boldsymbol{\lambda}$, so $\mathcal{L}$ has the form

$$\mathcal{L} = \boldsymbol{p}\dot{\boldsymbol{q}} - H(\boldsymbol{p}, \boldsymbol{q}) - \sum_{a=1}^{m} \lambda_a \varphi^a(\boldsymbol{p}, \boldsymbol{q}). \tag{6.13}$$

In this case the Euler–Lagrange equations (6.12) become

$$\dot{\boldsymbol{p}} = -\frac{\partial H}{\partial \boldsymbol{q}} - \sum_{a=1}^{m} \lambda_a \frac{\partial \varphi^a}{\partial \boldsymbol{q}} = \{H + \sum \lambda_a \varphi^a, \boldsymbol{p}\}, \tag{6.14}$$

$$\dot{\boldsymbol{q}} = \frac{\partial H}{\partial \boldsymbol{p}} + \sum_{a=1}^{m} \lambda_a \frac{\partial \varphi^a}{\partial \boldsymbol{p}} = \{H + \sum \lambda_a \varphi^a, \boldsymbol{q}\}, \tag{6.15}$$

$$\varphi^a(\boldsymbol{p}, \boldsymbol{q}) = 0, \quad a = 1, \dots, m. \tag{6.16}$$

where $\{,\}$ is the Poisson bracket defined by the symplectic form $d\boldsymbol{p} \wedge d\boldsymbol{q}$.

Note that there are no equations for the time evolution of $\boldsymbol{\lambda}$; instead, $\boldsymbol{\lambda}$ plays the role of Lagrange multipliers, used in multivariable calculus for finding conditional extrema of functions. It is convenient to write

$$\widetilde{\mathcal{M}} = \mathcal{M} \times \mathbb{R}^m$$

(locally), where $\mathcal{M}$ is a $2n$-dimensional manifold with coordinates $\boldsymbol{p}, \boldsymbol{q}$, and $\boldsymbol{\lambda}$ are coordinates in $\mathbb{R}^m$, and consider equations (6.14)–(6.16) as describing a trajectory of a point in $\mathcal{M}$ which depends on additional parameters $\boldsymbol{\lambda}$.

The functions $\varphi^a(\boldsymbol{p}, \boldsymbol{q})$ are called *constraints*, and equations (6.16) determine a subset

$$\mathcal{M}_0 = \{(\boldsymbol{p}, \boldsymbol{q}) \mid \varphi^\alpha(\boldsymbol{p}, \boldsymbol{q}) = 0\} \subset \mathcal{M}. \tag{6.17}$$

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 $2n - m$. Thus, adding term $\sum \lambda_a \varphi^a(\boldsymbol{p}, \boldsymbol{q})$ to the first order Lagrangian effectively constrains the solutions to the submanifold $\mathcal{M}_0 \subset \mathcal{M}$.

To write down the corresponding Lagrangian system on $\mathcal{M}_0$, one needs to restrict the 1-form $\vartheta_{\mathcal{L}}$ on $\mathcal{M}_0 \times \mathbb{R}$ and use Darboux theorem to obtain corresponding coordinates $(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{\lambda})$. If the result is a non-degenerate Lagrangian system, than it corresponds to the Hamiltonian system and we are done. If it results in a singular Lagrangian, we need to repeat the above construction and obtain a submanifold $\mathcal{M}_1$ of $\mathcal{M}_0$, and so on. To apply Darboux theorem at each step of this procedure one needs to solve the constraints (6.16), which could be a very difficult problem. However, there is important class of constraints for which one does not need to solve equations (6.16).

6.3. First class constraints and reduced phase space

As in the previous section, consider the Lagrangian (6.13) on $\widetilde{\mathcal{M}} = \mathcal{M} \times \mathbb{R}^m$ and assume that the matrix $\left(\dfrac{\partial \varphi^a}{\partial p_i}, \dfrac{\partial \varphi^a}{\partial q^i}\right)$ has rank m. We have shown that solutions of equations of motion must be constrained to the submanifold $\mathcal{M}_0 \subset \mathcal{M}$, given by equations $\varphi^a(\boldsymbol{p}, \boldsymbol{q}) = 0$. To simplify further, we will assume that the inclusion map $i : \mathcal{M}_0 \to \mathcal{M}$ is a proper embedding.

This naturally leads to the following questions.

- Can one restrict to the manifold $\mathcal{M}_0$ equations (6.14)–(6.16)? In other words, do all trajectories $(\boldsymbol{p}(t), \boldsymbol{q}(t))$ of the Hamilton's equations (6.14)–(6.15) lie on $\mathcal{M}_0$ if $(\boldsymbol{p}(0), \boldsymbol{q}(0)) \in \mathcal{M}_0$?

- Can we describe the algebra of observables whose evolution does not depend on the arbitrary parameters $\lambda_1, \dots, \lambda_m$ in (6.14)–(6.15)?

Let $\mathscr{A} = C^\infty(\mathcal{M})$ be the algebra of observables of our system, and let $\mathscr{A}_0 = C^\infty(\mathcal{M}_0)$. Since $\mathcal{M}_0 \subset \mathcal{M}$ is a proper embedding, we have

$$\mathscr{A}_0 = \mathscr{A}/\mathcal{I}$$

where $\mathcal{I} \subset \mathscr{A}$ is the structure ideal of $\mathcal{M}_0$, the ideal consisting of functions on $\mathcal{M}$ that vanish on $\mathcal{M}_0$. For every function $f \in \mathscr{A}_0$ we have

$$f = \tilde{f}|_{\mathcal{M}_0},$$

where $\tilde{f} \in \mathscr{A}$ is defined $\bmod \mathcal{I}$.

Condition that the matrix $\left(\frac{\partial \varphi^a}{\partial p_i}, \frac{\partial \varphi^a}{\partial q^i}\right)$ has constant rank m on $\mathcal{M}_0$ leads to the following result.

Lemma 6.3. *Ideal $\mathcal{I}$ is generated by the constraints $\varphi^1, \dots, \varphi^m$.*

It is remarkable that, as was discovered by Dirac, an affirmative answer to both questions asked in the beginning of this section is provided by the following definition. Let $\{\ ,\ \}$ be the Poisson bracket on $\mathcal{M}$ associated with the symplectic form $d\boldsymbol{p} \wedge d\boldsymbol{q}$ (see Definition 4.13).

Definition 6.4. Constraints $\varphi^1, \dots, \varphi^m$ for the singular Lagrangian (6.13) are called *first class constraints* if $\{\varphi^a, \varphi^b\}, \{H, \varphi^a\} \in \mathcal{I}$ for all $a, b = 1, \dots, m$.

In other words, there are smooth functions g_c^{ab} and h_b^a on $\mathcal{M}$ such that

$$\{\varphi^a, \varphi^b\} = \sum_{c=1}^{m} g_c^{ab} \varphi^c \quad \text{and} \quad \{H, \varphi^a\} = \sum_{b=1}^{m} h_b^a \varphi^b. \tag{6.18}$$

Throughout the remainder of this section, we assume that $\varphi^a(\boldsymbol{p}, \boldsymbol{q})$ are the first class constraints.

Lemma 6.5. *For the first class constraints, trajectories $(\boldsymbol{p}(t), \boldsymbol{q}(t))$ of the Hamilton's equations (6.14)–(6.15) lie on $\mathcal{M}_0$ if $(\boldsymbol{p}(0), \boldsymbol{q}(0)) \in \mathcal{M}_0$.*

Proof. It follows from (6.14)–(6.15) that

$$\dot{\varphi}^a = \{H, \varphi^a\} + \sum_{b=1}^{m} \lambda_b \{\varphi^b, \varphi^a\},$$

and it follows from (6.18) that $\dot{\varphi}^a = 0$ on $\mathcal{M}_0$. Thus $\varphi^a(\boldsymbol{p}(t), \boldsymbol{q}(t)) = \varphi^a(\boldsymbol{p}(0), \boldsymbol{q}(0)) = 0$, $a = 1, \dots, m$. □

In general, according to (6.14)–(6.15), the evolution of arbitrary $f \in \mathscr{A}$ is given by

$$\dot{f} = \{H, f\} + \sum_{a=1}^{m} \lambda_a \{\varphi^a, f\}, \tag{6.19}$$

and it follows from (6.18) that restriction of this equation to $\mathcal{M}_0$ does not depend on the choice of a representative of $f \bmod \mathcal{I}$ and defines the evolution in the algebra $\mathscr{A}_0 = \mathscr{A}/\mathcal{I}$. Still, this evolution depend on the choice of parameters $\lambda_1, \dots, \lambda_m$.

Definition 6.6. Admissible observables are functions f on $\mathcal{M}_0$ whose extensions $\tilde{f}$ to $\mathcal{M}$ satisfy

(6.20) $$\{\tilde{f}, \varphi^a\}|_{\mathcal{M}_0} = 0, \quad a = 1, \ldots, m.$$

We will denote the space of admissible observables by $\mathscr{A}_0^*$:

(6.21) $$\mathscr{A}_0^* = \{f \in \mathscr{A}_0 \mid \{\tilde{f}, \varphi^a\}|_{\mathcal{M}_0} = 0, \quad a = 1, \ldots, m\}.$$

In particular, $H|_{\mathcal{M}_0}$ is an admissible observable. It follows from Lemma 6.3 and (6.18) that condition (6.20) does not depend on the choice of extension $\tilde{f}$. The Poisson bracket of admissible observables is defined by

$$\{f, g\}_0 = \{\tilde{f}, \tilde{g}\}\Big|_{\mathcal{M}_0},$$

where $\tilde{f}, \tilde{g}$ are extensions of f, g respectively. It follows from (6.20) that $\mathscr{A}_0^*$ is a Poisson algebra. For admissible observables equation (6.19) takes the form

(6.22) $$\dot{f} = \{H|_{\mathcal{M}_0}, f\}_0$$

and no longer depends on on the choice of arbitrary parameters $\lambda_1, \ldots, \lambda_m$.

According to the above definition, the algebra of admissible observables $\mathscr{A}_0^*$ is a subalgebra in the quotient $\mathscr{A}/\mathcal{I}$. Alternatively, $\mathscr{A}_0^*$ can also be defined as a quotient of a subalgebra in $\mathscr{A}$. Namely, let

$$\mathscr{A}^* = \{f \in \mathscr{A} \mid \ \{f, \varphi^a\}|_{\mathcal{M}_0} = 0, \quad a = 1, \ldots, m\}.$$

Lemma 6.7. *$\mathscr{A}^*$ is a Poisson subalgebra of $\mathscr{A}$: if $f, g \in \mathscr{A}^*$, then $fg \in \mathscr{A}^*$ and $\{f, g\} \in \mathscr{A}^*$. Moreover, $\mathcal{I} \subset \mathscr{A}^*$ is a Poisson algebra ideal of $\mathscr{A}^*$ and*

$$\mathscr{A}^*/\mathcal{I} \simeq \mathscr{A}_0^*.$$

Proof. Follows from Lemma 6.3, equations (6.18) and the Jacobi identity. □

In other words, we have the following commutative diagram, where columns are algebra embeddings and horizontal lines are surjective algebra morphisms (restricting to $\mathcal{M}_0$).

(6.23) $$\begin{array}{ccc} \mathscr{A} & \longrightarrow & \mathscr{A}_0 \\ \cup & & \cup \\ \mathscr{A}^* & \longrightarrow & \mathscr{A}_0^* \end{array}$$

The functions $f \in \mathscr{A}_0^*$ depend on $2n - m - m = 2(n - m)$ parameters and in many cases can be thought of as functions on the *reduced phase space* — symplectic manifold $\mathcal{M}_0^*$ 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 part (4) of Theorem 4.14 in Chapter 4,

(6.24) $$[X_{\varphi^a}, X_{\varphi^b}] = X_{\{\varphi^a, \varphi^b\}}, \quad a, b = 1, \ldots, m.$$

We also have

$$\omega(X_{\varphi^a}, X_{\varphi^b}) = \{\varphi^a, \varphi^b\},$$

so that

$$\omega(X_{\varphi^a}, X_{\varphi^b})|_{\mathcal{M}_0} = 0.$$

Denote by Y_a the vector fields X_{φ^a} restricted to $\mathcal{M}_0$. It follows from (6.18) 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 (6.18) and (6.24) that

$$[X_{\varphi^a}, X_{\varphi^b}] = \sum_{c=1}^{m} g_c^{ab} X_{\varphi^c}.$$

This means that the vector fields $Y_1, \dots, Y_m$ generate 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 Frobenius theorem, $\mathcal{M}_0$ is a foliation with m-dimensional leaves given by the integral manifolds of the distribution $\mathcal{P}$.

Theorem 6.8. *Assume that the foliation $\mathcal{P}$ defined above is a fibration: there is a smooth manifold $\mathcal{M}_0^*$ and a map $\pi\colon \mathcal{M}_0 \to \mathcal{M}_0^*$ such that π is a fiber bundle and fibers of π are exactly the integral manifolds of $\mathcal{P}$.*

Then $\mathcal{M}_0^*$ is a manifold of dimension $2n-2m$ which has a closed non-degenerate 2-form ω^* such that the 2-form $\omega|_{\mathcal{M}_0}$ defined above is the pullback of ω^*: $\omega|_{\mathcal{M}_0} = \pi^*(\omega^*)$.

Moreover, $C^\infty(\mathcal{M}_0^*) = \mathscr{A}_0^*$ is the Poisson algebra defined above.

In other words, $(\mathcal{M}_0^*, \omega^*)$ is a symplectic manifold, the reduced phase space. Though not every foliation is a fibration, one can always find an open subset in $U \subset \mathcal{M}_0$ such that the restriction of the foliation to U is a fibration. Thus, we can always define the reduced phase space locally.

One can show that one can also define (locally) space $\mathcal{M}^*$ so that we have the following geometric analog of diagram (6.23).

$$\begin{array}{ccc} \mathcal{M} & \supset & \mathcal{M}_0 \\ \downarrow & & \downarrow \\ \mathcal{M}^* & \supset & \mathcal{M}_0^* \end{array} \tag{6.25}$$

Remark 6.9. Hamiltonian reduction, considered in Theorem 5.8 in Chapter 5, is a special case of the Dirac formalism. Namely, let $\mathcal{M}$ be a symplectic manifold with a proper Hamiltonian action of a real Lie group G and let $\mu\colon \mathcal{M} \to \mathfrak{g}^*$ be the corresponding moment map. Choose a basis $\xi^a \in \mathfrak{g}$ and define the constraints

$$\varphi^a(m) = \langle \mu(m), \xi^a \rangle.$$

Then it is easy to see that φ^a are first class constraints; in this case,

$$\mathcal{M}_0 = \{m \in \mathcal{M} \mid \varphi^a(m) = 0\} = \mu^{-1}(0).$$

The integral manifolds of the distribution $\mathcal{P}$ are the orbits of G-action, so in this case

$$\mathcal{M}_0^* = \mu^{-1}(0)/G$$

which coincides with the definition of Hamiltonian reduction.

Note that the construction above defines the reduced phase space $\mathcal{M}_0^*$ as a quotient of $\mathcal{M}_0$ by an appropriate equivalence relation. However, it is often convenient to define $\mathcal{M}_0^*$ as a submanifold in $\mathcal{M}_0$ (at least locally). It can be done by choosing a transverse slice, a (local) submanifold $S \subset \mathcal{M}_0$ which is transversal to the integral manifolds of $\mathcal{P}$. Then since locally every integral submanifold intersects S at exactly one point, one can identify $\mathcal{M}_0^* \simeq S$. We will later see that this construction is very common in gauge theories, in particular in Maxwell's theory of electromagnetism and in the Yang-Mills theory, where it is usually referred to as "gauge fixing". It is important to note that the choice of such transeverse slice is not unique: there are many ways to choose it (in the language of gauge theory, there are many different ways to fix the gauge).

The easiest way to define a transverse slice is by imposing additional m equations

$$\chi_a(\boldsymbol{p}, \boldsymbol{q}) = 0, \quad a = 1, \dots, m, \tag{6.26}$$

called *additional constraints*. Condition

$$\det \left(\{\chi_a, \varphi^b\} \right)_{a,b=1}^{m} \neq 0 \tag{6.27}$$

guarantees that the submanifold $S \subset \mathcal{M}_0$, defined by (6.26), intersects transversally the integral manifolds of the distribution $\mathcal{P}$.

If we choose the additional constraints so that

$$\{\chi_a, \chi_b\} = 0, \quad a, b = 1, \dots, m, \tag{6.28}$$

then one can easily find canonical coordinates on $\mathcal{M}_0^*$. Indeed, by Darboux theorem, one find canonical coordinates p_i, q^i, $i = 1, \dots, n$ on $\mathcal{M}$ such that $p_{n-m+a} = \chi_a$ and

$$\omega = \sum_{i=1}^{n} dp_i \wedge dq^i.$$

Transversality condition (6.27) becomes

$$\det \left(\frac{\partial \varphi^a}{\partial q^{n-m+b}} \right)_{a,b=1}^{m} \neq 0.$$

By the implicit function theorem, it means that one can use constraints $\varphi^a(\boldsymbol{p}, \boldsymbol{q}) = 0$ to express q^{n-m+a} in terms of remaining variables. Thus, the reduced phase space

$$\mathcal{M}_0^* \simeq S = \{(\boldsymbol{p}, \boldsymbol{q}) \mid \varphi^a(\boldsymbol{p}, \boldsymbol{q}) = \chi_a(\boldsymbol{p}, \boldsymbol{q}) = 0\}$$

is given by the equations

$$p_{m-n+a} = 0, \quad q^{n-m+a} = f^a(\boldsymbol{p}^*, \boldsymbol{q}^*), \quad a = 1, \dots, m,$$

where $\boldsymbol{p}^* = (p_1, \dots, p_{n-m})$, $\boldsymbol{q}^* = (q^1, \dots, q^{n-m})$. Thererore, $\boldsymbol{p}^*, \boldsymbol{q}^*$ are coordinates on $\mathcal{M}_0^*$; in these coordinates,

$$\omega^* = d\boldsymbol{p}^* \wedge d\boldsymbol{q}^*.$$

We also have $f^*(\boldsymbol{p}^*, \boldsymbol{q}^*) = f(0, \boldsymbol{p}^*, q^a(\boldsymbol{p}^*, \boldsymbol{q}^*), \boldsymbol{q}^*)$ and

$$\{f^*, g^*\} = \frac{\partial f^*}{\partial \boldsymbol{p}^*} \frac{\partial g^*}{\partial \boldsymbol{q}^*} - \frac{\partial f^*}{\partial \boldsymbol{q}^*} \frac{\partial g^*}{\partial \boldsymbol{p}^*}. \tag{6.29}$$

Remark 6.10. For the Hamiltonian reduction, considered in Example 5.4 in Chapter 5, additional constraints χ_a determine an embedding of $\mathcal{M}_0^* = \mu^{-1}(0)/G$ into $\mathcal{M}_0 = \mu^{-1}(0)$ by choosing a representative in each orbit.

6.4. Exercises

Exercise 6.1. Prove that when the matrix $\left\{\dfrac{\partial^2 H}{\partial\lambda_a\partial\lambda_b}\right\}_{a,b=1}^{m}$ has constant rank $k < m$, one can locally choose coordinates $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)$ such that $\lambda_1, \dots, \lambda_k$ are expressed in terms of $\boldsymbol{p}$ and $\boldsymbol{q}$, and the Hamiltonian H is a linear function in the remaining $\lambda_{k+1}, \dots, \lambda_m$.

Exercise 6.2. Prove Lemma 6.3.

Exercise 6.3. Prove (6.29) by computing Poisson bracket $\{f, g\}$ on $\mathcal{M}$ in coordinates $\eta = (p_a, \boldsymbol{p}^*, \varphi^a, \boldsymbol{q}^*)$.

Exercise 6.4. Constraints φ^a for which $\det\left(\{\varphi^a, \varphi^b\}\right)_{a,b=1}^{m} \neq 0$ (this implies $m = 2k$ is even) are called the *second class constraints*. Define the *Dirac bracket* on $\mathcal{M}$ by the formula

$$\{f, g\}_{\mathrm{DB}} = \{f, g\} - \sum_{a,b=1}^{2k} \{f, \varphi^a\} C_{ab} \{\varphi^b, g\},$$

where $\{\ ,\ \}$ is a Poisson bracket on $\mathcal{M}$ associated with the symplectic form ω, and C_{ab} is the inverse matrix to $(\{\varphi^a, \varphi^b\})$.

(1) Prove that Dirac bracket is a Poisson bracket on $\mathcal{M}$.

(2) Prove that Dirac bracket is degenerate and its center consists of the functions $F(\varphi^1, \dots, \varphi^{2k})$, where $F\colon \mathbb{R}^{2k} \to \mathbb{R}$.

(3) Prove that Dirac bracket restricts to $\mathcal{M}_0$ as a non-degenerate Poisson bracket that corresponds to the symplectic form ω_0.

Exercise 6.5. Let $\varphi^a(\boldsymbol{p}, \boldsymbol{q})$ be the first class constraints on the symplectic manifold $\mathcal{M}$ and $\chi_a(\boldsymbol{p}, \boldsymbol{q})$ be the additional constraints.

(1) Show that φ^a and χ_a can be combined into the second class constraints.

(2) Prove that Poisson bracket on the reduced phase space $\mathcal{M}^*$ for the first class constraints φ^a coincides with the restriction to $\mathcal{M}^*$ of the Dirac bracket for the second class constraints φ^a, χ_a.

Chapter 7

Special Relativity

In this chapter, we formulate the main principles of Special Relativity, in which the Newtonian spacetime $E^3 \times \mathbb{R}$, described in Chapter 1, is replaced by the Minkowski spacetime $M^4 = \mathbb{R}^{1,3}$ with the Minkowski metric. Correspondingly, the Galileo's principle of relativity is replaced by the *Special Principle of Relativity*, which we state in the modern form below.

7.1. Minkowski spacetime and the Lorentz group

Recall that in Section 1.1, we have defined the Newtonian spacetime $E^3 \times \mathbb{R}$; choosing a frame, we can identify it with $\mathbb{R}^3 \times \mathbb{R}$, where $\mathbb{R}^3$ is considered with the usual Euclidean metric. We also defined the Galilean group $G = E(3) \ltimes \mathbb{R}^4$, where $E(3) \simeq \mathrm{O}(3) \ltimes \mathbb{R}^3$ is the homogeneous Galilean group, consisting of transformations

$$(\boldsymbol{r}, t) \mapsto (g \cdot \boldsymbol{r} + \boldsymbol{v}t, t)$$

Finally, we formulated one of the main principles of Newtonian mechanics: the laws of motion are invariant under the action of the Galilean group.

Note, however, that the Newtonian spacetime was not given a metric: we only defined the notion of distance for two points in spacetime (events) which are at the same time. Moreover, the action of the Galilean group preserves the notion of simultaneous events: time in Newtonian mechanics is absolute.

However, by the end of 19th century it became clear that the Galileo's relativity principle is not compatible with Maxwell's formulation of electrodynamics. The latter is not invariant with respect to the Galilean group; instead, it is invariant with respect to a deformation of the Galielan group — so-called Lorentz group (see below).

The solution was found by Einstein, who formulated the special theory of relativity in 1905. It was observed by Minkowski in 1908, based on the earlier 1904 paper by Poincaré, that in the special theory of relativity the space and time are unified into a four-dimensional continuum, the Minkowski spacetime.

Namely, denote by $\mathbb{R}^{1,3}$ the 4-dimensional real vector space with the Minkowski metric

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2, \tag{7.1}$$

where $\eta = \mathrm{diag}(1,-1,-1,-1)$. We will commonly use Greek letters for the spacetime indices ranging $0,1,2,3$, and Latin letters for the spacial indices ranging $1,2,3$. Minkowski metric allows us to identify the dual vector space to $\mathbb{R}^{1,3}$ with $\mathbb{R}^{1,3}$. Denoting corresponding vectors by (x_0, x_1, x_2, x_3), as it is commonly used by physicists, we have

$$x_\mu = \eta_{\mu\nu} x^\nu, \quad \text{and} \quad x^2 = \eta_{\mu\nu} x^\mu x^\nu = x_\mu x^\mu.$$

Correspondingly, $x^\mu = \eta^{\mu\nu} x_\nu$, where $\eta^{\mu\nu}$ is the inverse matrix to $\eta_{\mu\nu}$. It is also common to introduce a variable t (time) such that $x^0 = ct$, where c is the speed of light in vacuum; thus, in terms of $t, \boldsymbol{x}$ the metric is given by

$$ds^2 = c^2(dt)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2. \tag{7.2}$$

A vector $x \in \mathbb{R}^{1,3}$ is called time-like if $x^2 = x^\mu x_\mu > 0$, and space-like if $x^2 < 0$. Vectors with $x^2 = 0$ are called light-like; the set of all such vectors is called the lightcone. The picture below illustrates this (for $\mathbb{R}^{1,2}$).

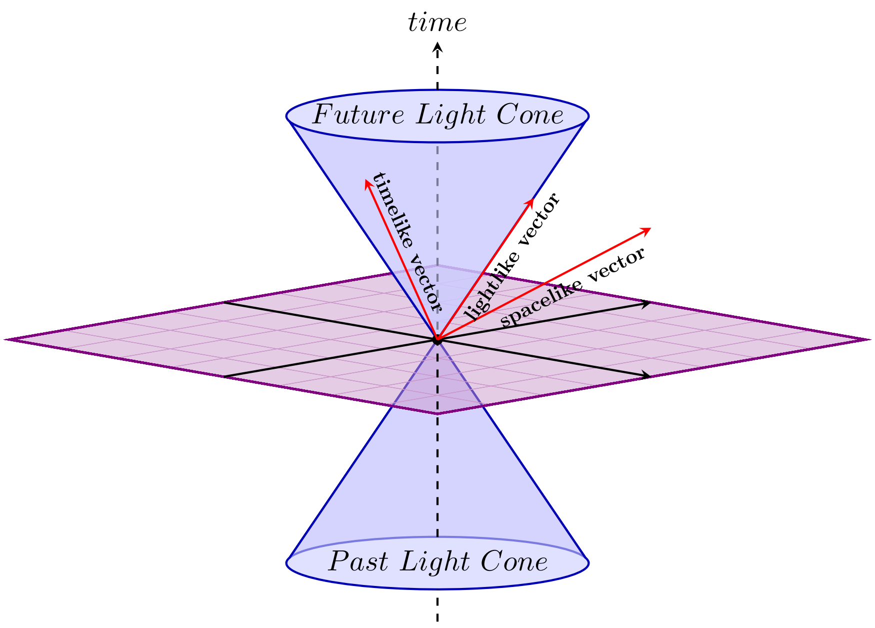

Figure 7.1. Light cone

Definition 7.1. The Minkowski spacetime M^4 is the pseudo-Riemannian manifold which is isometric to $\mathbb{R}^{1,3}$.

We will refer to a choice of isometry $M^4 \simeq \mathbb{R}^{1,3}$ as an *inertial frame*, or simply a frame.

By definition, any two frames are related by an isometry of $\mathbb{R}^{1,3}$. This group plays an important role in special relativity and has a special name.

Definition 7.2. The *Poincaré group* $\mathfrak{P}$ is the group of all isometries of $\mathbb{R}^{1,3}$.

It is easy to show any such isometry has the form

$$x \mapsto g \cdot x + a$$

for some $a \in \mathbb{R}^4$ and $g \in \mathrm{O}(1,3)$; thus,

$$\mathfrak{P} \simeq \mathrm{O}(1,3) \ltimes \mathbb{R}^4. \tag{7.3}$$

In particular, this implies that M^4 has a canonical structure of an affine space. Points in the spacetime M^4 are thought of as representing *events*. We call a pair of events $P_1, P_2 \in M^4$ *space-like* if the vector $P_1 - P_2$ (the interval between events) is space-like, and similarly for time-like and light-like.

The group $\mathrm{O}(1,3)$ of *linear* transformations of $\mathbb{R}^{1,3}$ preserving the Minkowski metric is called the *Lorentz group* and is denoted by $\mathfrak{L}$:

$$\mathfrak{L} = \mathrm{O}(1,3). \tag{7.4}$$

As we will later show, in the limit $c \to \infty$ the Lorentz group becomes the homogeneous Galilean group $E(3)$ introduced in Section 1.1, and the Poincaré group becomes the full Galilean group. Exact meaning of the limit will be clarified in Section 7.4.

Orbits of the Lorentz group in $\mathbb{R}^{1,3}$ have the form

$$\mathcal{O}_m = \{x \in \mathbb{R}^{1,3} \mid x^2 = m\}, \qquad m \in \mathbb{R}.$$

For $m > 0$, these orbits are two-sheeted hyperboloids; for $m < 0$, they are one-sheeted hyperboloids, and for $m = 0$, they are cones (see Fig. 7.1).

It follows from the transitivity of the $\mathfrak{L}$-action on orbits that for any pair of timelike events $P_1, P_2 \in M^4$ there is a frame such that these two events take place in the same point in space: $P_1 = (ct_1, \boldsymbol{x})$, $P_2 = (ct_2, \boldsymbol{x})$ so that $P_2 - P_1 = (c(t_2 - t_1), 0, 0, 0)$. Similarly, for any two space-like events there is a frame such that they happen at the same time: $P_1 = (ct, \boldsymbol{x_1})$, $P_2 = (ct, \boldsymbol{x_2})$, so that $P_2 - P_1 = (0, \boldsymbol{x}_2 - \boldsymbol{x}_1)$.

Finally, the Galileo's principle of relativity is replaced by

Special Principle of Relativity. The laws of motion are invariant under the action of the Poincaré group in the Minkowski spacetime M^4.

In particular, it implies that if a particle is traveling at the speed of light in one frame (i.e. its velocity vector is light-like), it will also travel at the speed of light in any other frame: speed of light is the same in any frame[1].

7.2. Structure of the Lorentz group

In this section we discuss the structure of the Lorentz group $\mathfrak{L} = \mathrm{O}(1,3)$. By definition, this group consists of 4×4 matrices $\Lambda = \{\Lambda^\mu_\alpha\}$ satisfying

$$\Lambda^{\mathrm{t}} \eta \Lambda = \eta, \tag{7.5}$$

where $\eta = \mathrm{diag}(1, -1, -1, -1)$. The group $\mathfrak{L}$ acts linearly on $\mathbb{R}^{1,3}$, $x \mapsto x' = \Lambda x$, where $x'^{\mu} = \Lambda^\mu_\nu x^\nu$. It easily follows from the definition that

$$(\Lambda^0_0)^2 - (\Lambda^1_0)^2 - (\Lambda^2_0)^2 - (\Lambda^3_0)^2 = 1,$$

[1] Originally, Einstein formulated the special principle of relativity as invariance of the laws of physics under a shift of inertial reference frames and the invariance of the speed of light in vacuum. It is equivalent to the modern formulation stated in the text.

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 at least four connected components determined by the signs of Λ^0_0 and $\det \Lambda$.

Definition 7.3. The restricted Lorentz group $\mathfrak{L}^\uparrow_+$ is the subgroup of $\mathfrak{L}$ which preserves orientation and time direction:

$$\mathfrak{L}^\uparrow_+ = \{\Lambda \in \mathfrak{L} \mid \Lambda^0_0 > 0,\ \det \Lambda = 1\}.$$

It can be shown (we skip the proof) that $\mathfrak{L}^\uparrow_+$ is connected and thus is the connected component of identity in $\mathfrak{L}$. Other connected 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 a six-dimensional connected Lie group. It can be shown that it is generated by the subgroup SO(3) of spacial rotations and the *Lorentz boosts*, described below.

Consider the two-dimensional x^0x^1-plane. Denoting for simplicity $x^1 = x$, $x^0 = ct$, consider the transformation of this plane given by $(t, x) \mapsto (t', x')$, where

$$\begin{aligned} x &= x' \cosh\psi + ct' \sinh\psi, \\ ct &= x' \sinh\psi + ct' \cosh\psi. \end{aligned} \tag{7.6}$$

for some 'angle' $\psi \in \mathbb{R}$. It is easy to see that this transformation preserves the form $c^2t^2 - x^2$. It can be trivially extended to $\mathbb{R}^{1,3}$, leaving x^2, x^3 unchanged; so defined transformation lies in the Lorentz group. Replacing x^1 by x^2 or x^3, we get similar transformations acting in planes x^0x^2 and x^0x^3, respectively. Transformations of this form are called *Lorentz boosts.*

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

$$x = \frac{x' + vt'}{\sqrt{1 - \dfrac{v^2}{c^2}}}, \qquad t = \frac{t' + \dfrac{v}{c^2}x'}{\sqrt{1 - \dfrac{v^2}{c^2}}}. \tag{7.7}$$

This transformation relates coordinates (t, x) in the inertial reference frame K with the coordinates (t', x') in the inertial reference frame K' moving relative to K with velocity v along the x-axis. The formula for (t', x') in terms of (t, x) is given by replacing v by $-v$.

Remark 7.4. Transformations (7.7) have a number of elementary but rather surprising corollaries such as the Lorentz contraction (length of a rod measured in a moving frame is less than in a frame in which it is at rest), the time delay (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) and many others, which are discussed in great detail in introductory physics textbooks.

7.3. Lie algebra of the Lorentz and Poincare groups

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 (7.5) by setting $\Lambda = e^{sX} = I + sX + O(s^2)$. Equivalently, it means that $\mathfrak{so}(1,3)$ consists of matrices of the form

$$\begin{pmatrix} 0 & \boldsymbol{v}^{\mathrm{t}} \\ \boldsymbol{v} & u \end{pmatrix}, \qquad \boldsymbol{v} \in \mathbb{R}^3, \quad u \in \mathfrak{so}(3). \tag{7.8}$$

This implies that as a vector space (but not as a Lie algebra!), $\mathfrak{so}(1,3)$ is the direct sum $\mathfrak{so}(3) \oplus \mathbb{R}^3$. In particular, $\dim \mathfrak{so}(1,3) = 6$.

Thus, we can easily define a basis of $\mathfrak{so}(1,3)$. Namely, recall the basis J_i of $\mathfrak{so}(3)$ which was defined in Section 1.1. Using the embedding $\mathfrak{so}(3) \hookrightarrow \mathfrak{so}(1,3)$, we can consider J_i as elements in $\mathfrak{so}(1,3)$:

$$J_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \qquad J_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \qquad J_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{7.9}$$

In addition, taking the vector $\boldsymbol{v}$ in (7.8) to be one of the standard basis vectors in $\mathbb{R}^3$, we get three more basis vectors in $\mathfrak{so}(1,3)$:

$$K_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad K_2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad K_3 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \tag{7.10}$$

Using

$$\exp\left(s\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right) = \begin{pmatrix} \cosh s & \sinh s \\ \sinh s & \cosh s \end{pmatrix}$$

it is easy to see that the one-parameter subgroup generated by K_1 is exactly the group of Lorentz boosts in x^0x^1 plane described by (7.6); similarly, K_2, K_3 generate boosts in x^0x^2 and x^0x^3 planes.

Elements J_i, K_i form a basis of $\mathfrak{so}(1,3)$. An explicit computation shows that in this basis, the commutation relations are given by

$$\begin{aligned} [J_i, J_j] &= \epsilon_{ijl} J_l, \\ [J_i, K_j] &= \epsilon_{ijl} K_l, \\ [K_i, K_j] &= -\epsilon_{ijl} J_l, \end{aligned} \tag{7.11}$$

where ϵ_{ijk} is the fully antisymmetric tensor, see (1.17).

We can also give a basis in the Lie algebra $\mathfrak{p}$ of the Poincaré group $\mathfrak{P} = \mathrm{O}(1,3) \ltimes \mathbb{R}^4$. Denoting by P_μ the generators of $\mathfrak{p}$ corresponding to spacetime translations, we obtain the following set of relations:

$$(7.12)\qquad \begin{aligned}[][P_i, P_j] &= [P_i, P_0] = 0,\\ [J_i, P_0] &= 0, \quad [J_i, P_j] = \epsilon_{ijk}P_k,\\ [K_i, P_0] &= P_i, \quad [K_i, P_j] = \delta_{ij}P_0.\end{aligned}$$

It is also common in physics literature to use generators $M_{\alpha\beta}$ of $\mathfrak{so}(1,3)$, $\alpha, \beta = 0, 1, 2, 3$, defined by the following formulas

$$(7.13)\qquad M_{ij} = -\epsilon_{ijk}J_k, \qquad M_{0i} = -M_{i0} = K_i, \quad i, j = 1, 2, 3.$$

This is a special case of a more general definition, which works for any non-degenerate bilinear form η, see Exercise 7.1. Then relations (7.11), (7.12) can be rewritten in the unified form:

$$(7.14)\qquad \begin{aligned}[][P_\mu, P_\nu] &= 0,\\ [M_{\lambda\mu}, P_\nu] &= \eta_{\lambda\nu}P_\mu - \eta_{\mu\nu}P_\lambda,\\ [M_{\lambda\mu}, M_{\rho\sigma}] &= \eta_{\lambda\rho}M_{\mu\sigma} - \eta_{\lambda\sigma}M_{\mu\rho} + \eta_{\mu\sigma}M_{\lambda\rho} - \eta_{\mu\rho}M_{\lambda\sigma},\end{aligned}$$

often used by physicists. (We leave derivation of these relations to the reader; see Exercise 7.1).

Remark 7.5. In Example 2.6, we used generators M_{ij} defined by $M_{ij} = \epsilon_{ijk}J_k$, which differs by sign from the formula above. The reason for this sign change is that now we use a different metric, so now $\eta_{ii} = -1$ for $i = 1, 2, 3$.

It is also common to "raise indices", defining $P^\mu = \eta^{\mu\nu}P_\nu$, $M^{\mu\nu} = \eta^{\mu\alpha}\eta^{\nu\beta}M_{\alpha\beta}$ so that

$$(7.15)\qquad M^{ij} = M_{ij} = -\epsilon_{ijk}J_k, \qquad M^{0i} = -M_{0i} = -K_i, \quad i, j = 1, 2, 3.$$

We leave it to the reader to check that so defined generators satisfy

$$(7.16)\qquad \begin{aligned}[][P^\mu, P^\nu] &= 0,\\ [M^{\lambda\mu}, P^\nu] &= \eta^{\lambda\nu}P^\mu - \eta^{\mu\nu}P^\lambda,\\ [M^{\lambda\mu}, M^{\rho\sigma}] &= \eta^{\lambda\rho}M^{\mu\sigma} - \eta^{\lambda\sigma}M^{\mu\rho} + \eta^{\mu\sigma}M^{\lambda\rho} - \eta^{\mu\rho}M^{\lambda\sigma}.\end{aligned}$$

Remark 7.6. Note that for Lie algebra of the Galilean group, we had a relation $[K_i, P_j] = 0$ in $\mathfrak{g}$, but that Lie algebra allowed a central extension, in which that relation was replaced by $[K_i, P_j] = \delta_{ij}z$ (see Remark 1.9). For the Lie algebra of Poincaré group, we now have $[K_i, P_j] = \delta_{ij}P_0$, so this commutator is non-zero even without taking a central extension. In fact, one can show that $\mathfrak{p}$ has no non-trivial central extensions (see Section 7.4 below).

Formulas (7.11) look almost like two copies of the relations for generators J_i of Lie algebra $\mathfrak{so}(3)$. More precisely, we have the following result. Let $\mathfrak{sl}(2,\mathbb{C})_\mathbb{R}$ denote the Lie algebra $\mathfrak{sl}(2,\mathbb{C})$ considered as a 6-dimensional **real** Lie algebra.

Lemma 7.7. *Consider the map*

$$(7.17)\qquad \begin{aligned}\mathfrak{so}(1,3) &\to \mathfrak{sl}(2,\mathbb{C})_\mathbb{R}\\ J_a &\mapsto \tfrac{1}{2i}\sigma_a\\ K_a &\mapsto \tfrac{1}{2}\sigma_a\end{aligned}$$

where σ_a, $a = 1, 2, 3$, are Pauli matrices (see Exercise 1.7). Then this map is an isomorphism of real Lie algebras.

The proof is immediate from an explicit computation (compare with Exercise 1.7).

It can be shown that this isomorphism lifts to an isomorphism of **real** Lie groups

$$\mathrm{SL}(2,\mathbb{C})/\mathbb{Z}_2 \cong \mathfrak{L}_+^{\uparrow}. \tag{7.18}$$

We give an explicit construction of this isomorphism in Exercise 7.4. Note that since $\mathrm{SL}(2,\mathbb{C})$ is simply-connected, this shows that $\mathrm{SL}(2,\mathbb{C})$ is the universal cover of the restricted Lorentz group.

As an immediate corollary, we get the following description of the complexified Lie algebra of the Lorentz group

$$\mathfrak{so}(1,3) \otimes \mathbb{C} \cong \mathfrak{sl}(2,\mathbb{C}) \oplus \mathfrak{sl}(2,\mathbb{C}) \tag{7.19}$$

which will be very useful in the future. In particular, this shows that $\mathfrak{so}(1,3)$ is semisimple. An explicit construction of this isomorphsim is given in Exercise 7.2.

7.4. Lorentz group as the deformation of Galilean group

If instead of using coordinates $x^0, \dots, x^3$ in $\mathbb{R}^{1,3}$ we use (t, x^1, x^2, x^3), where $x^0 = ct$, then in these coordinates the Lie algebra of the Lorentz group is given by

$$X^{\mathrm{t}}\eta_c + \eta_c X = 0,$$

where $\eta_c = \mathrm{diag}(c^2, -1, -1, -1)$. Solutions of this equation are matrices of the form

$$\begin{pmatrix} 0 & \frac{1}{c^2}\boldsymbol{v}^{\mathrm{t}} \\ \boldsymbol{v} & u \end{pmatrix}, \qquad \boldsymbol{v} \in \mathbb{R}^3, \quad u \in \mathfrak{so}(3). \tag{7.20}$$

Thus, we can choose a basis J_i, $\tilde{K}_i$, where $\tilde{K}_i$ are defined by

$$\tilde{K}_1 = \begin{pmatrix} 0 & 1/c^2 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \tilde{K}_2 = \begin{pmatrix} 0 & 0 & 1/c^2 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \tilde{K}_3 = \begin{pmatrix} 0 & 0 & 0 & 1/c^2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \tag{7.21}$$

It is easy to see that these generators satisfy commutation relations

$$\begin{aligned} [J_i, J_j] &= \epsilon_{ijl} J_l, \\ [J_i, \tilde{K}_j] &= \epsilon_{ijl} \tilde{K}_l, \\ [\tilde{K}_i, \tilde{K}_j] &= -\tfrac{1}{c^2}\epsilon_{ijl} J_l. \end{aligned} \tag{7.22}$$

One can view these relations as defining a family of Lie algebras, depending on parameter c, on the same vector space, with basis J_i, $\tilde{K}_i$.

Note that for any $c \neq 0$ these Lie algebras are isomorphic (indeed, each of them is obtained from the usual $\mathfrak{so}(1,3)$ by a change of basis in $\mathbb{R}^{1,3}$). However, in the limit $c \to \infty$, we get a different Lie algebra structure: in this limit, the commutation relations become

$$[J_i, J_j] = \epsilon_{ijl} J_l,$$
$$[J_i, \tilde{K}_j] = \epsilon_{ijl} \tilde{K}_l,$$
$$[\tilde{K}_i, \tilde{K}_j] = 0$$

and coincide with the commutation relation in the Lie algebra $\mathfrak{se}(3)$ of the homogeneous Galilean group $E(3) = \mathrm{O}(3) \ltimes \mathbb{R}^3$, see (1.19).

This is a special case of a general situation commonly called *deformation of Lie algebras.* Namely, a *formal deformation* of a finite-dimensional real Lie algebra $\mathfrak{g}$ with a Lie bracket $[\ ,\ ]$ is a Lie algebra $\tilde{\mathfrak{g}} = \mathfrak{g} \otimes \mathbb{R}[[h]]$ over $\mathbb{R}[[h]]$, a ring of formal power series in the variable h, with the Lie bracket

$$[x, y]_h = [x, y] + hm_1(x, y) + h^2 m_2(x, y) + \cdots$$

In particular, setting $h = 1/c^2$, we see that the family of Lie algebra structures given by (7.22) is a deformation of Lie algebra $\mathfrak{se}(3)$.

It is natural to ask whether one can classify all deformations, up to a suitable notion of isomorphism. Namely, we say that two formal deformations $[\ ,\ ]_h$, $[\ ,\ ]'_h$ are equivalent if there is an automorphism

$$F(x) = x + hf(x) + \cdots \in \mathrm{End}(\mathfrak{g})[[h]]$$

such

$$[x, y]'_h = F^{-1}([F(x), F(y)]_h).$$

Replacing in the definitions above $\mathbb{R}[[h]]$ by the quotient $\mathbb{R}[[h]]/(h^2)$, i.e. keeping only terms of degrees 0 and 1 in h, we get a definition of the *infinitesimal* deformation. It turns out that the set of isomorphism classes of infinitesimal deformations admits a description in terms of the Lie algebra cohomology.

Remark 7.8. The relation between infinitesimal deformations and formal deformations is non-trivial: for example, not every infinitesimal deformation can be extended to a formal one. We will not discuss the details here, restricting ourselves to the study of infinitesimal deformations.

Recall that for any $\mathfrak{g}$-module M, one can define the cohomology of $\mathfrak{g}$ with values in M using the general formalism of Ext functors:

$$H^i(\mathfrak{g}, M) = \mathrm{Ext}^i(\mathbb{R}, M), \tag{7.23}$$

where Ext functors are computed in the category of $\mathfrak{g}$-modules (or, equivalently, category of modules over the universal enveloping algebra $U\mathfrak{g}$).

These cohomology also admits a more explicit definition using the Chevalley-Eilenberg complex:

$$C^k(M) = \mathrm{Hom}(\Lambda^k \mathfrak{g}, M)$$

with the differential $\delta_k\colon C^k \to C^{k+1}$ given by

$$(\delta_k f)(x_1,\dots,x_{k+1}) = \sum_{i=1}^{k+1}(-1)^{i+1}x_i\cdot f(x_1,\dots,\hat{x}_i,\dots,x_{k+1}) + \sum_{1\le i<j\le k+1}(-1)^{i+j}f([x_i,x_j],x_1,\dots,\hat{x}_i,\dots,\hat{x}_j,\dots x_{k+1}).$$

Cohomology of this complex coincides with the cohomology defined before via Ext functors (this follows from using a very special resolution for computing the cohomology).

Remark 7.9. Consider the case when $\mathfrak{g} = \mathrm{Vect}(X)$ is the Lie algebra of vector fields on a compact smooth manifold X, and $M = C^\infty(X)$. Since $\mathfrak{g}$ is infinite-dimensional, the space of all linear maps $\mathrm{Hom}(\Lambda^k\mathfrak{g}, M)$ is too large. Instead, notice that we have a natural pairing $\Lambda^k\mathfrak{g}\otimes\Omega^k(X)\to C^\infty(X)$, where $\Omega^k(X)$ is the space of differential forms of degree k on X. Thus, we have an embedding $\Omega^k(X)\hookrightarrow \mathrm{Hom}(\Lambda^k\mathfrak{g}, M)$. Restriction of the Chevalley-Eilenberg differential to the subcomplex $\Omega^\bullet(X)$ coincides with the exterior derivative on differential forms. Thus, in this case (properly understood) Chevalley-Eilenberg complex is exactly the de Rham complex $\Omega^\bullet_{\mathrm{dR}}(X,\mathbb{R})$.

Theorem 7.10. *Isomorphism classes of infinitesimal deformations of a finite-dimensional Lie algebra $\mathfrak{g}$ are in bijection with the space $H^2(\mathfrak{g},\mathfrak{g})$, where $\mathfrak{g}$ is considered with the adjoint $\mathfrak{g}$-action.*

Proof. Any infinitesimal deformation of $\mathfrak{g}$

$$[x,y]_h = [x,y] + hm(x,y) \mod h^2$$

is determined by the linear function $m(x,y)$, which clearly must be skew-symmetric and thus can be considered as an element of $C^2(\mathfrak{g}) = \mathrm{Hom}(\Lambda^2\mathfrak{g},\mathfrak{g})$.

Jacobi equation on $[\ ,\ ]_h$ gives the following condition on m:

$$\begin{gathered}[m(x,y),z] + m([z,x],y) + [m(y,z),x]\\ +[m(z,x),y] + m([z,x],y) + m([y,z],x) = 0\end{gathered} \tag{7.24}$$

for all $x,y,z\in\mathfrak{g}$. Comparing it with the differential of the Chevalley-Eilenberg complex, we see that this condition is equivalent to $\delta_2 m = 0$. Thus, m defines a Lie algebra structure if and only if it is a 2-cocycle.

Similarly, if two cocycles m, m' define isomorphic deformations:

$$[x,y]'_h = F^{-1}([F(x),F(y)]_h) \mod h^2$$

for $F(x) = x + hf(x) \mod h^2$, then

$$m'(x,y) - m(x,y) = [x,f(y)] - [y,f(x)] - f([x,y]) = \delta_1 f(x,y)$$

is a coboundary, so $[m]=[m']$ in $H^2(\mathfrak{g},\mathfrak{g})$. Conversely, if $m'-m$ is a coboundary, then the corresponding deformations are isomorphic. □

As an immediate corollary, we see that if $H^2(\mathfrak{g},\mathfrak{g}) = 0$, then $\mathfrak{g}$ has no non-trivial infinitesimal deformations. Such Lie algebras are called *stable*. It can be shown that for stable algebras, there are also no non-trivial formal deformations: small deformations of a stable Lie algebra gives an isomorphic Lie algebra.

Any semisimple Lie algebra is stable. However, for the Lie algebra $\mathfrak{g} = \mathfrak{se}(3)$ we have $H^2(\mathfrak{g}, \mathfrak{g}) = \mathbb{R}$; the generator of H^2 is the 2-cocycle m with the only non-zero values $m(\tilde{K}_i, \tilde{K}_j) = \epsilon_{ijk} J_k$.

The corresponding infinitesimal deformation is given by

$$\begin{aligned} [J_i, J_j] &= \epsilon_{ijl} J_l, \\ [J_i, \tilde{K}_j] &= \epsilon_{ijl} \tilde{K}_l, \\ [\tilde{K}_i, \tilde{K}_j] &= h\epsilon_{ijl} J_l. \end{aligned}$$

Setting $h = -1/c^2$, we get exactly the relations (7.22) of Lorentz Lie algebra. In this case, this is not only an infinitesimal deformation but an actual deformation defined over $\mathbb{R}[h]$: Jacobi identity holds perfectly, and not just $\mod h^2$.

The Lorentz Lie algebra $\mathfrak{so}(1,3)$ is semisimple and therefore is stable. Thus, the passage from the Newtonian spacetime to the Minkowski spacetime is a deformation from the unstable structure to the stable one, so the special relativity is a natural deformation of the Newtonian mechanics.

In a similar way, one can show that central extensions of a real Lie algebra $\mathfrak{g}$ are classified by cohomology $H^2(\mathfrak{g}, \mathbb{R})$. For the Lie algebra of Galilean group, we have $H^2(\mathfrak{g}, \mathbb{R}) = \mathbb{R}$, so this Lie algebra has a non-trivial central extension, which is unique (up to rescaling the central element); this central extension has been constructed in Remark 1.9. For the Lie algebra of the Poincaré group, we have $H^2(\mathfrak{p}, \mathbb{R}) = 0$ (which is related to the fact that $\mathfrak{p} = \mathfrak{so}(1,3) \ltimes \mathbb{R}^4$, and $\mathfrak{so}(1,3)$ is semisimple); thus, it has no non-trivial central extensions.

7.5. Exercises

Exercise 7.1. Let V be n-dimensional real or complex vector space with non-degenerate symmetric bilinear form η. Let $\mathrm{SO}(\eta)$ be the group of automorphisms preserving η and the orientation and let $\mathfrak{so}(\eta)$ be the corresponding Lie algebra.

(1) For any $v_1, v_2 \in V$, define the linear operator $M_{v_1,v_2}\colon V \to V$ by

$$M_{v_1,v_2}(w) = \eta(v_1, w)v_2 - \eta(v_2, w)v_1. \tag{7.25}$$

Prove that $v_1 \wedge v_2 \mapsto M_{v_1,v_2}$ gives an isomorphism

$$\Lambda^2(V) \xrightarrow{\sim} \mathfrak{so}(\eta).$$

(2) Let e_α be a basis in V, and let $\eta_{\mu\nu}$ be matrix of η in this basis: $\eta_{\alpha\beta} = \eta(e_\alpha, e_\beta)$. Let $\eta^{\mu\nu}$ be the inverse matrix.

Define $M_{\alpha\beta}, M^{\alpha\beta} \in \mathfrak{so}(\eta)$ by

$$\begin{aligned} M_{\alpha\beta} &= M_{e_\alpha,e_\beta}, \\ M^{\alpha\beta} &= \eta^{\alpha\mu}\eta^{\beta\nu} M_{\mu\nu}. \end{aligned}$$

Show that then matrix entries of $M^{\alpha\beta}$ are given by

$$\begin{aligned} (M_{\alpha\beta})^\mu_\nu &= \eta_{\alpha\nu}\delta^\mu_\beta - \eta_{\beta\nu}\delta^\mu_\alpha \\ (M^{\alpha\beta})^\mu_\nu &= \eta^{\beta\mu}\delta^\alpha_\nu - \eta^{\alpha\mu}\delta^\beta_\nu. \end{aligned} \tag{7.26}$$

In particular, for $\eta = \mathrm{diag}(1, -1, -1, -1)$ this coincides with (7.13), and for $V = \mathbb{R}^n$ with the standard inner product, so defined $M_{\alpha\beta}$ coincide with generators introduced in Exercise 2.6.

(3) Show that $M^{\alpha\beta}$, $M_{\alpha\beta}$ satisfy the following commutation relations:

$$\begin{aligned}[M_{\lambda\mu}, M_{\rho\sigma}] &= \eta_{\lambda\rho}M_{\mu\sigma} - \eta_{\lambda\sigma}M_{\mu\rho} + \eta_{\mu\sigma}M_{\lambda\rho} - \eta_{\mu\rho}M_{\lambda\sigma} \\ [M^{\lambda\mu}, M^{\rho\sigma}] &= \eta^{\lambda\rho}M^{\mu\sigma} - \eta^{\lambda\sigma}M^{\mu\rho} + \eta^{\mu\sigma}M^{\lambda\rho} - \eta^{\mu\rho}M^{\lambda\sigma}\end{aligned} \tag{7.27}$$

(compare with Exercise 2.6).

Exercise 7.2. Consider the following basis of $\mathfrak{so}(1,3)\otimes\mathbb{C}$:

$$J_k^{(\pm)} = \tfrac{1}{2}(J_k \pm iK_k).$$

Show that so defined elements satisfy

$$[J_i^{(+)}, J_j^{(+)}] = \epsilon_{ijl}J_l^{(+)}, \quad [J_i^{(-)}, J_j^{(-)}] = \epsilon_{ijl}J_l^{(-)}, \quad [J_i^{(+)}, J_j^{(-)}] = 0,$$

and deduce from this that the map

$$\begin{aligned}\mathfrak{so}(1,3)_{\mathbb{C}} &\to \mathfrak{so}(3,\mathbb{C}) \oplus \mathfrak{so}(3,\mathbb{C}) \\ J_k^{(+)} &\mapsto (J_k, 0) \\ J_k^{(-)} &\mapsto (0, J_k)\end{aligned}$$

is an isomorphism of Lie algebras.

Since $\mathfrak{so}(3,\mathbb{C}) \cong \mathfrak{sl}(2,\mathbb{C})$ (see Exercise 1.7), this gives another construction of the isomorphism $\mathfrak{so}(1,3)\otimes\mathbb{C} \cong \mathfrak{sl}(2,\mathbb{C}) \oplus \mathfrak{sl}(2,\mathbb{C})$ (compare with (7.19)).

Exercise 7.3. The goal of this problem is to describe the group Spin(4) – the universal cover of SO(4).

Recall that the algebra of quaternions $\mathbb{H}$ is the four-dimensional algebra over $\mathbb{R}$ with basis $1, \boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ and relations

$$\begin{aligned}&\boldsymbol{i}^2 = \boldsymbol{j}^2 = \boldsymbol{k}^2 = \boldsymbol{ijk} = -1 \\ &\boldsymbol{ij} = -\boldsymbol{ji}, \quad \boldsymbol{ik} = -\boldsymbol{ki}, \quad \boldsymbol{jk} = -\boldsymbol{kj}.\end{aligned}$$

The complex conjugation and norm in $\mathbb{H}$ are defined by letting for $h = a + b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k}$

$$\begin{aligned}\bar{h} &= a - b\boldsymbol{i} - c\boldsymbol{j} - d\boldsymbol{k}, \\ |h|^2 &= h\bar{h} = a^2 + b^2 + c^2 + d^2.\end{aligned}$$

(1) Consider the embedding of $\mathbb{H}$ into the algebra of 2×2 complex matrices given by

$$a + b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k} \mapsto \begin{pmatrix} a+bi & c+di \\ -c+di & a-bi\end{pmatrix}$$

Show that this embedding identifies the group of unit quaternions $\mathbb{U} = \{h \in \mathbb{H} \mid |h| = 1\}$ with SU(2).

(2) Consider the action of $\mathbb{U}\times\mathbb{U}$ on $\mathbb{H}$ given by

$$(X, Y)\cdot h = Xh\overline{Y}$$

Show that this action preserves the norm and thus defines a morphism $\mathbb{U}\times\mathbb{U} = \mathrm{SU}(2)\times\mathrm{SU}(2) \to \mathrm{SO}(4)$.

(3) The morphism defined in the previous part gives rise to a morphism of Lie algebras $\mathfrak{su}(2)\oplus\mathfrak{su}(2) \to \mathfrak{so}(4)$ and thus to a morphism of their complexiofications

$$\mathfrak{sl}(2,\mathbb{C}) \oplus \mathfrak{sl}(2,\mathbb{C}) \to \mathfrak{so}(4,\mathbb{C}).$$

Show that this morphism is an isomorphism. (After identifying $\mathfrak{so}(4,\mathbb{C}) \cong \mathfrak{so}(1,3)_{\mathbb{C}}$, this isomorphism coincides with the isomorphism constructed in the previous problem.)

(4) Show that the morphism defined in part (2) is surjective and its kernel is $\{(I, I), (-I, -I)\} \cong \mathbb{Z}_2$, thus establishing an isomorphism
$$\mathrm{SO}(4) \cong (\mathrm{SU}(2) \times \mathrm{SU}(2))/\mathbb{Z}_2.$$

Exercise 7.4. In this problem, we describe the group Spin(1, 3) — the universal cover of the Lorentz group.

Consider the following isomorphism between vector space $\mathbb{R}^{1,3}$ and the space of Hermitian 2×2 matrices:
$$\boldsymbol{x} = (x^0, x^1, x^2, x^3) \mapsto X = x^0\sigma_0 + x^1\sigma_1 + x^2\sigma_2 + x^3\sigma_3 = \begin{pmatrix} x^0 + x^3 & x^1 - ix^2 \\ x^1 + ix^2 & x^0 - x^3 \end{pmatrix},$$
where σ_μ are the Pauli matrices defined in Exercise 1.7):
$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{7.28}$$

(1) Show that $\boldsymbol{x}^2 = \det X$.

(2) Use the previous part and the action of $\mathrm{SL}(2,\mathbb{C})$ (considered as a real Lie group) on the space of Hermitian matrices given by $X \mapsto AXA^\dagger$ (here † stands for the Hermitian conjugation) to define a morphism of real Lie groups $\mathrm{SL}(2,\mathbb{C}) \to \mathfrak{L}_+^\uparrow$, given explicitly by $\Lambda^\mu_\nu = \frac{1}{2}\,\mathrm{Tr}(\sigma_\mu A \sigma_\nu A^\dagger)$.

(3) Show that the morphism constructed in the previous part gives rise to isomorphism of real Lie algebras defined in Lemma 7.7.

(4) Show that the morphism constructed in part (2) gives an isomorphism of real Lie groups
$$\mathfrak{L}_+^\uparrow \cong \mathrm{SL}(2,\mathbb{C})/\{\pm I\},$$
so that $\mathrm{SL}(2,\mathbb{C})$, considered as a real Lie group, is the universal cover of $\mathfrak{L}_+^\uparrow$.

Chapter 8

Relativistic Particle

In this chapter we give Lagrangian and Hamiltonian descriptions of a free relativistic particle, i.e. the particle whose motion is invariant under the action of the Poincaré group.

We show that the Poincaré group acts naturally on the phase space of a free particle and this action is Hamiltonian. Finally, we explain that special principle of relativity imposes strong restriction on the interaction between the particles, which can be only described through their interaction with a field and not by potential forces.

Here we continue to use the Minkowski spacetime M^4, defined in the previous chapter. We will often assume that we have chosen an inertial frame, so there is an identification $M^4 \cong \mathbb{R}^{1,3}$ with the standard coordinates $x^0 = ct, x^1, x^2, x^3$ in $\mathbb{R}^{1,3}$. However, we will be careful that all our constructions are actually independent of the choice of frame.

8.1. World line of a particle

A motion of a particle in M^4 is described by a *world line*, connecting timelike events $P_0 = (ct_0, \boldsymbol{r}_0)$ and $P_1 = (ct_1, \boldsymbol{r}_1)$. By definition, it is a map $\gamma\colon [\tau_0, \tau_1] \to M^4$, $\gamma(\tau) = x^\mu(\tau)$, such that $\gamma(\tau_0) = P_0$, $\gamma(\tau_1) = P_1$, and at each $\tau \in [\tau_0, \tau_1]$ the tangent vector $\gamma'(\tau)$ is timelike (see Section 7.1).

Note that in this case τ should not be considered as time: it is just a parameter along the trajectory. Moreover, trajectories which only differ by a reparametrization are physically equivalent. Thus, it is common to choose the parameter in a special way.

The simplest choice is to use time t (relative to a chosen inertial frame) as parameter so that $\gamma(t) = (ct, \boldsymbol{r}(t))$. In this case, $\gamma'(t) = (c, \boldsymbol{v}(t))$, $\boldsymbol{v}(t) = \dot{\boldsymbol{r}}(t)$, and

$$|\gamma'(t)| = \sqrt{c^2 - v^2(t)} = c\sqrt{1 - \frac{v^2}{c^2}},$$

where $v(t) = |\boldsymbol{v}(t)|$ is the Euclidean length of a vector $\boldsymbol{v}(t)$ in $\mathbb{R}^3$. Another choice is to use the natural parameter s, the Minkowski length along the trajectory:

$$s = \int |\gamma'(\tau)|\, d\tau$$

which manifestly doesn't depend on the choice of a frame or any other choices (up to adding a constant).

The relation between these parametrizations is given by

$$s(t) = c\int_{t_0}^{t}\sqrt{1-\frac{v^2(\tau)}{c^2}}\,d\tau.$$

In particular, if we choose a frame in which the particle is stationary ($\boldsymbol{r}(t)$ doesn't depend on t), then $s(t) = c(t-t_0)$. For this reason, the quantity $c^{-1}s(\tau)$ is frequently referred to as *proper time* along the trajectory. In terms of the natural parameter the tangent vector of unit length is

$$u^\mu = \frac{dx^\mu(s)}{ds} = \left(\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}, \frac{\boldsymbol{v}}{c\sqrt{1-\frac{v^2}{c^2}}}\right), \quad u^2 = u_\mu u^\mu = 1. \tag{8.1}$$

8.2. The principle of the least action

As in Chapter 1, our goal is to define trajectories of motion as extremals of the action

$$S(\gamma) = \int_{\tau_0}^{\tau_1} L(\gamma(\tau), \gamma'(\tau))\,d\tau.$$

Note however that in this case, we want the Lagrangian to be independent of reparametrization of the trajectory and invariant under the action of the Poincaré group. Also, since space and time are now unified into the Minkowski spacetime, the configuration space of a relativistic particle is M^4 and not $\mathbb{R}^3$; time in a given reference frame is just one of possible parameterizations of a world line.

The simplest and most natural choice of a Poincaré invariant action is

$$S(\gamma) = -\alpha\int_{P_0}^{P_1} ds = -\alpha\int_{\tau_0}^{\tau_1}|\partial_\tau x|\,d\tau, \quad |\partial_\tau x| = \sqrt{\partial_\tau x^\mu \partial_\tau x_\mu}, \tag{8.2}$$

where $\partial_\tau x^\mu = \dfrac{dx^\mu}{d\tau}$. Here integration goes over the world line γ with timelike velocity vectors that connects initial and final events P_0 and P_1, and α is a constant.

If we use $\tau = t$ as the parameter, then this action can be rewritten as

$$S(\gamma) = \int_{t_0}^{t_1} L(\gamma'(t))dt, \quad \text{where} \quad L = -\alpha c\sqrt{1-\frac{v^2}{c^2}} \quad \text{and} \quad v = |\dot{\boldsymbol{r}}| < c \tag{8.3}$$

(note that $|\dot{\boldsymbol{r}}|^2 = \dot{\boldsymbol{r}}\cdot\dot{\boldsymbol{r}}$ stands for the Euclidean metric in $\mathbb{R}^3$).

We can now consider the extremals of this action on the space of trajectories with fixed endpoints. As before, we only consider the situation when the interval between the endpoints is timelike.

Theorem 8.1. *Extremals of action the* (8.2) *are straight lines in Minkowski space with timelike velocity vectors.*

Proof. Choosing $\tau = t$ as a parameter, from (8.3) we immediately obtain that in this reference frame

$$\int_{P_0}^{P_1} ds = c\int_{t_0}^{t_1}\sqrt{1-\frac{v^2(t)}{c^2}}dt \le c(t_1-t_0).$$

Thus, if $\alpha > 0$, then $S(\gamma) \ge -\alpha c(t_1-t_0)$, with the equality only if $v(t) = 0$ on γ. This corresponds to straight line trajectory $\boldsymbol{r}(t) = const$. Thus for $\alpha > 0$, such a trajectory is the absolute minimum of the action functional. □

To compare L with the Lagrangian for a free particle in classical mechanics, let us consider the limit $c \to \infty$. In this limit,

$$L = -\alpha c\sqrt{1-\frac{v^2}{c^2}} = -\alpha c + \frac{\alpha v^2}{2c} + O(\alpha c^{-3}).$$

Thus we see that letting $\alpha = mc$ and omitting the constant term $-\alpha c$, which does not affect the equations of motion, this expression matches the Lagrangian of a free non-relativistic paricle:

$$L = -\alpha c\sqrt{1-\frac{v^2}{c^2}} = -mc^2 + \frac{mv^2}{2} + O(c^{-2}), \quad \alpha = mc. \tag{8.4}$$

Therefore, we set $\alpha = mc$ in (8.2) and define the action for a free relativistic particle of mass m by

$$S(\gamma) = -mc\int_{P_0}^{P_1} ds = \int_{\tau_0}^{\tau_1} L(\gamma'(\tau))d\tau, \tag{8.5}$$

with the Lagrangian function

$$L = -mc\sqrt{\partial_\tau x_\mu \partial_\tau x^\mu}. \tag{8.6}$$

In particular, if we choose parametrization $\tau = t$, this becomes

$$L = -mc^2\sqrt{1-\frac{v^2}{c^2}}. \tag{8.7}$$

As in Section 1.3, we can easily write down corresponding Euler-Lagrange equations in relativistic invariant form.

Proposition 8.2. *The Euler-Lagrange equations of a free relativistic particle are*

$$\frac{du^\mu}{ds} = 0$$

where u^μ is the unit tangent vector (8.1).

Thus, solutions of these equations describe motion with constant velocity.

Proof. Since $ds = \sqrt{dx_\mu dx^\mu}$, we have along the worldline 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 $\delta x_\mu(P_0) = \delta x_\mu(P_1) = 0$, we obtain

$$\delta S = -mc \int_{P_0}^{P_1} \delta(ds) = mc \int_{P_0}^{P_1} \frac{du^\mu}{ds} \delta x_\mu ds. \qquad \square$$

We also have the variational formula with free ends (cf. Remark 1.6),

$$\delta S = mc \int_{P_0}^{P_1} \frac{du^\mu}{ds} \delta x_\mu ds - mc\, u^\mu \delta x_\mu \big|_{P_0}^{P_1}. \tag{8.8}$$

As in Chapter 2, we define the momentum covector $(p_\mu) \in T^*\mathbb{R}^4$ by

$$p_\mu = \frac{\partial L}{\partial(\partial_\tau x^\mu)}.$$

Since the Lagrangian L with arbitrary parameter τ along the world-line is given by $L = -mc\sqrt{\partial_\tau x^\mu \partial_\tau x_\mu}$, we can rewrite the previous formula as

$$p_\mu = -\frac{mc\, \partial_\tau x^\mu}{|\partial_\tau x|}. \tag{8.9}$$

Choosing $\tau = t$ as a parameter, we get

$$\begin{aligned} p_i &= \frac{\partial L}{\partial v^i} = \frac{mv^i}{\sqrt{1 - \dfrac{v^2}{c^2}}}, \quad i = 1, 2, 3 \\ p_0 &= \frac{-mc}{\sqrt{1 - \dfrac{v^2}{c^2}}} \end{aligned} \tag{8.10}$$

where $v^i = \dot{x}^i$ and $v = \sqrt{\sum (v^i)^2}$ is the particle speed (computed in the usual Euclidean metric on $\mathbb{R}^3$).

The component p_0 can be interpreted in terms of the energy of the particle. Following Definition 2.2, we define the energy of a relativistic particle by

$$\mathscr{E} = \frac{\partial L}{\partial \boldsymbol{v}} \boldsymbol{v} - L = \frac{mv^2}{\sqrt{1 - \dfrac{v^2}{c^2}}} + mc^2\sqrt{1 - \frac{v^2}{c^2}} = \frac{mc^2}{\sqrt{1 - \dfrac{v^2}{c^2}}}. \tag{8.11}$$

At $v = 0$ we obtain the *rest energy* $\mathscr{E}_0$ of the particle,

$$\mathscr{E}_0 = mc^2,$$

which is proportional to the mass of the particle. At small velocities v 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 non-relativistic particle.

Comparing the formula for p_0 with (8.11), we see that $p_0 = -\mathscr{E}/c$.

It is common in physics to define the energy-momentum 4-vector $p = (p^\mu) \in T\mathbb{R}^4$ by

$$p^\mu = -\eta^{\mu\nu} p_\nu = -\eta^{\mu\nu} \frac{\partial L}{\partial(\partial_t x^\nu)} \tag{8.12}$$

(note the minus sign!), which gives us $p = (p^0, \boldsymbol{p})$, where

$$p^0 = \mathscr{E}/c = \frac{mc}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad \boldsymbol{p} = \frac{m\boldsymbol{v}}{\sqrt{1 - \frac{v^2}{c^2}}} \quad \text{and} \quad p^2 = \eta_{\mu\nu} p^\mu p^\nu = (p^0)^2 - \boldsymbol{p}^2 = m^2 c^2. \tag{8.13}$$

Equations (8.13) imply the following identities, which we will frequently use:

$$p^0 = \sqrt{\boldsymbol{p}^2 + m^2 c^2}, \qquad \boldsymbol{p} = \frac{p^0}{c} \boldsymbol{v}. \tag{8.14}$$

Remark 8.3. There are several reasons for introduction of negative sign in (8.12). First, it makes the component p^0 positive: $p^0 = \mathscr{E}/c$. More importantly, it means that the spacial momentum vector $\boldsymbol{p}$ is given by the familiar formula $\boldsymbol{p} = \frac{\partial L}{\partial \boldsymbol{v}}$, where, as in classical mechanics, we identify 3-dimensional vectors and covectors using the usual Euclidean norm in $\mathbb{R}^3$; this guarantees that in the limit $c \to \infty$, our formulas for $\boldsymbol{p}$ match non-relativistic formulas (see Example 2.1).

Remark 8.4. For fixed $P_0 \in M^4$ we define the classical action $S(P)$ (cf. Remark 1.8) as $-mc$ times the Minkowski length of the timelike interval between points P_0 and P. In other words,

$$S(P) = -mc \int_{P_0}^{P} ds,$$

where integration goes along the straight line connecting points P_0 and P in M^4. Denoting by x^μ the coordinates of P, we get from (8.8):

$$\frac{\partial S}{\partial x^\mu} = -mcu_\mu, \tag{8.15}$$

so

$$p_\mu = -\frac{\partial S}{\partial x^\mu}. \tag{8.16}$$

As was observed in Remark 4.6, there is a difference in sign in the formulas expressing the energy and momentum of a non-relativistic particle in terms of the classical action. However, in formula (8.16) the time and spacial components of the energy-momentum 4-vector are introduced in a relativistic invariant way.

8.3. Noether integrals

Here we give a more invariant derivation of integrals of motion. Recall that the Lagrangian (8.7) of a free relativistic particle is invariant under the Poincaré group action

$$x^\mu \mapsto x'^\mu = \Lambda^\mu_\nu x^\nu + a^\mu$$

on $\mathbb{R}^{1,3}$.

Since Poincaré group is 10-dimensional, according to Noether theorem — Theorem 2.5 in Chapter 2 — there are ten integrals of motion corresponding to the generators P_μ and $M_{\lambda\mu}$ of the Lie algebra $\mathfrak{p}$ of the Poincaré group in Section 7.3.

Namely, the Lagrangian L with arbitrary parameter τ along the world-line is

$$L = -mc|\partial_\tau x| = -mc\sqrt{\partial_\tau x^\mu \partial_\tau x_\mu}.$$

Therefore, the Noether integrals of motion for the abelian Lie algebra $\mathbb{R}^4$ are given by

$$I_\mu = \frac{\partial L}{\partial(\partial_\tau x^\mu)} = -\frac{mc\partial_\tau x_\mu}{|\partial_\tau x|},$$

and are reparametrization invariant.

Choosing $\tau = t$, we see that $I_\mu = p_\mu$ are exactly the components of the momentum covector defined by (8.9).

We can also define the integrals of motion corresponding to generators $M^{\mu\nu}$ of the Lorentz Lie algebra, defined in (7.15). Namely, the corresponding vector field on $\mathbb{R}^{1,3}$ is

$$X^{\mu\nu} = (M^{\mu\nu} \cdot x)^\sigma \frac{\partial}{\partial x^\sigma} = (\eta^{\sigma\nu} x^\mu - \eta^{\sigma\mu} x^\nu)\frac{\partial}{\partial x^\sigma},$$

and the Noether integral associated with this vector field is

$$I^{\mu\nu} = (\eta^{\sigma\nu} x^\mu - \eta^{\sigma\mu} x^\nu)\left(\frac{\partial L}{\partial \dot{x}^\sigma}\right) = (\eta^{\sigma\nu} x^\mu - \eta^{\sigma\mu} x^\nu)p_\sigma = x^\nu p^\mu - x^\mu p^\nu.$$

Thus we obtain components of the total angular momentum

$$-I^{12} = x^1 p^2 - x^2 p^1, \quad -I^{23} = x^2 p^3 - x^3 p^2, \quad -I^{31} = x^3 p^1 - x^1 p^3,$$

and the integrals of motion corresponding to Lorentz boosts

$$I^{0i} = p^0 x^i - x^0 p^i.$$

Note that though I^{0i} depend explicitly on t, they are still integrals of motion:

$$\frac{dI^{0i}}{dt} = v^i p^0 - cp^i = 0,$$

due to the relation (8.14).

8.4. The Hamiltonian formulation

Since in the special relativity all velocities are less than the speed of light, instead of a tangent bundle $T\mathbb{R}^3$ in Newtonian mechanics, one uses a subbundle of $T\mathbb{R}^3$ whose fibers are open balls $\mathbb{B}(0,c)$ of radius c in $\mathbb{R}^3$, centered at the origin. Of course, this fibre bundle is just the product $\mathbb{R}^3 \times \mathbb{B}(0,c)$. According to the Definition 4.2, the Legendre transform is the map $\tau_L : \mathbb{R}^3 \times \mathbb{B}(0,c) \to T^*\mathbb{R}^3$,

$$\mathbb{R}^3 \times \mathbb{B}(0,c) \ni (\boldsymbol{r}, \boldsymbol{v}) \mapsto (\boldsymbol{p}, \boldsymbol{r}) \in T^*\mathbb{R}^3,$$

where canonically conjugated momentum $\boldsymbol{p} = \tau_L(\boldsymbol{v})$ is given by[1]

$$\boldsymbol{p} = \frac{\partial L}{\partial \boldsymbol{v}} = \frac{m\boldsymbol{v}}{\sqrt{1 - \dfrac{\boldsymbol{v}^2}{c^2}}}, \tag{8.17}$$

---

[1]Here we again identify a momentum 3-covector $\left(\frac{\partial L}{\partial v^1}, \frac{\partial L}{\partial v^2}, \frac{\partial L}{\partial v^3}\right)$ — an element of $T^*\mathbb{R}^3$ — with a 3-vector $\boldsymbol{p}$ — an element of $T\mathbb{R}^3$ — using Euclidean inner product in $\mathbb{R}^3$ and not the metric η.

which is the formula (8.13). It is remarkable that τ_L is a diffeomorphism between $\mathbb{R}^3 \times \mathbb{B}(0,c)$ and $T^*\mathbb{R}^3$.

As in Section 4.2, $T^*\mathbb{R}^3$ is a symplectic manifold with the symplectic form

$$\omega = d\boldsymbol{p} \wedge d\boldsymbol{r} = dp^1 \wedge dx^1 + dp^2 \wedge dx^2 + dp^3 \wedge dx^3$$

and the Poisson bracket

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

The Euler-Lagrange equations for a free relativistic particle are the Hamilton's equations

$$\dot{\boldsymbol{p}} = \{\mathscr{H}, \boldsymbol{p}\} = -\frac{\partial \mathscr{H}}{\partial \boldsymbol{r}}, \quad \dot{\boldsymbol{r}} = \{\mathscr{H}, \boldsymbol{r}\} = \frac{\partial \mathscr{H}}{\partial \boldsymbol{p}} \tag{8.18}$$

with the Hamiltonian function $\mathscr{H} = \mathscr{E} \circ \tau_L^{-1}$. Explicitly,

$$\mathscr{H} = cp^0 = c\sqrt{\boldsymbol{p}^2 + m^2c^2},$$

and for small $\boldsymbol{p}$ we have

$$\mathscr{H} = mc^2\sqrt{1 + \frac{\boldsymbol{p}^2}{m^2c^2}} \simeq mc^2 + \frac{\boldsymbol{p}^2}{2m}.$$

8.5. Hamiltonian action of Poincaré group on $T^*\mathbb{R}^3$

Recall that in Section 5.3 we defined the action of the Galilean group G on the phase space $T^*\mathbb{R}^3$ of a free particle, and have shown that this is a Hamiltonian action of the central extension of G. In this section, we give a relativistic analog of this result: Poincaré group $\mathfrak{P}$ also acts naturally on the phase space $T^*\mathbb{R}^3$ of free relativistic particle, and its action is Hamiltonian.

Namely, as in Section 5.3, let $\mathscr{X}$ be the set of all classical trajectories of a free relativistic particle, the set of timelike straight lines in $\mathbb{R}^{1,3}$. Since the Poincaré group $\mathfrak{P}$ group acts on $\mathbb{R}^{1,3}$ preserving equations of motion of a free particle, this action descends to an action on $\mathscr{X}$; thus, we have a natural action of $\mathfrak{P}$ on $\mathscr{X}$. On the other hand, for each $(\boldsymbol{p}, \boldsymbol{r}) \in T^*\mathbb{R}^3$, there exists a unique solution satisfying the initial conditions $\boldsymbol{r}(0) = \boldsymbol{r}$, $\left.\frac{\partial L}{\partial \dot{\boldsymbol{r}}}\right|_{t=0} = \boldsymbol{p}$. Thus, we have an identification of the phase space and $\mathscr{X}$; therefore, the action of G on $\mathscr{X}$ automatically gives an action of $\mathfrak{P}$ on $T^*\mathbb{R}^3$.

It is easy to write this action explicitly in coordinates. Solutions of classical equations of motion for a free particle are straight lines; thus, every solution has the form

$$\boldsymbol{r}(t) = \boldsymbol{r} + t\boldsymbol{v}, \quad |\boldsymbol{v}| < c$$

so $\mathscr{X} = \{(\boldsymbol{r}, \boldsymbol{v})\} = \mathbb{R}^3 \times \mathbb{B}(0,c)$. According to (8.14), the correspondence between $T^*\mathbb{R}^3$ and $\mathscr{X}$ is given by

$$(\boldsymbol{p}, \boldsymbol{r}) \mapsto l = \left\{\left(\boldsymbol{r} + t\frac{c\boldsymbol{p}}{p^0}, t\right)\right\}_{t\in\mathbb{R}} \in \mathscr{X}. \tag{8.19}$$

Correspondingly, this defines the action of $\mathfrak{P}$ on $T^*\mathbb{R}^3$. We won't write down an explicit formula for that action since it is rather messy; we just note that this action is not linear. The action of $\mathrm{SO}(3) \subset \mathrm{SO}(1,3)$ is, of course, just the regular action of $\mathrm{SO}(3)$ on $T^*\mathbb{R}^3$.

It is also easy to see that the elements of $\mathfrak{P}$ preserve the Hamiltonian $\mathscr{H} = cp^0$ of a free relativistic particle, except for the pure Lorentz boosts which do not preserve $\mathscr{H}$.

Theorem 8.5. *The action of Poincaré group $\mathfrak{P}$ on the phase space $T^*\mathbb{R}^3$ of free relativistic particle with mass m is Hamiltonian.*

The Hamiltonian functions corresponding to generators J_i, K_i, P_i of the Lie algebra of the Poincaré group (see Section 7.3) are given by

$$\hat{J}_i = -\epsilon_{ijk}x^j p^k,$$
$$\hat{K}_i = x^i\sqrt{\boldsymbol{p}^2 + m^2c^2},$$
$$\hat{P}_0 = p^0 = \sqrt{\boldsymbol{p}^2 + m^2c^2},$$
$$\hat{P}_i = -p^i, \quad i = 1, 2, 3.$$

The Poisson brackets of these functions are given by

$$\{\hat{P}_i, \hat{P}_j\} = \{\hat{P}_i, \hat{P}_0\} = \{\hat{J}_i, \hat{P}_0\} = 0, \quad \{\hat{J}_i, \hat{J}_j\} = \epsilon_{ijk}\hat{J}_k, \tag{8.20}$$

$$\{\hat{K}_i, \hat{K}_j\} = -\epsilon_{ijk}\hat{J}_k, \quad \{\hat{J}_i, \hat{K}_j\} = \epsilon_{ijk}\hat{K}_k, \tag{8.21}$$

$$\{\hat{K}_i, \hat{P}_0\} = \hat{P}_i, \quad \{\hat{K}_i, \hat{P}_j\} = \delta_{ij}\hat{P}_0, \quad \{\hat{J}_i, \hat{P}_j\} = \epsilon_{ijk}\hat{P}_k. \tag{8.22}$$

This theorem is an analog of Theorem 5.6, in which we had established a similar result for the Galilean group. Note, however, that for the Galilean group we needed to introduce a central extension in order to get a Hamiltonian action. For the Poincaré group, this problem does not arise: the Poisson brackets of the generating functions exactly match the commutation relations (7.11)–(7.12) in the Poincaré Lie algebra $\mathfrak{p}$. This is to be expected, as $\mathfrak{P}$ has no non-trivial central extensions, see remark at the end of Section 7.4.

Proof. As we already have seen in Section 5.3, the action of the Euclidean subgroup E(3) of the Poincaré group $\mathfrak{P}$ is Hamiltonian with the Hamiltonian functions[2] $\hat{P}_i = -p^i$ and $\hat{J}_i = -\epsilon_{ijk}x^j p^k$, $i = 1, 2, 3$.

The one-parameter subgroup of translations in x^0 direction acts on $\mathscr{X}$ by $l \mapsto l + (s, 0, 0, 0)$, or $(\boldsymbol{r}, \boldsymbol{v}) \mapsto \left(\boldsymbol{r} - \frac{s\boldsymbol{v}}{c}, \boldsymbol{v}\right)$. Using (8.14), this gives the following action on $T^*\mathbb{R}^3$:

$$\boldsymbol{p} \mapsto \boldsymbol{p}, \quad \boldsymbol{r} \mapsto \boldsymbol{r} - \frac{s\boldsymbol{p}}{p^0}.$$

Thus the corresponding vector field is $\xi_0 = -\dfrac{\boldsymbol{p}}{p^0}\dfrac{\partial}{\partial \boldsymbol{r}}$. On the other hand we have

$$\frac{\partial p^0}{\partial \boldsymbol{p}} = \frac{\boldsymbol{p}}{p^0},$$

so the vector field $-\xi_0$ is generated by the Hamiltonian function $\hat{P}_0 = p^0 = \sqrt{\boldsymbol{p}^2 + m^2c^2}$.

Computation of Hamiltonian functions $\hat{K}_i$ for the generators of Lorentz boosts is more involved, so we outline it in the Exercise 8.2.

[2] Note that canonically conjugated momenta to x^i are components p^i of the 3-vector $\boldsymbol{p}$.

The Poisson brackets of the Hamiltonian functions are verified by direct computation; the key step is the following relation:

$$\{x^i, p^0\} = -\frac{p^i}{p^0}.$$

We leave the details of this computation to the reader. □

Remark 8.6. As in the non-relativistic case (see Remark 5.7), we can identify $\mathscr{X}$ with $T^*\mathbb{R}^3$ by setting $\boldsymbol{r}(t_0) = \boldsymbol{r}$, $\left.\dfrac{\partial L}{\partial \dot{\boldsymbol{r}}}\right|_{t=t_0} = \boldsymbol{p}$. The Hamiltonian functions will not change, except for those associated with the Lorentz boosts, which will become

$$\hat{K}_i = x^i\sqrt{\boldsymbol{p}^2 + m^2c^2} + ct_0\hat{P}_i = x^ip^0 - ct_0p^i.$$

Observables $\hat{K}_i = x^ip^0 - ct_0p^i$ are not integrals of motion, while $x^ip^0 - x^0p^i$ are Noether integrals.

We can also easily compute the Poisson brackets of the generators of Poincaré Lie algebra and coordinate functions on $\mathbb{R}^3$:

$$\{\hat{J}_i, x^j\} = \epsilon_{ijk}x^k, \tag{8.23}$$

$$c\{\hat{K}_i, x^j\} = x^i\{\mathscr{H}, x^j\} = cx^i\frac{p^j}{p^0}, \tag{8.24}$$

$$\{\hat{P}_i, x^j\} = -\delta_{ij}, \quad i, j = 1, 2, 3. \tag{8.25}$$

These Poisson brackets exemplify that $\mathbb{R}^6$ is a phase space of a relativistic particle.

8.6. No-interaction theorem

Recall that the interaction of particles is described by a field of force. In classical mechanics, one can say that particle creates a field around itself, a certain force then acts instantaneously on every other particle in the system. The field is just a mode of describing the interaction of particles. The situation changes drastically in special relativity, since in a given inertial frame of reference, a change in position of one particle influences other particles only after a certain period of time. The special relativity principle imposes a very strong restriction on Hamiltonian systems: it implies that the interaction of finitely many relativistic particles is not possible!

For precise formulation, we introduce the notion of a Hamiltonian system of n relativistic particles. It is a system with the phase space $\mathbb{R}^{6n} = (T^*\mathbb{R}^3)^n$, the canonical 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 $\mathscr{H}$. The Hamiltonian system $(\mathbb{R}^{6n}, \omega, \mathscr{H})$ describes n relativistic particles, if the principle of relativity holds in the following precise form.

a) There exists a set of ten generators of the Poincaré Lie algebra — ten functions $\hat{P}_0 = \mathscr{H}/c$, $\hat{P}_i$, $\hat{J}_i$ and $\hat{K}_i$ on $\mathbb{R}^{6n}$ with Poisson brackets (8.20)–(8.22).

b) The coordinates of each particle transform correctly under the Poincaré group — coordinates $\boldsymbol{r}_a$ and the generators of the Poincaré Lie algebra have Poisson brackets (8.23)–(8.25).

In addition, the system is called non-degenerate, if

$$\det\left\{\frac{\partial^2\mathscr{H}}{\partial p_a^i \partial p_b^j}\right\} \neq 0.$$

Theorem 8.7. *Let $(\mathbb{R}^{6n},\omega,\mathscr{H})$ be a non-degenerate system of n relativistic particles. 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

$$\begin{aligned}
\hat{\boldsymbol{P}} &= -\sum_{a=1}^{n} \tilde{\boldsymbol{p}}_a,\\
\mathscr{H} &= \sum_{a=1}^{n} c\sqrt{\tilde{\boldsymbol{p}}_a^2 + m_a^2 c^2},\\
\hat{J}_i &= -\sum_{a=1}^{n} \epsilon_{ijk} x_a^j \tilde{p}_a^k,\\
\hat{K}_i &= \sum_{a=1}^{n} x_a^i \sqrt{\tilde{\boldsymbol{p}}_a^2 + m_a^2 c^2}.
\end{aligned}$$

This result is a manifestation of the fundamental fact that relativistic invariant Hamiltonian systems of interacting particles in Minkowski spacetime should have infinitely many degrees of freedom, and the interaction is described by a field theory. As physicists say, the field itself becomes a physical reality. The basic example of classical field theory is the theory of electromagnetism, where charged relativistic particles interact with the external electromagnetic field. Another fundamental example in classical physics is the theory of gravity, where massive relativistic particle interacts with the external gravitational field.

8.7. Exercises

Exercise 8.1. Verify that Hamilton's equations (8.18) give equations of motion of a free relativistic particle.

Exercise 8.2. The goal of this exercise is to prove the formulas for the generators $\hat{K}_i$ of the Lorentz boosts given in Theorem 8.5. Denote by $\Lambda(\psi) = \exp(\psi K_1) \in \mathfrak{L}$ the one-parameter subgroup of Lorentz boosts in x^0x^1-plane:

$$\Lambda(\psi)x = (x^0\cosh\psi + x^1\sinh\psi, x^0\sinh\psi + x^1\cosh\psi, x^2, x^3), \quad \psi \in \mathbb{R}.$$

(1) Show that the action of $\Lambda(\psi)$ on $\mathscr{X}$ is given by $(\boldsymbol{r}, \boldsymbol{v}) \mapsto (\Lambda(\psi)(\boldsymbol{r}), \Lambda(\psi)(\boldsymbol{v}))$, where

$$\Lambda(\psi)(\boldsymbol{r}) = \left(\frac{cx^1}{\tilde{v}},\, x^2 - \frac{x^1 v^2 \sinh\psi}{\tilde{v}},\, x^3 - \frac{x^1 v^3 \sinh\psi}{\tilde{v}}\right), \qquad \tilde{v} = v^1 \sinh\psi + c\cosh\psi,$$

$$\Lambda(\psi)(\boldsymbol{v}) = \left(\frac{cv^1\cosh\psi + c^2\sinh\psi}{\tilde{v}},\, \frac{cv^2}{\tilde{v}},\, \frac{cv^3}{\tilde{v}}\right).$$

(2) Show that under identification $X \simeq T^*\mathbb{R}^3$ given by (8.19), the action of $\Lambda(\psi)$ on $T^*\mathbb{R}^3$ is given by

$$\Lambda(\psi)(\boldsymbol{r}) = \left(\frac{x^1 p^0}{\tilde{p}},\, x^2 - \frac{x^1 p^2 \sinh\psi}{\tilde{p}},\, x^3 - \frac{x^1 p^3 \sinh\psi}{\tilde{p}}\right), \qquad \tilde{p} = p^1 \sinh\psi + p^0\cosh\psi,$$

$$\Lambda(\psi)(\boldsymbol{p}) = (p^1\cosh\psi + p^0\sinh\psi,\, p^2, p^3).$$

(3) Show that the vector field ξ_{K_1} on $T^*\mathbb{R}^3$ is given by

$$\xi_{K_1} = \frac{x^1\boldsymbol{p}}{p^0}\frac{\partial}{\partial\boldsymbol{r}} - p^0\frac{\partial}{\partial p^1}$$

and dervie from it that the corresponding Hamiltonian function is $\hat{K}_1 = x^1\sqrt{\boldsymbol{p}^2 + m^2c^2}$. In a similar way one gets the expressions for $\hat{K}_2$, $\hat{K}_3$.

Exercise 8.3. Prove the no-interaction theorem for $n = 1$.

Chapter 9

Spinors and Dirac Operator

In this section, we introduce spinors — special representations of group $\mathrm{SO}(n)$ which play important role in physics. Informally, they are sometimes described as "square roots of vectors"; exact meaning of that is clarified below.

Spinors are especially useful in constructing the Dirac operator — "square root of Laplace operator", described in Section 9.5 below.

9.1. Spin group and definition of spinors

Let $G = \mathrm{SO}(n)$; for simplicity, we will always assume that $n \geq 3$. It is a compact semisimple Lie group, which is a compact form of the complex semisimple Lie group $G_{\mathbb{C}} = \mathrm{SO}(n, \mathbb{C})$. We will denote by $\mathfrak{g} = \mathfrak{so}(n, \mathbb{R})$ and $\mathfrak{g}_{\mathbb{C}} = \mathfrak{so}(n, \mathbb{C})$ Lie algebras of G, $G_{\mathbb{C}}$ respectively. As is well-known, $\mathfrak{g}_{\mathbb{C}}$ is a simple Lie algebra (except for $n = 4$, in which case it is only semisimple); if $n = 2k$ is even, the corresponding Dynkin diagram is D_k, and if $n = 2k + 1$ is odd, then the Dynkin diagram is of type B_k, as shown below. For $n = 4$, the Dynkin diagram is $A_1 \sqcup A_1$.

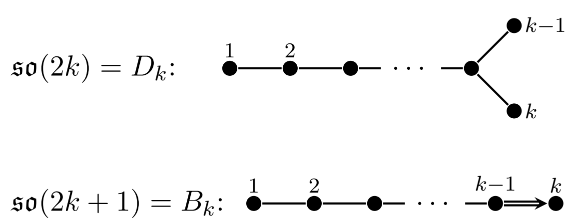

Figure 9.1. Dynkin diagrams for $\mathfrak{so}(n)$. Numbers on vertices are indices used to number roots.

Group G (and thus Lie algebra $\mathfrak{g}$) naturally acts on $\mathbb{R}^n$; this is usually called the defining, or tautological, representation of G. It is convenient to pass to complex representation $V = \mathbb{C}^n$, which is naturally a representation of $G_{\mathbb{C}}$ and $\mathfrak{g}_{\mathbb{C}}$; abusing the language, we will also refer to it as the tautological representation.

The tautological representation V is irreducible (one can show that for $n > 4$, the tautological representations is the irreducible representation whose highest weight is the fundamental weight ω_1 if the roots are indexed as in Figure 9.1).

We will be interested in the irreducible representations corresponding to the fundamental weights from the opposite end of the Dynkin diagram. Namely, for $\mathfrak{so}(2k+1) = B_k$, consider the fundamental weight ω_k corresponding to the righmost node in the diagram and denote the corresponding complex irreducible representation by S; for $\mathfrak{so}(2k) = D_k$, consider the fundamental weights ω_{k-1}, ω_k corresponding to the two endpoins of the "forked" end of the Dynkin diagram and denote the corresponding irreducible complex representations S^+, S^-.

Definition 9.1. Representations S (for odd n), $S^{\pm}$ (for even n) are called *spinor representations* of $\mathfrak{so}(n, \mathbb{C})$.

These representations have a number of useful properites. Most importantly, one can show that the fundamental representation V is a subrepresentation of $S \otimes S$ (for odd n) or $S^+ \otimes S^-$ (for even n); for example, for $n = 3$ one has $V = \mathrm{Sym}^2 S$ (see more low-dimensional examples below). For this reason, one frequently refers to spinors as square roots of vectors.

An important feature of spinor representations is that they do not lift to a representation of $\mathrm{SO}(n)$. Recall that the group $\mathrm{SO}(n)$ is not simply-connected; instead, we have $\pi_1(\mathrm{SO}(n)) = \pi_1(\mathrm{SO}(n, \mathbb{C})) = \mathbb{Z}_2$ (for $n \geq 3$). Thus, one can write

$$\mathrm{SO}(n) = \mathrm{Spin}(n)/\mathbb{Z}_2,$$
$$\mathrm{SO}(n, \mathbb{C}) = \mathrm{Spin}^{\mathbb{C}}(n)/\mathbb{Z}_2$$

where $\mathrm{Spin}(n)$ is the universal cover of $\mathrm{SO}(n)$ and thus is a connected simply-connected Lie group with Lie algebra $\mathfrak{so}(n)$; similarly, $\mathrm{Spin}^{\mathbb{C}}(n)$ is a connected simply-connected complex Lie group with Lie algebra $\mathfrak{so}(n, \mathbb{C})$. These groups are called the *spin groups*. It is easy to see that $\mathrm{Spin}(n)$ is a compact real form of $\mathrm{Spin}^{\mathbb{C}}(n)$. In some low-dimensional cases, these groups admit more explicit description, see the next section.

By general results of Lie theory, any representation of $\mathfrak{so}(n)$ lifts to a representation of $\mathrm{Spin}(n)$; however, it turns out that for spinor representations, the non-trivial element of $\mathbb{Z}_2 \subset \mathrm{Spin}(n)$ acts on S by -1 and thus these representations do not descend to representations of $\mathrm{SO}(n)$ or $\mathrm{SO}(n, \mathbb{C})$.

9.2. Low-dimensional examples

In some special cases, the spin group and spinor representation can be described very explicitly.

If $n = 3$, then it is well-known that $\mathrm{SO}(3) \cong \mathrm{SU}(2)/\mathbb{Z}_2$ (compare with Exercise 1.7) and thus in this case $\mathrm{Spin}(3) = \mathrm{SU}(2)$, $\mathrm{Spin}^{\mathbb{C}}(3) = \mathrm{SL}(2, \mathbb{C})$. For this group, the spinor representation $S = \mathbb{C}^2$ is the defining representation of $\mathrm{SU}(2)$; equivalently, it can also be described as the irreducible representation of $\mathfrak{sl}(2, \mathbb{C})$ with highest weight 1, or in the physical language, with spin $\frac{1}{2}$. The tautological representation $V = \mathbb{C}^3$ of $\mathrm{SO}(3,\mathbb{C})$ is the irreducible representation with highest weight 2, or with spin 1, and it is related to the spinor representation by

$$V = \mathrm{Sym}^2 S.$$

If $n = 4$, then the Dynkin diagram is $D_2 = A_1 \sqcup A_1$, and the corresponding Lie algebra is $\mathfrak{so}(4,\mathbb{C}) \simeq \mathfrak{sl}(2,\mathbb{C}) \oplus \mathfrak{sl}(2,\mathbb{C})$. Therefore, in this case we have $\mathrm{Spin}^{\mathbb{C}}(4) = \mathrm{SL}(2,\mathbb{C}) \times \mathrm{SL}(2,\mathbb{C})$, $\mathrm{Spin}(4) = \mathrm{SU}(2) \times \mathrm{SU}(2)$. Explicit construction of isomorphism

$$\mathrm{SO}(4) \simeq (\mathrm{SU}(2) \times \mathrm{SU}(2))/\mathbb{Z}_2$$

was given in Exercise 7.3.

In this case the spinor representations $S^{\pm}$ are both 2-dimensional. On each of them, one of the copies of $\mathrm{SU}(2)$ acts trivially and the other acts by the tautological representation:

$$S^+ \simeq \mathbb{C}^2 \otimes \mathbb{C},$$
$$S^- \simeq \mathbb{C} \otimes \mathbb{C}^2$$

where we denote by $\mathbb{C}^2$ the tautological representation of $\mathrm{SL}(2,\mathbb{C})$.

One can check that the relation between the tautological representation $V = \mathbb{C}^4$ of $\mathrm{SO}(4,\mathbb{C})$ and the spinor representations is given by

$$V = S^+ \otimes S^-.$$

One can also give an explicit description of the spin group and spinor representations for $n = 5$ (in which case the spin group is the symplectic group due to equality of Dynkin diagrams $B_2 = C_2$) and for $n = 6$ (in which case the Dynkin diagram is $D_3 = A_3$ so that $\mathrm{Spin}(6) = \mathrm{SU}(4)$).

9.3. Clifford algebra construction

Spinor representations admit an explicit construction in terms of so-called Clifford algebra. In this section we give an overview of this construction.

Let V be a real vector space with a symmetric bilinear form B; we denote by $Q(v) = B(v,v)$ the corresponding quadratic form. Let $n = \dim V$.

Definition 9.2. Clifford algebra $\mathrm{Cl}(V)$ is the unital associative algebra over $\mathbb{R}$ generated by vectors $v \in V$ with defining relations

$$v \cdot w + w \cdot v = 2B(v,w) \cdot 1. \tag{9.1}$$

It is easy to check that these relations are equivalent to relations

$$v^2 = Q(v) \cdot 1, \qquad v \in V.$$

Equivalently, one can define the Clifford agebra by

$$\mathrm{Cl}(V) = T(V)/J$$

where $T(V)$ is the tensor algebra of V and J is the two-sided ideal generated by elements $v \otimes v - Q(v) \cdot 1$, $v \in V$.

In a similar way one defines the Clifford algebra for a complex vector space with a symmetric bilinear form B, using the same defining relations (9.1). In this case, $\mathrm{Cl}(V)$ is a complex associative algebra.

In particular, we will denote by $\mathrm{Cl}(n)$ the Clifford algebra corresponding to an n-dimensional complex vector space with a non-degenerate quadratic form. Since any two non-degenerate quadratic forms are equivalent over $\mathbb{C}$, this implies that up to isomorphism, $\mathrm{Cl}(n)$ doesn't depend on the choice of the quadratic form.

Clifford algebra has a number of useful properties, which we summarize below. Unless specified otherwise, these properties hold both in real and complex case.

(1) It is a filtered algebra: if we denote by $F_p \subset \mathrm{Cl}(V)$ the subspace spanned by products of at most p elements of V, then
$$\mathbb{C} = F_0 \subset F_1 \subset \dots, \qquad \bigcup_p F_p = \mathrm{Cl}(V),$$
and $F_p \cdot F_q \subset F_{p+q}$.

The corresponding graded algebra $\mathrm{gr}(\mathrm{Cl}(V)) = \bigoplus F_p/F_{p-1}$ is canonically isomorphic to the exterior algebra ΛV. In particular, this implies that $\dim \mathrm{Cl}(V) = 2^n$ if $\dim V = n$.

(2) The Clifford algebra is naturally $\mathbb{Z}_2$-graded: $\mathrm{Cl}(V) = \mathrm{Cl}_0(V) \oplus \mathrm{Cl}_1(V)$; this grading is uniquely determined by the requirement that elements $v \in V$ are odd: $v \in \mathrm{Cl}_1(V)$. As usual, we will refer to Cl_0, Cl_1 as even and odd parts of Cl.

(3) Given two vector spaces V_1, V_2 with quadratic forms, one has a natural isomorphism $\mathrm{Cl}(V_1 \oplus V_2) \simeq \mathrm{Cl}(V_1) \otimes \mathrm{Cl}(V_2)$, where the algebra structure on the tensor product is given by the usual sign rule:
$$(v_1 \otimes v_2) \cdot (w_1 \otimes w_2) = (-1)^{pq}\, v_1 w_1 \otimes v_2 w_2, \qquad v_2 \in \mathrm{Cl}_p, w_1 \in \mathrm{Cl}_q\,.$$

This allows one to give an explicit description of the Clifford algebra. Indeed, given a vector space W of dimension k, consider $V = W \oplus W^*$ with the bilinear form given by
$$\begin{aligned} &B(w,f) = B(f,w) = \tfrac{1}{2}\langle f, w\rangle, \qquad w \in W,\ f \in W^*,\\ &B(w_1,w_2) = B(f_1,f_2) = 0, \qquad w_i \in W,\ f_i \in W^* \end{aligned}$$
where $\langle\ ,\ \rangle$ is the natural pairing between W and W^*. It is easy to see that this form is non-degenerate.

Theorem 9.3. *Let $V = W \oplus W^*$ as above. Then one has a natural algebra isomorphism $\mathrm{Cl}(V) \simeq \mathrm{End}(\Lambda W)$.*

Proof. For $w \in W$, $f \in W^*$, consider operators $a_w, i_f \colon \Lambda W \to \Lambda W$ defined by
$$\begin{aligned} &a_w \colon \eta \mapsto w \wedge \eta,\\ &i_f \colon w_1 \wedge \dots \wedge w_m \mapsto \sum_{i=1}^{m} (-1)^{i-1} \langle f, w_i\rangle w_1 \wedge \dots \wedge \widehat{w_i} \wedge \dots \wedge w_m, \end{aligned}$$
where as usual, the hat means that the corresponding factor in the wedge product should be skipped. Explicit computation shows that these operators satisfy the relations of the Clifford algebra $\mathrm{Cl}(W \oplus W^*)$ and thus define a morphism $\mathrm{Cl}(W \oplus W^*) \to \mathrm{End}(\Lambda W)$. It

is easy to show that this morphism is surjective. Since $\dim \operatorname{End}(\Lambda W) = (2^k)^2 = 2^{2k} = \dim \operatorname{Cl}(W \oplus W^*)$, this map is an isomorphism. □

Note that the vector space ΛW, and thus the algebra $\operatorname{End}(\Lambda W)$, are naturally $\mathbb{Z}_2$ graded:

$$\operatorname{End}(\Lambda W)_0 = \operatorname{End}(\Lambda^{even} W) \oplus \operatorname{End}(\Lambda^{odd} W)$$
$$\operatorname{End}(\Lambda W)_1 = \operatorname{Hom}(\Lambda^{even} W, \Lambda^{odd} W) \oplus \operatorname{Hom}(\Lambda^{odd} W, \Lambda^{even} W).$$

It is easy to see that the isomorphism $\operatorname{Cl}(W \oplus W^*) \simeq \operatorname{End}(\Lambda W)$ constructed above is an isomorphism of $\mathbb{Z}_2$-graded algebras.

Since any two non-degenerate quadratic forms over $\mathbb{C}$ are equivalent, we immediately get the following corollary.

Theorem 9.4. *Denote by* $\operatorname{Mat}_n(\mathbb{C})$ *the algebra of complex* $n \times n$ *matrices.*

(1) *If* $n = 2k$ *is even, then* $\operatorname{Cl}(n) \simeq \operatorname{Mat}_{2^k}(\mathbb{C})$ *(as a non-graded algebra) and* $\operatorname{Cl}_0(n) \simeq \operatorname{Mat}_{2^{k-1}}(\mathbb{C}) \oplus \operatorname{Mat}_{2^{k-1}}(\mathbb{C})$.
(2) *If* $n = 2k+1$ *is odd, then* $\operatorname{Cl}(n) = \operatorname{Cl}(1) \otimes \operatorname{Cl}(2k)$. *In this case,* $\operatorname{Cl}(n) \simeq \operatorname{Mat}_{2^k}(\mathbb{C}) \oplus \operatorname{Mat}_{2^k}(\mathbb{C})$ *as a non-graded algebra, and* $\operatorname{Cl}_0(n) \simeq \operatorname{Mat}_{2^k}(\mathbb{C})$.

Indeed, the first part is immediate from Theorem 9.3. To prove the second part, note that $\operatorname{Cl}(1) = \mathbb{C}[\varepsilon]/(\varepsilon^2 = 1)$ is isomorphic as a non-graded algebra to $\mathbb{C} \oplus \mathbb{C}$, and for any $\mathbb{Z}_2$ graded algebra A, we have $(\operatorname{Cl}(1) \otimes A)_0 = A_0 \oplus \varepsilon A_1 \simeq A$ as an (ungraded) algebra.

Since it is well known that the matrix algebra $\operatorname{Mat}_d(\mathbb{C})$ has a unique simple module, namely the tautological representation $\mathbb{C}^d$, Theorem 9.4 immediately implies the following result.

Corollary 9.5. *Let* $n = 2k$ *be even. Then the algebra* $\operatorname{Cl}_0(n)$ *has two different simple modules, each of dimension* 2^{k-1}. *For odd* $n = 2k+1$, $\operatorname{Cl}_0(n)$ *has a unique simple module of dimension* 2^k.

The relevance of Clifford algebra to representations of $\mathrm{SO}(n, \mathbb{C})$ is explained by the following simple fact. Since the action of the group $\mathrm{SO}(V)$ preserves the quadratic form on V, it also gives an action of $\mathrm{SO}(V)$ by automorphisms on the Clifford algebra $\operatorname{Cl}(V)$, and an action of $\mathfrak{so}(V)$ by parity-preserving derivations of $\operatorname{Cl}(V)$ which we will denote by $x \mapsto a \cdot x$, $x \in \operatorname{Cl}(V)$, $a \in \mathfrak{so}(V)$.

In the case when $\operatorname{Cl}(V)$ is a matrix algebra, since it is known that every derivation of $\operatorname{Mat}_n(\mathbb{C})$ is inner, it gives a map $f\colon \mathfrak{so}(V) \to \operatorname{Cl}(V)$ such that for $a \in \mathfrak{so}(V)$, $x \in \operatorname{Cl}(V)$ we have

$$a \cdot x = [f(a), x].$$

This only defines $f(a)$ up to addition of a scalar which we can fix by requiring that $f(x) \in \operatorname{Cl}(V) \simeq \operatorname{Mat}_n(\mathbb{C})$ is traceless (or, equivalently, $f(a) \in [\operatorname{Cl}(V), \operatorname{Cl}(V)]$). This also guarantees that $f([a, b]) = [f(a), f(b)]$; in other words, $f\colon \mathfrak{so}(V) \to \operatorname{Cl}(V)$ is a morphism of Lie algebras. Moreover, since the derivation $x \mapsto a \cdot x$ must be parity preserving, it follows that $f(a) \in \operatorname{Cl}_0(V)$ for any a.

In the case when $\operatorname{Cl}(V)$ is direct sum of two matrix algebras, this argument requires a little modification which we leave to the reader. However, one can also define the morphism f explicitly as follows.

Lemma 9.6. *Let $V = \mathbb{C}^n$ with the standard symmetric bilinear form $B(x, y) = \sum_{i=1}^n x_i y_i$. Let e_i be the standard basis of $\mathbb{C}^n$ and E_{ij} the corresponding basis of matrix units in $\mathrm{Mat}_n(\mathbb{C})$.*

Define a linear map $f\colon \mathfrak{so}(n,\mathbb{C}) \to \mathrm{Cl}_0(V)$ by

$$f\colon M_{ij} \mapsto -\tfrac{1}{4}[e_i, e_j], \tag{9.2}$$

where $M_{ij} = E_{ji} - E_{ij}$ are the generators of $\mathfrak{so}(n,\mathbb{C})$, see Exercise 2.6.

Then

1. *For any $a \in \mathfrak{so}(n,\mathbb{C})$, $x \in \mathrm{Cl}(V)$, we have $a \cdot x = [f(a), x]$.*
2. *The map f extends to a morphism of associative algebras $U\mathfrak{so}(n,\mathbb{C}) \mapsto \mathrm{Cl}_0(n)$, where $U\mathfrak{so}(n,\mathbb{C})$ is the universal enveloping algebra of $\mathfrak{so}(n,\mathbb{C})$.*

The proof is by explicit computation which we leave to the reader. Note that for $i \neq j$, elements e_i, e_j anticommute in the Clifford algebra, so one can also write (9.2) as

$$M_{ij} \mapsto -\tfrac{1}{2}e_i e_j, \qquad i \neq j.$$

This lemma implies that any module over $\mathrm{Cl}_0(n)$ is automatically a module over $U\mathfrak{so}(n,\mathbb{C})$, and thus is a representation of $\mathfrak{so}(n,\mathbb{C})$ and of the spin group $\mathrm{Spin}^{\mathbb{C}}(n)$. A more general version of this lemma, which works for arbitrary non-degenerate symmetric bilinear form B, is given in Exercise 9.3.

We can now use this lemma to give a construction of spinor representations.

Theorem 9.7. *Let $n = 2k$ be even. Then the two simple modules over the Clifford algebra $\mathrm{Cl}_0(n)$ constructed in Corollary 9.5, considered as $\mathfrak{so}(n,\mathbb{C})$ modules via map (9.2), are exactly the spinor modules $S^{\pm}$ described in Definition 9.1.*

Similarly, for $n = 2k+1$, the unique simple module over $\mathrm{Cl}_0(n)$ considered as $\mathfrak{so}(n,\mathbb{C})$ module is the spinor module S.

We omit the proof of this theorem.

One can also use the Clifford algebra and the group $\mathrm{Cl}_0^{\times}$ of invertible elements in Cl_0 to give an explicit construction of the spin group $\mathrm{Spin}^{\mathbb{C}}(n)$. We omit this construction as it is not necessary for our purposes.

9.4. Lorentz group and Weyl spinors

So far we were mostly discussing spinor representations of the complex group $\mathrm{SO}(n,\mathbb{C})$ or its compact form $\mathrm{SO}(n)$. However, one can also consider spinor representations for other real forms of $\mathrm{SO}(n,\mathbb{C})$. For example, since $\mathrm{SO}(p,q)$ is a real form of $\mathrm{SO}(p,q;\mathbb{C}) \simeq \mathrm{SO}(p+q,\mathbb{C})$, we can restrict a spinor representation S of $\mathrm{SO}(n,\mathbb{C})$ to $\mathrm{SO}(p,q)$, $n = p+q$.

We will be mostly interested in the restricted Lorentz group $\mathfrak{L}_+^{\uparrow}$; recall that it is the connected component of $\mathrm{O}(1,3)$ (see Section 7.2). Combining the inclusion $\mathfrak{L}_+^{\uparrow} \subset \mathrm{SO}(1,3) \subset \mathrm{SO}(1,3;\mathbb{C})$ with the isomorphism $\mathrm{SO}(1,3;\mathbb{C}) \cong \mathrm{SO}(4,\mathbb{C})$, we get an inclusion

$$\mathfrak{L}_+^{\uparrow} \subset \mathrm{SO}(4,\mathbb{C}).$$

Recall that $\mathrm{Spin}^{\mathbb{C}}(4) \simeq \mathrm{SL}(2,\mathbb{C}) \times \mathrm{SL}(2,\mathbb{C})$; thus, we have the universal cover $\mathrm{SL}(2,\mathbb{C}) \times \mathrm{SL}(2,\mathbb{C}) \to \mathrm{SO}(4,\mathbb{C})$ (which is two-fold); restricting it to $\mathfrak{L}_+^{\uparrow} \subset \mathrm{SO}(4,\mathbb{C})$ we get a two-fold cover of $\mathfrak{L}_+^\uparrow$. It can be shown (we skip this argument) that this is the universal cover of $\mathfrak{L}_+^\uparrow$. On the other hand, we had shown in Exercise 7.4 that the universal cover of $\mathfrak{L}_+^\uparrow$ is $\mathrm{SL}(2,\mathbb{C})$ (considered as a real group). Thus, we have a commutative diagram

$$\begin{array}{ccc} \mathrm{SL}(2,\mathbb{C})_{\mathbb{R}} & \subset & \mathrm{SL}(2,\mathbb{C}) \times \mathrm{SL}(2,\mathbb{C}) \\ \downarrow & & \downarrow \\ \mathfrak{L}_+^\uparrow & \subset & \mathrm{SO}(1,3;\mathbb{C}) \cong \mathrm{SO}(4,\mathbb{C}) \end{array}$$

in which horizontal lines are real forms of complex groups, and vertical arrows are (two-fold) universal covers.

By results of the previous section, the group $\mathrm{Spin}^{\mathbb{C}}(4) \cong \mathrm{SL}(2,\mathbb{C}) \times \mathrm{SL}(2,\mathbb{C})$ has two spinor representations:

$$S^+ \cong \mathbb{C}^2 \otimes \mathbb{C},$$
$$S^- \cong \mathbb{C} \otimes \mathbb{C}^2,$$

called the *Weyl spinors* (the left-handed and right-handed spinors).

Definition 9.8. The module $\hat{S} = S^+ \oplus S^-$ over $\mathrm{Spin}^{\mathbb{C}}(4)$ is called *Dirac bispinor* module.

Note that it is different from the defining (or tautological) representation V of $\mathrm{SO}(4,\mathbb{C})$: the tautological representation is given by $V = S^+ \otimes S^-$, while Dirac bispinor representation is $\hat{S} = S^+ \oplus S^-$ and is only defined for the universal cover $\mathrm{Spin}^{\mathbb{C}}(4)$, not for $\mathrm{SO}(4,\mathbb{C})$.

At the level of Lie algebras, Dirac bispinor module can be described through Clifford algebra construction: by Lemma 9.6, we have a morphism

$$f\colon U\mathfrak{so}(4,\mathbb{C}) \to \mathrm{Cl}_0(4) \subset \mathrm{Cl}(4).$$

In this case, $\mathrm{Cl}(4) \cong \mathrm{Mat}_4(\mathbb{C})$ (as an ungraded algebra) and has a unique nonzero simple module $\hat{S} \cong \mathbb{C}^4$ so that $\mathrm{Cl}(4) \cong \mathrm{End}\,\hat{S}$. Restricting it to $\mathrm{Cl}_0(4) \cong \mathrm{Mat}_2(\mathbb{C}) \oplus \mathrm{Mat}_2(\mathbb{C})$, we get $\hat{S} = S^+ \oplus S^-$, where $S^\pm$ is the defining module over the two copies of $\mathrm{Mat}_2(\mathbb{C})$; and pulling back $S^\pm$ via f to $\mathfrak{so}(4,\mathbb{C})$ gives us the spinor modules.

For future use, we can give explicit formulas for the action of the Clifford algebra and of $\mathfrak{so}(1,3;\mathbb{C})$ in the spinor representation. Let γ_μ, $\mu = 0,1,2,3$, be the generators of the Clifford algebra $\mathrm{Cl}(1,3)$ (over $\mathbb{C}$):

$$\gamma_\lambda \gamma_\mu + \gamma_\mu \gamma_\lambda = 2\eta_{\lambda\mu} \tag{9.3}$$

where $\eta = \mathrm{diag}(1,-1,-1,-1)$.

Lemma 9.9. *The following formulas define a 4-dimensional representation of* $\mathrm{Cl}(1,3)$*:*

$$\gamma_\mu \mapsto \begin{pmatrix} 0 & \sigma_\mu \\ \sigma^\mu & 0 \end{pmatrix} \tag{9.4}$$

where σ_μ are Pauli matrices (see (1.24)) and $\sigma^\mu = \eta^{\mu\nu}\sigma_\nu$, so that

$$\sigma^0 = \sigma_0, \qquad \sigma^i = -\sigma_i, \qquad i = 1,2,3.$$

The proof is by explicit computation which easily follows from

$$\begin{aligned}&\sigma_\mu^2 = I,\\ &[\sigma_0, \sigma_a] = 0,\\ &[\sigma_a, \sigma_b] = 2i\epsilon_{abc}\sigma_c,\\ &\sigma_a\sigma_b + \sigma_b\sigma_a = 0, \quad a \neq b.\end{aligned}$$

Since we already know that $\mathrm{Cl}(1,3)$ is isomorphic to the matrix algebra $\mathrm{Mat}_4(\mathbb{C})$ and thus has a unique non-trivial module $\hat{S}$ which has dimension 4, we see that in a suitable basis action of $\mathrm{Cl}(1,3)$ in $\hat{S}$ is given by (9.4).

In physics literature, 4×4 matrices γ_μ satisfying (9.3) are called *Dirac γ-matrices*. The matrices (9.4) give the so-called *Weyl* or *chiral basis* of γ-matrices; the *Dirac basis* is given by the choice $\gamma_0 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}$ and the same matrices γ_i, $i = 1, 2, 3$.

We also have an explicit formula for the morphism $U\mathfrak{so}(1,3) \to \mathrm{Cl}(1,3)$, obtained by a minor modification of Lemma 9.6 (see Exercise 9.3):

$$M_{\lambda\mu} \mapsto -\tfrac{1}{4}[\gamma_\lambda, \gamma_\mu]. \tag{9.5}$$

In terms of the generators of $\mathfrak{so}(1,3)$, introduced in Section 7.3, we have

$$K_a \mapsto -\tfrac{1}{2}\gamma_0\gamma_a, \quad J_a \mapsto -\tfrac{1}{4}\epsilon_{abc}[\gamma_b, \gamma_c].$$

9.5. Dirac equation

Let $M \cong \mathbb{R}^{1,3}$ be the Minkowski spacetime. Consider the space of functions $C^\infty(M, \hat{S})$ with values in the Dirac bispinor module defined in the previous section. As before, we use notation γ_μ for generators of $\mathrm{Cl}(1,3)$; they act in $\hat{S}$ as described in Lemma 9.9.

Definition 9.10. Differential operator on $C^\infty(M, \hat{S})$ defined by the formula below is called the *Dirac operator*

$$\not{\partial} = \gamma^\mu\partial_\mu$$

where, as usual, we denote $\partial_\mu = \dfrac{\partial}{\partial x^\mu}$ and $\gamma^\mu = \eta^{\mu\nu}\gamma_\nu$.

From (9.3) we immediately get

$$\gamma^\lambda\gamma^\mu + \gamma^\mu\gamma^\lambda = 2\eta^{\lambda\mu} \tag{9.6}$$

which gives us the following lemma.

Lemma 9.11. *Dirac operator satisfies*

$$\not{\partial}^2 = \partial_\mu\partial^\mu I_4 = \left(\partial_0^2 - \sum_{i=0}^{3}\partial_i^2\right) I_4,$$

where I_4 is 4×4 identity matrix.

The scalar operator in the right-hand side is commonly called the *D'Alembert operator* (or simply box operator) and denoted by $\Box$:

$$\Box = \partial_\mu \partial^\mu = \partial_0^2 - \sum_{i=0}^{3} \partial_i^2. \tag{9.7}$$

The equation

$$(i\partial\!\!\!/ - m)\psi = 0, \tag{9.8}$$

is called the *Dirac equation*. Here $\psi \in C^\infty(M, \hat{S})$, the classical spinor, describes spin $\frac{1}{2}$ particle of mass m.

Using $\hat{S} = S^+ \oplus S^-$, we can write $\psi = \psi^+ + \psi^-$, $\psi^\pm \in C^\infty(M, S^\pm)$. The components $\psi^\pm$ are called Weyl spinors; note that since $S^\pm$ are not modules over the Clifford algebra, in general we can not restrict Dirac equation to $\psi^\pm$. Instead, observe that Lemma 9.11 implies the following fundamental property

$$(m + i\partial\!\!\!/)(m - i\partial\!\!\!/) = (\Box + m^2)I_4$$

which shows that each spinor component $\psi^\pm$ satisfies the Klein-Gordon equation[1]

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

Thus Dirac equation can be thought of as the square root of the Klein-Gordon equation.

In the massless case $m = 0$, the Dirac equation decouples into the *Weyl equation*

$$\sigma^\mu \partial_\mu \psi^+ = (I\partial_0 - \sigma_1\partial_1 - \sigma_2\partial_2 - \sigma_3\partial_3)\psi^+ = 0,$$

and its conjugate equation

$$\sigma_\mu \partial_\mu \psi^- = (I\partial_0 + \sigma_1\partial_1 + \sigma_2\partial_2 + \sigma_3\partial_3)\psi^- = 0$$

for the left-handed (respectively right-handed) Weyl spinors $\psi^\pm \in C^\infty(M, S^\pm)$.

9.6. Exercises

Exercise 9.1. Let γ_μ be generators of $\mathrm{Cl}(1,3)$ as in (9.3). Define $\gamma_5 = \gamma_0\gamma_1\gamma_2\gamma_3$. Prove that $\gamma_5^2 = -1$ and

$$\gamma_5\gamma_\mu + \gamma_\mu\gamma_5 = 0, \quad \mu = 0, 1, 2, 3.$$

Exercise 9.2. Prove the so-called Pauli lemma: if γ_μ and γ'_μ are two sets of 4×4 matrices satisfying (9.3), then there is a non-degenerate 4×4 matrix S such that

$$S\gamma_\mu S^{-1} = \gamma'_\mu.$$

Equivalently, Clifford algebra $\mathrm{Cl}(1,3)$ has a unique (up the equivalence) 4-dimensional representation.

Exercise 9.3. Prove more general form of Lemma 9.6: if η is a symmetic non-degenerate bilinear form on a finite-dimensional complex vector space V, and $M_{v_1,v_2} \in \mathfrak{so}(\eta)$ is defined as in Exercise 7.1, then the linear map $f\colon M_{v_1,v_2} \mapsto -\frac{1}{4}[v_1, v_2]$ extends to a morphism of associative algebras $U\mathfrak{so}(\eta) \to \mathrm{Cl}_0(V)$, satisfying $a \cdot x = [f(a), x]$.

[1]It should be better called Klein-Fock-Gordon equation.

Chapter 10

Notes and References

We refer the reader to the classic textbooks [**Arn1989**] and [**LL1976**], which are written, respectively, from mathematics and physics perspectives. The elegance of [**LL1976**] is supplemented by the attention to detail in [**Gol1980**], another physics classic.

The treatise [**AM1978**] and the encyclopedia surveys [**AG1990**], [**AKN1997**] provide a comprehensive exposition of classical mechanics, which includes the history of the subject, references to the classical works and to the recent development.

Arnold's book [**Arn1989**] is based on his lectures in Moscow State University in 1967-69, where he was the first to use systematically the language of differential symplectic geometry. These lectures were aimed at the undergraduate students, so [**Arn1989**] covers the calculus of differential forms and vector fields on smooth manifolds. Nowadays, this material is a part of the standard graduate mathematics curriculum, and we use it freely in the main text. The reader can find necessary differential geometry background in the numerous textbooks and monographs; we just mention [**War1983**], [**DFN1984**], [**DFN1985**] and [**Ste1983**], as well as the lecture notes [**Bry1995**]. Same applies to the background in the theory of Lie groups and Lie algebras; all necessary material can be found in [**Kir2004**] and [**Kir2008**]. In addition, the lectures [**God1969**] provide an introduction to the differential geometry of classical mechanics. In particular, the important role the second tangent bundle in Lagrangian mechanics is emphasized in [**God1969**] and [**Bry1995**].

Our exposition follows the traditional outline in [**LL1976**], [**Arn1989**] and in [**Tak2008**]. It starts with the Lagrangian formalism and introduces the Hamiltonian formalism through the Legendre transform. As in [**Arn1989**], we made a special emphasis on precise mathematical formulation of the basic principles, starting with the principle of the least action and the notion of Newtonian spacetime and the Galileo principle of relativity in Chapter 1. Exposition in Chapter 2 follows [**Arn1989**] and [**Tak2008**], and some of the nontrivial exercises have been taken from [**Bry1995**]. Material in Chapter 3 is pretty standard; since our goal is introduce the reader to the classical field theories — systems with infinitely many degrees of freedom — we skip many topics, including the Euler equations and rotation of a rigid body, Kepler problem, theory of oscillations, etc. Some of these topics have been relegated to the exercises; in particular, Exercises 3.4–3.6 introduce the celebrated Laplace-Runge-Lenz vector, whose components are extra integrals of motions for the Kepler problem.

Example 3.1 in Chapter 3 plays an important role in the explanation of the passage to the systems with with infinitely many degrees of freedom in Part 2. Chapter 4 is a succinct introduction to the Hamiltonian formalism and the basics of symplectic geometry — the study of symplectic (and, more generally, Poisson) manifolds. The notion of a Poisson bracket will play a fundamental role in classical field theories. We also just touched a very important notion of completely integrable Hamiltonian systems and formulate the celebrated Liouville-Arnold theorem, proved in [**Arn1989**]. We refer the reader to [**AKN1997**] and references therein for a comprehensive exposition, and to the monograph [**FT2007**] for the Hamiltonian approach to the theory of integrable systems, especially for the case of infinitely many degrees of freedom. Exercises 4.8 and 4.9 are very important: the former provides with a rich class of Poisson structures, while the latter introduces the reader to the beautiful orbit method in the representation theory. The basic references here are [**Kir1976**] and [**Kir2004**].

In Chapter 5 we introduce a very important notion of a Hamiltonian action of a Lie group and corresponding moment map, having many applications in mathematics. As an illustrative example, in Section 5.3 we consider the Hamiltonian action of the central extension of the Galilean group on the phase space of a free classical particle. This example is not widely known and may be considered as a non-relativistic limit of the Hamiltonian action of the Poincaré group, constructed in [**Are1971**]; note that the central extension of the Galilean group was introduced in the classic paper [**Bar1954**]. In Section 5.4 we discuss the Hamiltonian reduction, a powerful methods of obtaining a symplectic quotient of a symplectic manifold with the Hamiltonian group action. The main result — Theorem 5.8 — is due to Marsden and Weinstein; we refer the reader to [**AM1978**, Theorem 4.3.1] for the proof. In Chapter 6 we discuss Lagrangian and Hamiltonian systems with constraints. This formalism was developed by Dirac [**Dir1958**] for the Hamiltonian formulation of gauge theories and general relativity, the theories whose Lagrangians are singular. The Hamiltonian reduction in Section 5.4 is a special case of the Dirac formalism. Our exposition is based on the Faddeev's classic paper [**Fad1969**] and his lectures at Leningrad State University in the early 70s. The Dirac formalism, widely used by physicists, is practically unknown to the mathematicians, so we think that it worthwhile to make it popular. The Chapters 1–6 close our very brief exposition of the classical mechanics.

The Chapters 7 and 8 are meant to be a succinct introduction to the special relativity, written from a mathematics perspective. Thus in Chapter 7 we define the Minkowski spacetime, state the special principle of relativity and introduce Lorentz and Poincaré groups. Mathematicians will find explicit formulas for the generators of the corresponding Lie algebras of these groups, extensively used by physicists. In Section 7.4 we show, in a precise mathematical sense, that the Lorentz group is a deformation of the homogenous Galilean group, and introduce necessary notions from the Lie algebra cohomology. This very simple result has important philosophical consequences: the Lorentz Lie algebra is simple and therefore is stable under the deformations, so the passage from the Newtonian spacetime to the Minkowski spacetime is a deformation from the unstable physical structure to the stable one, and the special relativity is a natural deformation of the Newtonian mechanics. This point of view, advocated in [**Fla1982**] and [**Fad1998**], can be summarized by saying that the change of a fundamental physics paradigm is an 'evolution' rather than a 'revolution'. Mathematical background on the cohomology theory of Lie algebras and its relation to the deformation theory can be found in [**FF2000**] and in the monograph [**dAI1995**], written for physicists. The fact that Poincaré Lie algebra admits no non-trivial central extensions, in [**dAI1995**] was attributed to Wigner. We refer to [**Gal1967**] for a concise proof of this result.

Material in Sections 8.1–8.4 is rather standard and is a careful mathematical exposition of the corresponding sections in physics classic [**LL1971**], while the content of Sections 8.5 and 8.6 is not so well-known. Namely, in Section 8.5 we consider a Hamiltonian action of the Poincaré group on the phase space of a relativistic particle. Our exposition follows the paper [**Are1971**]; the corresponding Hamiltonian action of the Galilean group, described in Section 5.3, is a non-relativistic limit of this action. In Section 8.6 we consider a no-interaction theorem in the special relativity, which states that a system of several relativistic particles which admits a Poincaré group action[1] actually describes the free particles. In other words, the special principle of relativity implies that the interaction of finitely many relativistic particles is not possible! This result (see paper [**Leu1965**] for the proof) has an implication that one needs to pass from particles to fields in order to describe the relativistic nature of the physical world.

The material in Chapters 7 and 8 is needed for the relativistic classical fields theories In Part II. There are plenty of textbooks on special relativity, written from physics and mathematics perspectives. In addition to the physics classic [**LL1971**], we mention modern mathematical exposition in [**DFN1984**, **DFN1985**] and references therein. It should be noted that the choice of a signature $(+,-,-,-)$ for the Minkowski metric in Chapter 7 is just a convention. Equivalently, one can use the metric $\mathrm{diag}(-1,1,1,1)$ of signature $(-,+,+,+)$. Each choice of a signature has advantages and disadvantages. Thus our choice (called 'Landau-Lifshitz sign convention' or 'West Coast convention') is more convenient for the field theory (classical and quantum), while there are some extra negative signs for 3-vectors as in Chapter 8. The opposite choice of a signature (called 'Pauli convention' or 'East Coast convention') keeps Euclidean metric for the spacial variables, and is often used in general relativity.

In Chapter 9, we discuss the notion of spinors and Dirac equation, used for describing quantum particles with spin 1/2 (like electrons and positrons). Our presentation is very basic and elementary; detailed mathematical exposition can be found in the books [**LM1989**] and [**Del1999**]. For application to quantum field theory, we refer to the monograph [**BLOT1990**].

[1] For the validity of the no-interaction theorem it is essential that the Minkowski spacetime is four dimensional.

Part 2

Basics of Classical Field Theory

Chapter 11

Lagrangian Formulation of Field Theory

In this section, we lay a foundation of classical field theory. Informally fields are systems with infinitely many degrees of freedom, parameterized by the points in the physical space, as will be explained below. In analogy with the classical mechanics, where Euler-Lagrange equations are ordinary differential equations, equations of motion for classical fields are partial differential equations. This is reflected in the Lagrangian formalism, where corresponding Lagrangian densities should be *local*. It is described in Section 11.5, together with the derivation of corresponding Euler-Lagrange equations.

11.1. From particles to fields

Here we explain, at a physical level of rigor, how classical fields can be thought of as continuous limit of systems of classical particles when the number particles go to infinity. Namely, consider Example 3.1 in Section 3.3: a system of N coupled oscillators with the Lagrangian function

$$L(\boldsymbol{q}, \dot{\boldsymbol{q}}) = \sum_{j=1}^{N} \tfrac{1}{2}\dot{q}_j^2 - \sum_{j=1}^{N-1} \tfrac{1}{2}\omega_N^2 (q_{j+1} - q_j)^2, \quad q_{N+1} = q_1. \tag{11.1}$$

Corresponding equations of motion are

$$\ddot{q}_j = \omega^2 (q_{j+1} + q_{j-1} - 2q_j), \quad j = 1, \dots, N, \tag{11.2}$$

where it is understood that $q_0 = q_N$. A passage to the continuous limit is an assumption that there is a smooth function $\varphi(x,t)$, defined on the circle $\mathbb{R}/L\mathbb{Z}$ (or equivalently, on an interval $[0, L]$ with periodic boundary conditions), such that for each N

$$q_j(t) = \varphi(x_j, t), \quad \text{where} \quad x_j = j\delta, \quad \delta = \frac{L}{N} \quad \text{and} \quad \omega_N = \frac{c}{\delta}$$

for some constant $c > 0$. In the simultaneous limit $N \to \infty$ and $\delta \to 0$ such that $N\delta = L$ is fixed, the points x_j, $j = 1, \ldots, N$, become uniformly distributed on the circle $\mathbb{R}/L\mathbb{Z}$. By elementary calculus, we obtain that the system (11.2) of ordinary differential equations turns into the partial differential equation

$$\frac{\partial^2 \varphi}{\partial t^2} = c^2 \frac{\partial^2 \varphi}{\partial x^2} \tag{11.3}$$

— the wave equation for the function $\varphi(x,t)$ on the interval $[0, L]$ with periodic boundary conditions.

This example shows how the system with N degrees of freedom, describing N particles moving on the real line, in the continuous limit becomes a system with infinitely many degrees of freedom, described by a classical field — a real-valued smooth function $\varphi(x)$ on the circle. Note that in this limit the index j turns into the spatial variable x, which parameterizes infinitely many degrees of freedom.

Moreover, it follows from the basic calculus that the finite sum (11.1) turns into the integral over $[0, L]$ as $N \to \infty$, so that Lagrangian function $L(\boldsymbol{q}, \dot{\boldsymbol{q}})$ becomes the following functional

$$L(\varphi) = \int_0^L \mathscr{L}(\varphi(x,t))dx, \quad \text{where} \quad \mathscr{L}(\varphi(x,t)) = \frac{1}{2}\left((\partial_t \varphi)^2(x,t) - c^2(\partial_x \varphi)^2(x,t)\right) \tag{11.4}$$

and we put $\partial_t = \dfrac{\partial}{\partial t}$, $\partial_x = \dfrac{\partial}{\partial x}$. As it is well-known in the calculus of variations, the wave equation (11.3) corresponds to the critical points of the action functional

$$S[\varphi] = \int_{t_0}^{t_1} L(\varphi)dt, \tag{11.5}$$

defined on the space of fields $\varphi(x,t)$ with fixed values at the endpoints, $\varphi(x, t_0) = \varphi_0(x)$ and $\varphi(x, t_1) = \varphi_1(x)$.

11.2. Classical fields

Returning to the general situation, physicists think of a *field* as an assignment of a physical quantity to every point in the space-time. For a mathematical definition[1], we assume that we are given a $(n+1)$-dimensional orientable manifold $\mathcal{M}$, which plays the role of spacetime, and a real or complex vector bundle $\mathcal{E} \to \mathcal{M}$.

Definition 11.1. A field φ on $\mathcal{M}$ is a smooth section over $\mathcal{M}$ of the bundle $\mathcal{E}$, and the set $\mathscr{F} = \Gamma(\mathcal{M}, \mathcal{E})$ of all smooth global sections is the space of fields.

In some cases we will also impose some conditions of fast decay at infinity in spacial directions; this will be discussed separately.

It is standard to call φ a real or complex vector field[2] when $\mathcal{E} = \mathcal{M} \times V$ is a trivial vector bundle over $\mathcal{M}$, where V is an r-dimensional real or complex vector space; when $\mathcal{E}$ is a trivial line bundle, φ is called a scalar field. Choosing a basis in V allows to represent fields by smooth vector-valued functions on $\mathcal{M}$, $\varphi(x) = (\varphi^1(x), \ldots, \varphi^r(x))$, $x \in \mathcal{M}$.

[1] Later we extend this definition to include gauge fields and gravitational field.

[2] Not to be confused with a vector field on a manifold!

In most examples (with the notable exception of the theory of gravity) $\mathcal{M} = M \times \mathbb{R}$, where M is a n-dimensional orientable manifold, and the bundle $\mathcal{E}$ is a pull-back of some bundle E over M by the natural projection $\mathcal{M} \to M$. The manifold M plays a role of a physical space. It is usually assumed that the space-time $\mathcal{M} = M \times \mathbb{R}$ is equipped with a (pseudo) Riemannian metric, obtained from a Riemannian metric on M and the standard metric on $\mathbb{R}$; in a relativistic field theory $\mathcal{M} = \mathbb{R}^{1,3}$ is the Minkowski space-time. If $\boldsymbol{x} = (x^1, \dots, x^n)$ are local coordinates on M, then $x = (x^0, \boldsymbol{x})$, where we put $x^0 = t$, are local coordinates on $\mathcal{M}$. The set

$$\mathcal{F} = \Gamma(M, E)$$

of all smooth global sections of E is a field theory analog of the configuration space — the space of all fields at any given time $t = t_0$. Again, for non-compact M we will frequently impose additional conditions of fast decay at infinity; exact conditions depend on M.

Since configuration space $\mathcal{F}$ is infinite-dimensional for $n \geq 1$, classical field theory describes a system with infinitely many degrees of freedom. Finally, as in Section 1.2, we introduce the path space

$$P(\mathcal{F}) = P(\mathcal{F})_{\varphi_0, t_0}^{\varphi_1, t_1},$$

— the space of all smooth parameterized paths $\gamma\colon [t_0, t_1] \to \mathcal{F}$ with fixed endpoints, $\gamma(t_0) = \varphi_0$ and $\gamma(t_1) = \varphi_1$. Slightly abusing notation by using $\mathcal{E}$ for the restriction of the bundle $\mathcal{E}$ to $M \times I$, we have

$$P(\mathcal{F}) = \left\{\varphi \in \Gamma(M \times I, \mathcal{E}) : \varphi|_{t_0} = \varphi_0,\ \varphi|_{t_1} = \varphi_1\right\}, \quad \text{where} \quad I = [t_0, t_1].$$

Similarly, the space of fields $\mathscr{F}$ is a space of all smooth $\mathbb{R}$-parameterized paths in $\mathcal{F}$.

One can also define nonlinear fields as maps from $\mathcal{M}$ to another finite-dimensional manifold N; M is a called a source manifold and N — a target manifold. In case $\mathcal{M} = M \times \mathbb{R}$ the configuration space is $\mathcal{F} = C^\infty(M, N)$.

Remark 11.2. In case $M = \{\mathrm{pt}\}$ we have $\mathcal{F} = C^\infty(\mathrm{pt}, N) \simeq N$, which is exactly the configuration space in classical mechanics, so the fields are trajectories in N. Thus one can consider classical mechanics as a special case of classical field theory, namely a $(0+1)$ dimensional field theory.

11.3. Calculus of variations

The configuration space $\mathcal{F}$ and the space of fields $\mathscr{F}$ are infinite-dimensional Fréchet manifolds. We will later define action functional on the space of fields and talk about its critical points. To do so rigorously, we need to remind basic notions of the calculus of variations and introduce the de Rham complexes on $\mathcal{F}$ and $\mathscr{F}$ with the de Rham differential δ. For our purposes, it will be sufficient to explain the differential δ on 0-forms and to introduce the spaces of 1-forms and vector fields on $\mathscr{F}$.

Note that since $\mathscr{F}$ is a vector space, for any $\varphi \in \mathscr{F}$, the tangent space at φ is naturally identified with $\mathscr{F}$ itself: $T_\varphi \mathscr{F} \cong \mathscr{F}$. We define the cotangent space $T_\varphi^* \mathscr{F}$ to be the space $\mathscr{F}^*$ of all continuous linear functionals on $\mathscr{F}$.

Let $F : \mathscr{F} \to \mathbb{R}$ be a functional on $\mathscr{F}$. The directional derivative $\partial_\xi F$ of F at a point φ (the Gateaux derivative) in the direction $\xi \in T_\varphi\mathscr{F} \cong \mathscr{F}$ is defined by a calculus formula

$$(\partial_\xi F)[\varphi] = \frac{d}{ds}\bigg|_{s=0} F[\varphi + s\xi],$$

provided that the limit exists for all φ and ξ. We also assume that for $\partial_\xi F[\varphi]$ all φ is continuous linear functional of ξ. Similarly, we define higher order directional derivatives by

$$(\partial^k_{\xi_1\cdots\xi_k} F)[\varphi] = \frac{\partial^k}{\partial s_1 \cdots \partial s_k}\bigg|_{s_1=\cdots=s_k=0} F[\varphi + s_1\xi_1 + \cdots + s_k\xi_k], \tag{11.6}$$

assuming that the limit exists for all φ and $\xi_1, \ldots, \xi_k$ and is a continuous multlinear functional.

Definition 11.3. A functional $F\colon \mathscr{F} \to \mathbb{R}$ is called smooth if it has partial directional derivatives of all orders.

Example 11.1. Consider the simplest case of a trivial vector bundle $\mathcal{E} = \mathcal{M} \times V$, where V is r-dimensional real or complex vector space. Choosing a basis of V allows us to represent fields by smooth vector-valued functions on $\mathcal{M}$, $\varphi = (\varphi^1, \ldots, \varphi^r)$.

Choose a point $x_0 \in \mathcal{M}$. Then the evaluation functional

$$F[\varphi] = \varphi^a(x_0)$$

is smooth:

$$\partial_\xi F[\varphi] = \xi^a(x_0).$$

and higher derivatives vanish (it is a linear functional).

More generally, let $K_a(x)$, $a = 1, \ldots, r$, be a collection of compactly supported smooth functions on $\mathcal{M}$ and consider the functional

$$F[\varphi] = \int_{\mathcal{M}} \Big(\sum_a \varphi^a[x] K_a(x)\Big) d^{n+1}x$$

where $d^{n+1}x$ is a top degree differential form on $\mathcal{M}$. Then it is easy to see that so defined F is also smooth, with derivative

$$\partial_\xi F[\varphi] = \int_{\mathcal{M}} \Big(\sum_a \xi^a[x] K_a(x)\Big) d^{n+1}x$$

By definition, the 0-forms are smooth functionals $F\colon \mathscr{F} \to \mathbb{R}$. A 1-form ϑ on $\mathscr{F}$ assigns to each $\varphi \in \mathscr{F}$ a vector $\vartheta_\varphi \in T^*_\varphi\mathscr{F} \cong \mathscr{F}^*$. Thus, given a point $\varphi \in \mathscr{F}$ and a tangent vector $\xi \in T_\varphi\mathscr{F} \cong \mathscr{F}$, we get a real number $\vartheta_\varphi(\xi) = \langle \vartheta_\varphi, \xi\rangle$ where $\langle\ ,\ \rangle$ is the pairing between $\mathscr{F}^*$ and $\mathscr{F}$.

As in calculus, we define the total differential of a smooth functional $F\colon \mathscr{F} \to \mathbb{R}$ to be the 1-form δF, defined by

$$\delta F_\varphi(\xi) = \partial_\xi F[\varphi] \tag{11.7}$$

for all $\varphi \in \mathscr{F}$ and $\xi \in T_\varphi\mathscr{F} \cong \mathscr{F}$. We use δ for the exterior derivative on $\mathscr{F}$ so it is not confused with the exterior derivative on $\mathcal{M}$.

It is very convenient to use physics notation that stem from the calculus of variations. First, recall the familiar finite-dimensional situation, in which we are studying smooth functions on a finite-dimensional vector space V. In this case, we can choose coordinates v^i in V, writing each vector v as $v = (v^1, \dots, v^n)$. The corresponding one-forms dv^i form the basis (over $C^\infty(V)$) of one-forms on V, so for any $f \in C^\infty(V)$ and a tangent vector $\xi \in T_vV = V$, we can write

$$df = \sum \frac{\partial f}{\partial v^i} dv^i, \qquad \partial_\xi f = \sum \frac{\partial f}{\partial v^i} \xi^i. \tag{11.8}$$

Let us write an analog of this formula when we replace a finite-dimensional vector space by the infinite-dimensional space of fields $\mathscr{F}$. As before, consider the simple case when $\mathcal{E} = \mathcal{M} \times V$, where V is r-dimensional real or complex vector space. Choosing a basis of V allows to represent fields by smooth vector-valued functions on $\mathcal{M}$, $\varphi = (\varphi^1, \dots, \varphi^r)$. The values of $\varphi^a(x)$ for all $x \in \mathcal{M}$ and $a = 1, \dots, r$, play the role of coordinates on $\mathscr{F}$, similar to coordinates v^a in the finite-dimenional case, and $\delta\varphi^a(x)$ — variations of fields φ^a at points x — are 1-forms on $\mathscr{F}$, similar to dv^a: for $\xi \in T_\varphi\mathscr{F} = \mathscr{F}$, $\langle \delta\varphi^a(x), \xi \rangle = \xi^a(x)$.

To write an analog of formula (11.8), we need to sum over index a and also over all x; since $x \in \mathcal{M}$, it means we need to choose a volume form $d^{n+1}x$ on $\mathcal{M}$ and integrate over $\mathcal{M}$. Thus, we write

$$\delta F_\varphi = \sum_a \int_{\mathcal{M}} \frac{\delta F}{\delta\varphi^a(x)} \delta\varphi^a(x) d^{n+1}x, \qquad \partial_\xi F[\varphi] = \sum_a \int_{\mathcal{M}} \frac{\delta F}{\delta\varphi^a(x)} \xi^a(x) d^{n+1}x. \tag{11.9}$$

This formula should be taken as the definition of "partial derivatives" $\dfrac{\delta F}{\delta\varphi^a(x)}$. Indeed, by definition of a smooth functional, $\partial_\xi F[\varphi]$ is a continuous linear functional of ξ and thus can be written in the form $\int_{\mathcal{M}} \sum_a K_a(x)\xi^a(x) d^{n+1}x$ for some generalized functions (distributions) K_a.

Note that $\dfrac{\delta F}{\delta\varphi^a(x)}$ is not necessarily a smooth function - it is a distribution. For example, if $F[\varphi] = \varphi^a(x_0)$ is the evaluation functional considered in Example 11.1, then $\partial_\xi F[\varphi] = \xi^a(x_0) = \int_{\mathcal{M}} \xi^a(x)\delta(x - x_0) d^{n+1}x$, so

$$\frac{\delta\varphi^a(x_0)}{\delta\varphi^b(x)} = \delta^a_b \delta(x - x_0), \tag{11.10}$$

where $\delta(x)$ is the Dirac's delta-function.

11.4. Lagrangian density

Equations of motion in classical mechanics are Euler-Lagrange equations in the one variable calculus of variations. It is quite remarkable that equations of motion in classical field theory are also Euler-Lagrange equations in the 'multivariable' calculus of variations.

Namely, in classical mechanics Lagrangian is a function on the tangent bundle[3] of a configuration space, so it seems natural to define field-theoretical Lagrangian L as a functional on $T\mathcal{F}$ — a smooth map $L\colon T\mathcal{F} \to \mathbb{R}$ (in the Fréchet space topology). Correspondingly, the action S — a functional on the path space $P(\mathcal{F})$ — is defined by the same formula as in Section 1.2, namely

$$S[\varphi] = \int_{t_0}^{t_1} L(\gamma'(t))dt. \tag{11.11}$$

Here a path $\gamma \in P(\mathcal{F})$ is a field $\varphi(x,t)$, and its lift $\gamma'(t)$ to a path in $T\mathcal{F}$ is $(\varphi(x,t), \dot{\varphi}(x,t))$, where dot stands for the partial time derivative. As in classical mechanics (see Section 1.2), we formulate the principle of least action.

Principle of Least Action. Time evolution in classical field theory is described by critical points of the action functional $S[\varphi]$ on the set $P(\mathcal{F})$ of smooth paths in $\mathcal{F}$ with the fixed endpoints. This can can be written as

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} S[\varphi_\varepsilon] = 0 \tag{11.12}$$

for any one-parameter family of paths φ_ε in $P(\mathcal{F})$ such that $\varphi_0(x,t) = \varphi(x,t)$. Since $\mathcal{F}$ is a linear space, it is sufficient to consider only the paths $\varphi_\varepsilon = \varphi + \varepsilon u$, where 'variation' u has compact support on $M \times (t_0, t_1)$, so $u(x,t_0) = u(x,t_1) = 0$ for all $x \in M$.

Using the differential δ introduced in (11.7), we can rewrite (11.12) in a more compact form, commonly used by physicists:

$$\frac{\delta S}{\delta \varphi} = 0. \tag{11.13}$$

The following remarks are in order. First, the space of all Lagrangians — smooth functionals on $T\mathcal{F}$ — is 'too large' to give interesting time evolution. Second, in the formula (11.11) for the action, the time t plays a special role. However, in special relativity, time is no longer absolute and is included into the spacetime, so we would need to find a different, relativistic invariant way of writing the action.

To address both of these issues, we will only be studying theories where the Lagrangian is given by some kind of integrals over the space slices $M \times \{t\}$ of the spacetime $\mathcal{M}$:

$$L(\varphi, t)dt = \int_{M\times\{t\}} \mathcal{L}(\varphi). \tag{11.14}$$

(we will clarify what kind of object $\mathcal{L}$ is below) and thus, the action is given by

$$S[\varphi] = \int_{M\times I} \mathcal{L}(\varphi). \tag{11.15}$$

Here we tacitly assume that the integral over M converges. This is automatically so if M is compact; see Remark 11.8 for discussion of non-compact case.

The quantity $\mathcal{L}$ is called *Lagrangian density*. It should satisfy the following properties:

[3] Here we only consider Lagrangians that do not explicitly depend on time.

- In order for the integral (11.15) to make sense, for every $\varphi \in \mathscr{F}$ the expression $\mathcal{L}(\varphi)$ should be a top degree differential form[4] on $\mathcal{M} = M \times \mathbb{R}$. Thus, $\mathcal{L}$ should be a map
$$\mathcal{L}\colon \mathscr{F} \to \Omega^{n+1}(\mathcal{M}).$$
- (Locality) For every $\varphi \in \mathscr{F}$ the value of a differential form $\mathcal{L}(\varphi) \in \Omega^{n+1}(\mathcal{M})$ at a point $x \in \mathcal{M}$ depends only on the values at this point of the field φ and its partial derivatives up to an order k with respect to all variables. This condition can be reformulated in terms of the jet bundle; we do so below.
- Finally, we will usually also require that $\mathcal{L}$ has no explicit dependence on $x \in \mathcal{M}$ — the only dependence on $\mathcal{M}$ comes from the metric on $\mathcal{M}$. As we will show later, this automatically guarantees relativistic invariance of the action.

Since $\mathscr{F}$ is infinite-dimensional, one must be careful when talking about functions on it. However, there is an easy way to avoid analytic difficulties by using the notion of the jet bundle.

Definition 11.4. Given a point $x \in \mathcal{M}$ and a vector bundle $\mathcal{E}$ over $\mathcal{M}$, the k-jet space $J_k(\mathcal{E}, x)$ is the set of equivalence classes of sections φ of $\mathcal{E}$ over some neighborhood of x, where two sections φ, φ' are equivalent if
$$\frac{\partial^I \varphi}{\partial x^I}(x) = \frac{\partial^I \varphi'}{\partial x^I}(x)$$
for any multi-index I with $|I| \leq k$.

Note that this definition requires a choice of local coordinates near x and a choice of local trivialization of $\mathcal{E}$; however, it is easy to show that $J_k(\mathcal{E}, x)$ is independent of these choices. One immediately sees that the k-jet space is a finite-dimensional manifold. Moreover, it is not very difficult to show that there exists a canonical fiber bundle $J_k(\mathcal{E})$ over $\mathcal{M}$ such that its fiber at any point $x \in \mathcal{M}$ is the jet space $J_k(\mathcal{E}, x)$. This fiber bundle is called the *k-jet bundle* of $\mathcal{E}$. There is a natural map

$$\begin{aligned} \pi_k\colon \Gamma(M, \mathcal{E}) &\to J_k(\mathcal{E}) \\ \varphi &\mapsto [\varphi]_k \end{aligned} \tag{11.16}$$

which assigns to every section its class in the jet bundle.

Remark 11.5. Jet spaces can be also defined for the smooth maps of manifolds. In particular, for the maps $f\colon \mathbb{R} \to N$, the space of k-jets of f at any $x \in \mathbb{R}$ is called the k-th order tangent bundle $T^k_{f(x)}N$. Of course, $T^1N = TN$, the tangent bundle.

Then the locality condition for the Lagrangian density $\mathcal{L}$ can be stated as follows.

Definition 11.6. A a map $F\colon \mathscr{F} \to \Omega^q(\mathcal{M})$ is called local if there exists a number $k \geq 0$ such that the map F can be written as a composition
$$\mathscr{F} \to J_k(\mathcal{E}) \to \Omega^q(\mathcal{M})$$

[4] More accurately, it is a *density* on $\mathcal{M} = M \times \mathbb{R}$, and after fixing an orientation on M, it can be identified with the top degree form.

where the first map is the natural projection (11.16), and the second map is a smooth morphism of bundles over $\mathcal{M}$:

$$\tilde{F}\colon J_k(\mathcal{E}) \to \Omega^q(\mathcal{M}).$$

In a similar way, one defines locality condition for maps

$$\Lambda^p T\mathscr{F} \to \Omega^q(\mathcal{M})$$

as maps which are linear along on fibers of $\Lambda^p T\mathscr{F} \to \mathscr{F}$ and factor through appropriate k-jet space.

Note that $\tilde{F}$ is not required (and usually is not) to be a morphism of *vector* bundles: for example, it can depend exponentially on the value $\varphi(x)$.

This definition not only gives rigorous notion of locality, but also removes all analytic difficulties with defining forms and functions on infinite-dimensional space $\mathscr{F}$, reducing it to a smooth morphism of finite-dimensional bundles on $\mathcal{M}$.

Note that it is natural to interpret maps $\Lambda^p T\mathscr{F} \to \Omega^q(\mathcal{M})$ as (p, q) forms on $\mathscr{F} \times \mathcal{M}$. Thus, we will use the notation

$$\Omega^{p,q}(\mathscr{F} \times \mathcal{M}) = \{\text{local maps } \Lambda^p T\mathscr{F} \to \Omega^q(\mathcal{M})\}. \tag{11.17}$$

We define operators

$$\delta\colon \Omega^{p,q}(\mathscr{F} \times \mathcal{M}) \to \Omega^{p+1,q}(\mathscr{F} \times \mathcal{M})$$
$$d\colon \Omega^{p,q}(\mathscr{F} \times \mathcal{M}) \to \Omega^{p,q+1}(\mathscr{F} \times \mathcal{M})$$

as the exterior derivatives in $\mathscr{F}$ and $\mathcal{M}$ respectively, choosing signs so that δ, d anticommute. For example, for a functional $F \in \Omega^{(0,0)}(\mathscr{F} \times \mathcal{M})$, we have

$$(d\delta F)(\varphi, x) = dx^\mu \wedge (\partial_\mu \delta F)(\varphi, x) = -\partial_\mu \left(\frac{\delta F}{\delta\varphi(x)} \delta\varphi(x) \right) \wedge dx^\mu.$$

11.5. Euler–Lagrange equations

As in classical mechanics, the principle of least action leads to the Euler-Lagrange equations for classical fields.

As in all physics textbooks, we consider first the basic case, when M is Euclidean space $\mathbb{R}^n$ with coordinates $x^1, \dots, x^n$ and the volume form $d^n x = dx^1 \wedge \dots \wedge dx^n$. This also gives local coordinates $x^0, x^1, \dots, x^n$ and the volume form $d^{n+1}x$ on $\mathcal{M} = M \times \mathbb{R}$, where we let $x^0 = t$.

We consider a trivial rank r vector bundle $E = \mathbb{R}^n \times V$ over $\mathbb{R}^n$, where V is real or complex r-dimensional vector space, so that $\mathscr{F} = C^\infty(\mathbb{R}^{n+1}, V)$. Choosing a basis in V allows to represent fields by smooth vector-valued functions on the spacetime, $\varphi = (\varphi^1, \dots, \varphi^r)$.

We consider the Lagrangian density which only depends on the 1-jet of a field φ, i.e. on the value $\varphi(x, t)$ and first derivatives $\partial_\mu \varphi(x, t)$:

$$\mathcal{L}(\varphi) = \mathscr{L}([\varphi]_1) d^{n+1}x = \mathscr{L}(\varphi(x, t), \partial_\mu \varphi(x, t)) d^{n+1}x \tag{11.18}$$

for some function $\mathscr{L}$ on the 1-jet space.

Thus in this case we see that the basic formulas (11.14) and (11.15) for the Lagrangian and the action can be now written as follows

$$L(\varphi) = \int_{\mathbb{R}^n} \mathscr{L}(\varphi(x,t), \partial_\mu \varphi(x,t))\, d^n x \tag{11.19}$$

and

$$S[\varphi] = \int_{t_0}^{t_1} L(\varphi)\, dt = \int_{\mathbb{R}^n \times I} \mathscr{L}(\varphi, \partial_\mu \varphi) d^{n+1} x. \tag{11.20}$$

As before, here we assume that φ is such that the integral converges.

Using the multivariable calculus of variations, we obtain the following basic result.

Theorem 11.7. *A field φ is a critical point of the action (11.20) if and only if*

$$\frac{\partial \mathscr{L}}{\partial \varphi^a} - \frac{\partial}{\partial x^\mu} \frac{\partial \mathscr{L}}{\partial(\partial_\mu \varphi^a)} = 0, \quad a = 1, \dots, r. \tag{11.21}$$

(As usual, we use summation over repeated indices.)

Proof. Consider a family

$$\varphi_\varepsilon(x,t) = \varphi(x,t) + \varepsilon \delta\varphi(x,t),$$

where the variation $\delta\varphi$ has compact support on $\mathbb{R}^n \times I$ and $\varphi|_{t_0} = \varphi|_{t_1} = 0$, so that the integral in the definition of $S[\varphi_\varepsilon]$ is convergent if $S[\varphi]$ was convergent. We have, using summation over repeated Greek and Roman indices,

$$\begin{aligned} \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} S[\varphi_\varepsilon] &= \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \int_{\mathbb{R}^n \times I} \mathscr{L}(\varphi + \varepsilon\delta\varphi, \partial_\mu\varphi + \varepsilon\partial_\mu\delta\varphi) d^{n+1}x \\ &= \int_{\mathbb{R}^n \times I} \Big(\frac{\partial \mathscr{L}}{\partial \varphi^a} \delta\varphi^a + \frac{\partial \mathscr{L}}{\partial(\partial_\mu \varphi^a)} \frac{\partial \delta\varphi^a}{\partial x^\mu} \Big) d^{n+1}x, \end{aligned}$$

where the differentiation under the integral sign is easily justified. Using the Stokes' theorem (integration by parts), we can now rewrite it as

$$\frac{d}{d\varepsilon}\Big|_{\varepsilon=0} S[\varphi_\varepsilon] = \int_{\mathbb{R}^n \times I} \Big(\frac{\partial \mathscr{L}}{\partial \varphi^a} - \frac{\partial}{\partial x^\mu} \frac{\partial \mathscr{L}}{\partial(\partial_\mu \varphi^a)} \Big) \delta\varphi^a\, d^{n+1}x.$$

Note that there are no boundary terms because $\delta\varphi = (\delta\varphi^1, \dots, \delta\varphi^r)$ has compact support on $\mathbb{R}^n \times I$ and vanishes at $t = t_0$ and $t = t_1$.

Now it is immediate that if this integral is zero for any such $\delta\varphi$, then the expression in parentheses must be zero, which gives us the statement of the theorem. □

It follows from the chain rule that Euler-Lagrange equations (11.21) are generally non-linear partial differential equations of the second order. More generally, if the Lagrangian density $\mathscr{L}$ is a function on the jet space $J_k(\mathcal{E})$, then the corresponding Euler-Lagrange equations will be a system of partial differential equations of order $2k$. We leave it to the reader to write down these equations explicitly.

Remark 11.8. The definition of action (11.20) and derivation of Euler–Lagrange equations only make sense if the field φ is such that the integral in (11.20) converges. For example, one could require that φ be compactly supported.

However, restricting our consideration to such fields is not a viable option: in many cases, solutions of Euler–Lagrange equations are such that the integral in (11.20) does not converge.

In this case, we take the point of view that uses the Lagrangian density $\mathcal{L}$ (and not the action S) as the primary object of the theory. Then for any compact region $R \subset \mathcal{M}$ (i.e. a top-dimensional submanifold with boundary), we can define the action $S[\varphi]_R$ by the same formula as in (11.20), but with integration over R rather than $\mathcal{M}$, even if the full action $S[\varphi]$ over the whole $\mathcal{M}$ is not defined. We say that φ is a critical point of the action, if for any compact region $R \subset \mathcal{M}$ and for any variation $\delta\varphi$ whose support is contained in the interior of R, we have

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} S[\varphi + \varepsilon\delta\varphi]_R = 0.$$

It is easy to see that with this new definition of the critical point, Theorem 11.7 holds: a field φ a critical point if and only if it satisfies the Euler–Lagrange equations (11.21).

11.6. Coordinate-free form of Euler-Lagrange equations

It is not difficult to consider a general case when M is a Riemannian manifold. As before, the fields are the sections of a vector bundle $\mathcal{E}$ over spacetime $\mathcal{M} = M \times \mathbb{R}$, and a local Lagrangian density $\mathcal{L}\colon \mathscr{F} \to \Omega^{n+1}(\mathcal{M})$ is assumed to depend on 1-jet of φ; it defines the action functional by (11.15),

$$S[\varphi] = \int_{M\times I} \mathcal{L}(\varphi).$$

In order to derive the Euler–Lagrange equations, we introduce the variational derivative

$$\delta\mathcal{L}\colon T\mathscr{F} \to \Omega^{n+1}\mathcal{M}$$

which can also be considered as a $(1, n+1)$ form on $\mathscr{F} \times \mathcal{M}$. Following physics tradition, we write elements of $T\mathscr{F}$ as pairs $(\varphi, \delta\varphi)$. Note that $\delta\mathcal{L}$ is linear in $\delta\varphi$, but can include first derivatives of $\delta\varphi$.

Before stating the general result, let us summarize the derivation of Euler–Lagrange equations in coordinates, for simplicity just in the case of the scalar field. In this case, the key step in the proof of Theorem 11.7 is the following formula:

$$\begin{aligned}
\delta\mathcal{L} &= \left\{ \frac{\partial \mathscr{L}}{\partial\varphi}\delta\varphi + \frac{\partial \mathscr{L}}{\partial(\partial_\mu\varphi)}\partial_\mu\delta\varphi \right\} d^{n+1}x \\
&= \left\{ \left(\frac{\partial \mathscr{L}}{\partial\varphi} - \partial_\mu \frac{\partial \mathscr{L}}{\partial(\partial_\mu\varphi)} \right)\delta\varphi + \partial_\mu\left(\frac{\partial \mathscr{L}}{\partial(\partial_\mu\varphi)}\delta\varphi \right) \right\} d^{n+1}x \\
&= (\delta\mathcal{L})_l - d\gamma
\end{aligned}$$

where

$$(\delta\mathcal{L})_l = \left(\frac{\partial \mathscr{L}}{\partial\varphi} - \partial_\mu \frac{\partial \mathscr{L}}{\partial(\partial_\mu\varphi)} \right)\delta\varphi \, d^{n+1}x \tag{11.22}$$

and γ is a local $(1, n)$ form on $\mathscr{F} \times \mathcal{M}$ (or equivalently, a map $T\mathscr{F} \to \Omega^n\mathcal{M}$). Right now exact expression for γ is irrelevant, since in the integral defining the variation of action functional, it vanishes:

$$\delta S = \int_{M\times I} \delta\mathcal{L} = \int_{M\times I} (\delta\mathcal{L})_l.$$

The advantage of this rewriting is that since $(\delta\mathcal{L})_l$ is of the form

$$(\delta\mathcal{L})_l = F(\varphi, \partial_\mu\varphi)\delta\varphi\, d^{n+1}x \tag{11.23}$$

it is easy to conclude that $\delta S(\varphi) = 0$ for all $\delta\varphi$ if and only if $F(\varphi, \partial_\mu\varphi)$ is identically zero. Note that we couldn't use the same argument with original $\delta\mathcal{L}$ as it also included terms depending on derivatives $\partial_\mu\delta\varphi$ (these terms went into γ).

We can now formulate the invariant form of Theorem 11.7.

Theorem 11.9.

(1) *For any local Lagrangian density $\mathcal{L}\colon \mathscr{F} \to \Omega^{n+1}\mathcal{M}$ which only depends on 1-jets of φ, there exist a local form $\gamma \in \Omega^{1,n}(\mathscr{F}\times\mathcal{M})$, called the variational form, and a local form $(\delta\mathcal{L})_l \in \Omega^{1,n+1}(\mathscr{F}\times\mathcal{M})$ such that*

$$\delta\mathcal{L} = (\delta\mathcal{L})_l - d\gamma. \tag{11.24}$$

(2) *Locally in $\mathcal{M}$ the form $(\delta\mathcal{L})_l$ is given by*

$$(\delta\mathcal{L})_l = F_a(\varphi, \partial_\mu\varphi)\delta\varphi^a d^{n+1}x \tag{11.25}$$

for some functions F_a.

(3) *The critical points of the action functional are given by the Euler–Lagrange equations:*

$$(\delta\mathcal{L})_l(\varphi) = 0, \quad \textit{or} \quad F_a(\varphi, \partial_\mu\varphi) = 0. \tag{11.26}$$

It is important to note that condition that $(\delta\mathcal{L})_l$ locally has form (11.25) can be restated in coordinate-free way: it is equivalent to requiring that $(\delta\mathcal{L})_l$, considered as a map $T\mathscr{F} \to \Omega^{n+1}\mathcal{M}$, is linear over functions: $(\delta\mathcal{L})_l(\varphi, f(x)\delta\varphi) = f(x)(\delta\mathcal{L})_l(\varphi, \delta\varphi)$ for any $f \in C^\infty(\mathcal{M})$.

Proof. Since it is sufficient to consider only variations $\delta\varphi$ with compact support, by a partition of unity on M we can reduce finding a 1-form γ to a local question. Specifically, in a coordinate chart U use local coordinates x^μ on $U \times I$ and trivialize the bundle $\mathcal{E}$ as $(U \times I) \times V$. Introducing local coordinates in the fiber, we can write the field φ as $\varphi = (\varphi^1, \dots, \varphi^r)$ and the Lagrangian density as

$$\mathcal{L} = \mathscr{L}(\varphi^a, \partial_\mu\varphi^a)d^{n+1}x, \quad \text{where} \quad d^{n+1}x = d^n x \wedge dt,$$

and for simplicity we assumed that $\mathscr{L}$ does not depend explicitly on x^μ. Then we basically repeat the same computation as in the proof of Theorem 11.7:

$$\begin{aligned}\delta\mathcal{L} &= \left\{ \frac{\partial\mathscr{L}}{\partial\varphi^a}\delta\varphi^a + \frac{\partial\mathscr{L}}{\partial(\partial_\mu\varphi^a)}\partial_\mu\delta\varphi^a \right\} \wedge d^{n+1}x \\ &= \left\{ \left(\frac{\partial\mathscr{L}}{\partial\varphi^a} - \partial_\mu \frac{\partial\mathscr{L}}{\partial(\partial_\mu\varphi^a)} \right)\delta\varphi^a + \partial_\mu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\mu\varphi^a)}\delta\varphi^a \right) \right\} \wedge d^{n+1}x.\end{aligned}$$

If we now intoduce the local $(1,n)$-form γ at a point $(\varphi, x) \in \mathscr{F}\times\mathcal{M}$ by

$$\gamma = \left(\frac{\partial\mathscr{L}}{\partial(\partial_\mu\varphi^a)} \right)\delta\varphi^a \wedge \imath_\mu(d^{n+1}x), \tag{11.27}$$

where $\imath_\mu$ is the interior product with ∂_μ, then since d and δ anti-commute, we have

$$\begin{aligned} d\gamma &= -\delta(d\varphi^a)\wedge\left(\frac{\partial\mathscr{L}}{\partial(\partial_\mu\varphi^a)}\right)\imath_\mu(d^{n+1}x) - \delta\varphi^a\wedge d\left(\frac{\partial\mathscr{L}}{\partial(\partial_\mu\varphi^a)}\right)\imath_\mu(d^{n+1}x) \\ &= -\partial_\mu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\mu\varphi^a)}\delta\varphi^a\right)\wedge d^{n+1}x, \end{aligned}$$

so that

$$\begin{aligned} \delta\mathcal{L} &= \left(\frac{\partial\mathscr{L}}{\partial\varphi^a} - \partial_\mu\frac{\partial\mathscr{L}}{\partial(\partial_\mu\varphi^a)}\right)\delta\varphi^a\wedge d^{n+1}x - d\gamma \\ (\delta\mathcal{L})_l = \delta\mathcal{L} + d\gamma &= \left(\frac{\partial\mathscr{L}}{\partial\varphi^a} - \partial_\mu\frac{\partial\mathscr{L}}{\partial(\partial_\mu\varphi^a)}\right)\delta\varphi^a\wedge d^{n+1}x \end{aligned}$$

which completes the proof of part (1) and (2). The proof of part (3) is immediate. □

Note that in general case the Euler-Lagrange equations (11.21) are written in terms of local coordinates on M and a local trivialization of a vector bundle $\mathcal{E}$; we leave it as an exercise to the reader to verify that in fact the Euler-Lagrange equations are independent of these choices. We also leave to the reader to write down the Euler-Lagrange equations for the case when $\mathcal{L}$ depends on k-jets of φ for $k > 1$.

Example 11.2. It is instructive to see what these equations give for classical mechanics, considered as $0+1$ dimensional classical field theory as in Remark 11.2.

Namely, let $\mathcal{M} = \mathbb{R}$ be the 1-dimensional spacetime, and the space of fields is $\mathscr{F} = C^\infty(\mathbb{R}, N)$, where the manifold N is the configuration space. In this case, fields are the same as trajectories of a point in N; following the notation of Chapter 1, we will denote a field by $\boldsymbol{q}(t) \in \mathscr{F}$.

If $\mathcal{L} \in \Omega^{0,1}(\mathscr{F}\times\mathcal{M})$ is a local Lagrangian density which only depends on 1-jet of $\boldsymbol{q}(t)$, it must have the form $\mathscr{L}(\boldsymbol{q}, t) = L(\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t), t)$, where $L \in C^\infty(TN\times\mathbb{R})$ is a Lagrangian function considered in Section 1.3. The corresponding action functional is defined by

$$S[\boldsymbol{q}] = \int_{t_0}^{t_1}\mathscr{L}(\boldsymbol{q}, t)dt. \tag{11.28}$$

By Theorem 11.9, its critical points are given by the Euler-Lagrange equations (11.26),

$$(\delta\mathcal{L})_l = \delta\mathcal{L} + d\gamma = 0,$$

where $\gamma \in \Omega^{1,0}(\mathscr{F}\times\mathbb{R})$ is the variational 1-form. In this case, the definition (11.27) of the variational form γ becomes

$$\gamma = \frac{\partial L}{\partial\dot{\boldsymbol{q}}}\delta\boldsymbol{q} = \theta_L, \tag{11.29}$$

is the Liouville 1-form on TN, so

$$\delta\mathcal{L} + d\gamma = \left(\frac{\partial L}{\partial\boldsymbol{q}} - \frac{d}{dt}\frac{\partial L}{\partial\dot{\boldsymbol{q}}}\right)\delta\boldsymbol{q}\wedge dt.$$

Therefore, we recover Euler–Lagrange equations in Section 1.3:

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

11.7. Relativistic fields

In this section we consider fields on Minkowski spacetime M^4. A choice of the frame fidentifies M^4 with the vector space $\mathbb{R}^{1,3}$ with coordinates $x = (x^0, x^1, x^2, x^3)$ and Minkowski metric

$$\eta_{\mu\nu}dx^\mu dx^\nu = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2,$$

where we put $c = 1$ so that $x^0 = t$. Recall that by results of Section 7.1, we have a natural action of Poincaré group $\mathfrak{P} = \mathfrak{L} \ltimes \mathbb{R}^4$ on $\mathbb{R}^4$, where $\mathfrak{L} = \mathrm{O}(1,3)$ is the Lorentz group.

Leaving aside the gauge fields and gravitational field for the later discussion, here we define relativistic fields, which transform according to (irreducible) representations of the Lorentz group, introduced in Section 7.2. Namely, let (ρ, V) be a representation (single-valued or two-valued) of the Lorentz group $\mathfrak{L}$ in the real or complex vector space V, and let

$$\mathscr{F} = \{\varphi : \mathbb{R}^4 \to V\}$$

be the set of classical fields, sections of the trivial bundle $\mathcal{E} = \mathbb{R}^4 \times V$.

There is a natural action of the Poincaré group $\mathfrak{P}$ on the space $\mathscr{F}$, given by

$$\varphi(x) \mapsto (g \cdot \varphi)(x) = \rho(\Lambda)\varphi(g^{-1}x) \in \mathscr{F},$$

where $g = (\Lambda, a) \in \mathfrak{P}$. The space $\mathscr{F}$ of classical fields with this action of the Poincaré group is called the space of *relativistic classical fields* corresponding to the representation (ρ, V) of the Lorentz group. Clearly, this action extends to the action of the Poincaré group on the space $J_k(\mathscr{F})$ of k-jets.

Definition 11.10. A relativistic classical field theory is a theory with the Lagrangian function $\mathscr{L}$ on the 1-jet space $J_1(\mathcal{E})$ which is invariant under the action of the Poincaré group,

$$\mathscr{L}(g \cdot \varphi, \partial_\mu(g \cdot \varphi)) = \mathscr{L}(\varphi, \partial_\mu\varphi), \quad g \in \mathfrak{P}.$$

Remark 11.11. For the Poincare group action physicists use notation

$$\varphi(x) \mapsto \varphi'(x') = \rho(\Lambda)\varphi(x), \quad \text{where} \quad x' = \Lambda x + a \ \text{ and } \ (\Lambda, a) \in \mathfrak{P}.$$

If ρ is an ordinary (single-valued) finite-dimensional representation of Lorentz group $\mathfrak{L}$, the fields $\varphi(x)$ are called tensor, or pseudo-tensor fields; in case when representation ρ is two-valued, the fields are called spinor or pseudo-spinor fields. In case of single-valued irreducible representations, for the tensor fields the space inversion P and the time reversal T (see Section 7.2) act trivially on $\mathscr{F}$,

$$(P \cdot \varphi)(x) = \varphi(x) \quad \text{and} \quad (T \cdot \varphi)(x) = \varphi(x),$$

while for pseudo-tensor fields they act by multiplication by -1.

The difference between scalar and pseudo-scalar fields (or, more generally, between tensor of pseudo-tensor fields) is important only when one considers interaction with other fields, and will be ignored in this discussion.

11.8. Examples of relativistic field theories

Example 11.3 (Real scalar field)**.** The real scalar field $\varphi(x)$ corresponds to the trivial representation of the Lorentz group in the one-dimensional vector space $V = \mathbb{R}$. Thus the field $\varphi(x)$ is a smooth real-valued function on $\mathbb{R}^4$. The Lagrangian of the scalar field theory is

$$\mathscr{L}(\varphi(x), \partial_\mu\varphi(x)) = \frac{1}{2}\left(\partial_\mu\varphi(x)\partial^\mu\varphi(x) - m^2\varphi^2(x)\right) - V_{\text{int}}(\varphi(x)), \tag{11.30}$$

where $\partial^\mu\varphi = \eta^{\mu\nu}\partial_\nu\varphi$ and $\eta^{\mu\nu}$ is the inverse matrix to $\eta_{\mu\nu}$. From physics point of view, the term $V_{\text{int}}(\varphi(x))$ describes self-interaction of field φ; it is determined by a single smooth function $V_{\text{int}} \in C^\infty(\mathbb{R})$, called the interaction potential. If $V_{\text{int}} = 0$, we get the free scalar field theory. The coefficient $m \geq 0$ is called the *mass*; if $m = 0$, the theory is called massless.[5]

The corresponding Euler-Lagrange equation is

$$(\Box + m^2)\varphi + V'_{\text{int}}(\varphi) = 0, \quad \text{where} \quad \Box = \partial_\mu\partial^\mu = \partial_0^2 - \partial_1^2 - \partial_2^2 - \partial_3^2 \tag{11.31}$$

is the d'Alembert operator (compare with (9.7)).

In invariant notation,

$$\mathcal{L} = \frac{1}{2}d\varphi \wedge \star\, d\varphi - \frac{1}{2}(m^2\varphi^2 + 2V_{\text{int}}(\varphi))d^4x \in \Omega^{0,4}(\mathscr{F} \times M^4),$$

where $\star$ is the Hodge star operator of the Minkowski metric (we will discuss it in more detail in Section 15.2). Since d and δ anti-commute, we have

$$\begin{aligned}
\delta\mathcal{L} &= -d\delta\varphi \wedge \star d\varphi - \delta\varphi(m^2\varphi + V'_{\text{int}}(\varphi))d^4x \\
&= -\delta\varphi\left(d\star d\varphi + (m^2\varphi + V'_{\text{int}}(\varphi))d^4x\right) - d(\delta\varphi \wedge \star d\varphi) \\
&= -\delta\varphi(\Box\varphi + m^2\varphi + V'_{\text{int}})d^4x - d\gamma
\end{aligned}$$

where $\gamma = \delta\varphi \wedge \star\, d\varphi \in \Omega^{1,3}(\mathscr{F} \times M^4)$ is the variational 1-form.

In the case of free field theory ($V_{\text{int}}(\varphi) = 0$), (11.31) becomes the Klein-Gordon equation

$$(\Box + m^2)\varphi = 0. \tag{11.32}$$

Another commonly studied example is the so-called φ^4 theory, in which the interaction potential is given by $V_{\text{int}}(\varphi) = g\varphi^4/4!$; in this case, Euler–Lagrange equation becomes

$$(\Box + m^2)\varphi + g\frac{\varphi^3}{3!} = 0. \tag{11.33}$$

It is a nonlinear Klein-Gordon equation with cubic nonlinearity.

Example 11.4 (Complex scalar field)**.** Here $V = \mathbb{C}$, and the complex scalar field $\varphi(x)$ is a smooth complex-valued function on $\mathbb{R}^4$. Corresponding Lagrangian function for the free field has the form

$$\mathscr{L}(\varphi(x), \partial_\mu\varphi(x)) = \partial_\mu\varphi\overline{\partial^\mu\varphi(x)} - m^2\varphi(x)\overline{\varphi(x)}. \tag{11.34}$$

[5] The reason for this terminology can only be properly explained using quantum field theory, where m becomes the rest mass of the particles; we will not attempt to explain it here.

In addition to the invariance with respect to the Lorentz group action, this Lagrangian function is also invariant with respect to the so-called *gradient transformations* — the circle action, given by

$$\varphi(x) \mapsto e^{i\alpha}\varphi(x), \quad \bar{\varphi}(x) \mapsto e^{-i\alpha}\bar{\varphi}(x), \quad \alpha \in \mathbb{R}.$$

The Euler-Lagrange equations for free complex field are

$$(\Box + m^2)\varphi = 0, \quad (\Box + m^2)\bar{\varphi} = 0. \tag{11.35}$$

As we will see later, complex fields commonly appear in the theory of electromagnetism, where they describe *charged* particles that carry electric charge, while the real fields describe *neutral* particles that carry no charge.

It is trivial to generalize to the case $V = \mathbb{C}^n$ and to write down the Lagrangian for a free $\mathbb{C}^n$-valued field, which is invariant under the $\mathrm{U}(n)$ action.

Example 11.5 (Vector field)**.** The basic example is given by the vector (defining) representation of the Lorentz group when $V = \mathbb{R}^4$. Denoting the corresponding field by $U = (U_\mu)\colon \mathbb{R}^4 \to \mathbb{R}^4$, we have the Lagrangian function

$$\mathscr{L}(U, \partial_\mu U) = -\frac{1}{2}(\partial_\mu U_\nu \partial^\mu U^\nu - m^2 U_\nu U^\nu), \tag{11.36}$$

which gives the vector Klein-Gordon equation

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

Note the difference in signs between Lagrangians (11.30) and (11.36). This is because $\partial_\mu U_i \partial^\mu U^i = -\partial_\mu U_i \partial^\mu U_i$, and we want spacial components U_i to enter (11.36) in the same way as the scalar field φ enters (11.30). This causes the total energy of the field to be unbounded, since contribution of the component U_0 to the Hamiltonian H is negative (see Exercise 13.1 in Chapter 13). From physics perspective, this creates a number of problems. To avoid them, one can modify the Lagrangian, writing instead the following Lagrangian, introduced by Pauli:

$$\mathscr{L}_P(U, \partial_\mu U) = \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{m^2}{2}U_\nu U^\nu, \quad \text{where} \quad F_{\mu\nu} = \partial_\mu U_\nu - \partial_\nu U_\mu.$$

The corresponding Euler-Lagrange equations are

$$\partial_\mu F^{\mu\nu} + m^2 U^\nu = 0, \tag{11.37}$$

and one can show that for $m \neq 0$, they are equivalent to the system

$$\begin{aligned} &\partial_\mu U^\mu = 0 \\ &(\Box + m^2)U_\mu = 0. \end{aligned} \tag{11.38}$$

In other words, this new Lagrangain gives the same equations of motion as (11.36) but with additional constraint $\partial_\mu U^\mu = 0$.

Example 11.6 (Spinor field)**.** Here $V = \hat{S}$, the Dirac bispinor module described in Section 9.4. Corresponding equations of motion is the Dirac equation

$$(i\gamma^\mu \partial_\mu - m)\psi(x) = 0,$$

where γ_μ are Dirac γ-matrices (see Section 9.5), and $\psi \in V \cong \mathbb{C}^4$ is a Dirac bispinor.

Dirac equation is a first order partial differential equation, so it can not be derived from a variational principle involving only a spinor field ψ. In the quantum theory, Dirac spinors are fermions, so it is natural that in the classical theory components of $\psi(x)$ have anti-commuting values. We refer the reader to more advances sources for further information on this very interesting subject.

However, Dirac equation can be formally obtained from the Lagrangian

$$\mathscr{L}(\psi(x),\overline{\psi}(x)) = i\overline{\psi}(x)\gamma^\mu\partial_\mu\psi(x) - m\overline{\psi}(x)\psi(x), \tag{11.39}$$

where

$$\overline{\psi} = \psi^\dagger\gamma^0,$$

the adjoint spinor, is considered to be independent from the field ψ (here, using physics notation, $\psi^\dagger$ is a transpose of a complex conjugate $\psi \in \mathbb{C}^4$). Then clearly, the Euler-Lagrange equation

$$\frac{\partial\mathscr{L}}{\partial\overline{\psi}} = i\gamma^\mu\partial_\mu\psi - m\psi = 0$$

gives the Dirac equation, while

$$\frac{\partial\mathscr{L}}{\partial\psi} - \partial_\mu\frac{\partial\mathscr{L}}{\partial(\partial_\mu\psi)} = -i\partial_\mu\overline{\psi}\gamma^\mu - m\overline{\psi} = 0$$

is the adjoint Dirac equation.

Example 11.7 (General sigma model)**.** Let (M, g) and (N, h) be Riemannian (or pseudo-Riemannian) manifolds and let $\mathscr{F} = C^\infty(M, N)$ be the space of smooth maps $f\colon M \to N$. (When M is non-compact, one considers maps with compact support, or rapidly decaying at infinity if $M = \mathbb{R}^d$.) The metrics g and h define metrics in the corresponding tangent and cotangent spaces to M and N, so that for $df(x) \in T_{f(x)}N \otimes T_x^*M$ its norm $\|df(x)\|$ is well-defined. The corresponding action functional is

$$S(f) = \frac{1}{2}\int_M \|df(x)\|^2 dv(x),$$

where $dv(x)$ is the volume form on M. The physicists prefer to write down the action in terms of the local coordinates $(x^1,\dots,x^m)$ on M and $(y^1,\dots,y^n)$ on N. Denoting by g_{ab} and h^{ij} corresponding metric tensors, for the differential of the map $f(x) = (f^1(x),\dots f^n(x))$ we have

$$\|df(x)\|^2 = g^{ab}(x)h_{ij}(f(x))\partial_a f^i(x)\partial_b f^j(x)$$

and

$$S(f) = \frac{1}{2}\int_M g^{ab}(x)h_{ij}(f(x))\partial_a f^i(x)\partial_b f^j(x)\sqrt{\det g_{ab}(x)}\,d^m x.$$

Corresponding Euler-Lagrange equations describe *harmonic maps.*

When the 'target' N is a semisimple Lie group G with the pseudo-Riemannian (or Riemannian if G is compact) metric induced by the Cartan-Killing form, we have the so-called *principal chiral field model.*

11.9. Exercises

Exercise 11.1. Write down the Euler-Lagrange equations if the Lagrangian density $\mathscr{L}$ is a function on the k-jet space $J_k(\mathcal{E})$.

Exercise 11.2. Prove that the Euler-Lagrange equations (11.21) are independent of a choice of local coordinates on M and of a local trivialization of a vector bundle $\mathcal{E}$.

Exercise 11.3. Show that equations (11.38) in Example 11.5 are equivalent to (11.37). (Hint: apply ∂_ν to left-hand side of (11.37))

Exercise 11.4. Derive Euler-Lagrange equation for the harmonic maps in Example 11.7.

Chapter 12

Conservation Laws

As we have seen in Chapter 2, symmetries in classical mechanics lead to the integrals of motion. Similarly, symmetries in classical field theory lead to the conservation laws, to be described below.

We will continue using the notation of Chapter 11. In particular, we will denote by the $\mathscr{F}$ the space of fields of the theory (usually, sections of a vector bundle on spacetime $\mathcal{M}$). The theory is described by a Lagrangian density $\mathcal{L} \in \Omega^{0,n+1}(\mathscr{F} \times \mathcal{M})$. We also remind that we have introduced the variational form $\gamma \in \Omega^{1,n}(\mathscr{F} \times \mathcal{M})$ by the condition that the $(1, n+1)$ form

$$(\delta\mathcal{L})_l = \delta\mathcal{L} + d\gamma$$

has the form

$$(\delta\mathcal{L})_l = F(\varphi, \partial_\mu\varphi)\delta\varphi\, d^{n+1}x$$

(and thus, doesn't include derivatives $\delta(\partial_\mu\varphi)$).

12.1. Symmetries and Noether's theorem

Let us start with the classical mechanics, considered as $(0+1)$ classical field theory as in Example 11.2. In this case, the space of fields is $\mathscr{F} = C^\infty(\mathbb{R}, N)$, the Lagrangian density is $\mathcal{L}(\boldsymbol{q}, t) = L(\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t), t)\, dt$, and the variational form γ is the Liouville form $\gamma = \dfrac{\partial L}{\partial \dot{\boldsymbol{q}}}\delta\boldsymbol{q}$, so that

$$(\delta\mathcal{L})_l = \delta\mathcal{L} + d\gamma = \left(\frac{\partial L}{\partial \boldsymbol{q}} - \frac{d}{dt}\frac{\partial L}{\partial \dot{\boldsymbol{q}}}\right)\delta\boldsymbol{q} \wedge dt.$$

Let us now consider the symmetries of this system. Generalizing Definition 2.4 in Chapter 2, define a symmetry of such a system as a smooth automorphism $G\colon \mathscr{F} \to \mathscr{F}$ of the field space such that

$$\mathcal{L}(G(\boldsymbol{q}), t) = \mathcal{L}(\boldsymbol{q}, t). \tag{12.1}$$

We assume that G is time-preserving: values and derivatives of $G(\boldsymbol{q})$ at time t only depend on values and derivatives of $\boldsymbol{q}$ at same time t (and possibly on t itself).

Similarly, we say that a vector field ξ on $\mathscr{F}$ is an infinitesimal symmetry if

$$\partial_\xi \mathcal{L} = \delta\mathcal{L}(\xi) = 0. \tag{12.2}$$

It is easy to see that if G_s is a one-parameter group of symmetries, then $\xi = \left.\frac{d}{ds}\right|_{s=0} G_s$ is an infinitesimal symmetry. We have the following obvious result.

Lemma 12.1. *If ξ is an infinitesimal symmetry of $(0+1)$ field theory, then*

$$I = \imath_\xi \gamma \in C^\infty(TN \times \mathbb{R})$$

is an integral of motion.

Proof. By the definition of infinitesimal symmetry,

$$\imath_\xi \delta\mathcal{L} = \partial_\xi \mathcal{L} = 0.$$

Thus, on solutions of equations of motion $\delta\mathcal{L} + d\gamma = 0$ we have $\imath_\xi d\gamma = 0$. Using anti-commutativity of $\imath_\xi$ and d, for $I = \imath_\xi \gamma$ we have

$$dI = -\imath_\xi d\gamma = 0. \quad \square$$

Let us now consider a special case when the infinitesimal symmetry ξ comes from a vector field $X = v^i(\boldsymbol{q})\frac{\partial}{\partial q^i}$ on the configuration space N. Then we have

$$I = \frac{\partial L}{\partial \dot{\boldsymbol{q}}}\boldsymbol{v},$$

so I is precisely the Noether integral of motion described in Theorem 2.5.

In a similar way we can formulate the most general form of Noether's theorem in classical mechanics — Proposition 2.9 in Chapter 2. Namely, we define a *generalized infinitesimal symmetry* as a vector field ξ on $\mathscr{F}$ such that

$$\partial_\xi \mathcal{L} = d\alpha \tag{12.3}$$

with some $\alpha \in \Omega^{0,0}(\mathscr{F} \times \mathbb{R})$, where we use notation $\Omega^{p,q}(\mathscr{F} \times \mathcal{M})$ for the space of local forms, see (11.17). Using again equations of motion $\delta\mathcal{L} + d\gamma = 0$ and anti-commutativity of $\imath_\xi$ and d, we obtain

$$d(\imath_\xi \gamma - \alpha) = 0, \tag{12.4}$$

so that $I = \imath_\xi \gamma - \alpha$ is an integral of motion.

In particular, consider the case when ξ is induced by a vector field $X = v^i(\boldsymbol{q})\frac{\partial}{\partial q^i}$ on N, and α depends only on 1-jet of $\boldsymbol{q}$ and is given by a function K on TN. Then we have

$$I = \frac{\partial L}{\partial \dot{\boldsymbol{q}}}\boldsymbol{v} - K,$$

so I is precisely the Noether integral in Proposition 2.9 in Chapter 2.

It is remarkable that these arguments can be naturally generalized to classical field theories. However, in this case it makes sense to have symmetries that act non-trivially on the spacetime $\mathcal{M}$.

As in Chapter 11, consider the classical field theory on a spacetime $\mathcal{M}$ of dimension $n+1$, with the space of fields $\mathscr{F} = \Gamma(\mathcal{M}, \mathcal{E})$ for some vector bundle $\mathcal{E}$ over $\mathcal{M}$ and the Lagrangian density $\mathcal{L} \in \Omega^{(0,n+1)}(\mathscr{F} \times \mathcal{M})$. Suppose that a Lie group G acts on $\mathcal{M}$ and on the vector bundle $\mathcal{E}$, which gives a natural G-action on $\mathscr{F}$. We say that such action of G on $\mathscr{F}$ is local, if for all $g \in G$ the k-jet of $g \cdot \varphi$ at point $g\,x \in \mathcal{M}$ only depends on the k-jet of φ at x (compare with the Definition 11.6). This defines a G-action on $\mathscr{F} \times \mathcal{M}$, so for each $g \in G$ we have a map $g^*\colon \Omega^{(0,n+1)}(\mathscr{F} \times \mathcal{M}) \to \Omega^{(0,n+1)}(\mathscr{F} \times \mathcal{M})$, given by the pullback of differential forms.

Definition 12.2. The group G is a symmetry group of the theory with the Lagrangian density $\mathcal{L}$, if

$$g^*\mathcal{L} = \mathcal{L} \quad \text{for all} \quad g \in G.$$

Note that the locality of G-action on $\mathscr{F}$ is necessary for the map g^* to preserve locality condition in the definition of $\Omega^{p,q}(\mathscr{F} \times \mathcal{M})$.

In a similar way, one can define the notion of infinitesimal symmetry: given a vector field X on $\mathcal{M}$ and an endomorphism of $\mathcal{E}$, one can define a vector field ξ on $\mathscr{F}$ that naturally extends X (we leave it to the reader to write this explicitly). As before, we then define vector field $\hat{\xi} = \xi + X$ on $\mathscr{F} \times \mathcal{M}$. We say that such a vector field $\hat{\xi}$ is an *infinitesimal symmetry* if

$$L_{\hat{\xi}}(\mathcal{L}) = 0, \tag{12.5}$$

where $L_{\hat{\xi}}\colon \Omega^{(0,n+1)}(\mathscr{F} \times \mathcal{M}) \to \Omega^{(0,n+1)}(\mathscr{F} \times \mathcal{M})$ is the Lie derivative.

Explicitly, since $L_\xi f = \partial_\xi f$ for a function f on $\mathscr{F}$, and $L_X\omega = d(\imath_X\omega)$ for a top degree form ω on $\mathcal{M}$, we can rewrite (12.5) as

$$\partial_\xi \mathcal{L} + d(\imath_X \mathcal{L}) = 0. \tag{12.6}$$

More generally, we say that such a vector field $\hat{\xi} = \xi + X$ on $\mathscr{F} \times \mathcal{M}$ is a *generalized infinitesimal symmetry*, if

$$L_{\hat{\xi}}(\mathcal{L}) = \partial_\xi \mathcal{L} + d(\imath_X \mathcal{L}) = d\alpha \tag{12.7}$$

for some $\alpha \in \Omega^{0,n}(\mathscr{F} \times \mathcal{M})$.

In classical mechanics, symmetries (or infinitesimal symmetries) lead to conservation laws – integrals of motion. A similar statement is also true in classical field theory, with the proper definition of a "conservation law".

Definition 12.3. An n-form $J \in \Omega^{0,n}(\mathscr{F} \times \mathcal{M})$ is a *conservation law* for a classical field theory with a local Lagrangian density $\mathcal{L} \in \Omega^{0,n+1}(\mathscr{F} \times \mathcal{M})$, if $dJ = 0$ on the solutions of equations of motion.

Importance of a conservation law is that for $\mathcal{M} = M \times \mathbb{R}$, the quantity

$$I(t) = \int_{M \times \{t\}} J,$$

assuming that the integral is convergent, is a integral of motion:

$$\frac{dI}{dt} = 0$$

on the solutions of equations of motion. Indeed, we have by Stokes' theorem (again assuming that the integral is convergent)

$$0 = \int_{M\times[t_0,t_1]} dJ = I(t_1) - I(t_0).$$

We now have the following field theory version of the Noether theorem, Proposition 2.9 in Chapter 2.

Theorem 12.4. *Let $(\hat{\xi}, \alpha)$ be a generalized infinitesimal symmetry for a field theory with the local Lagrangian density $\mathcal{L} \in \Omega^{0,n+1}(\mathscr{F}\times\mathcal{M})$ and the variational 1-form $\gamma \in \Omega^{1,n}(\mathscr{F}\times\mathcal{M})$. Then*

$$J = \imath_\xi\gamma + \imath_X\mathcal{L} - \alpha \in \Omega^{0,n}(\mathscr{F}\times\mathcal{M})$$

is a conservation law.

Proof. By definition of infinitesimal symmetry, we have $\partial_\xi\mathcal{L} + d(\imath_X\mathcal{L}) = d\alpha$. On the other hand, on solutions of equations of motion, $\delta\mathcal{L} + d\gamma = 0$, so $\partial_\xi\mathcal{L} = \imath_\xi\delta\mathcal{L} = -\imath_\xi d\gamma$. Using anti-commutativity of $\imath_\xi$ and d, we get

$$dJ = d\imath_\xi\gamma + d\imath_X\mathcal{L} - d\alpha = \partial_\xi\mathcal{L} + d\imath_X\mathcal{L} - d\alpha = 0. \quad \square$$

Note that once we fix a volume form on $\mathcal{M}$ — a nowhere vanishing $(n+1)$-form ω — we can identify n-forms on $\mathcal{M}$ with vector fields. Every vector field X defines an n-form $\imath_X\omega$; in local coordinates, this form is given by $X^\mu\imath_\mu d^{n+1}x$, and conversely, representing a n-form in this way, we get a vector field X. Thus, we can consider J as a vector field $\boldsymbol{J} = J^\mu\partial_\mu$, where J^μ are defined by

$$J(\varphi) = J^\mu\imath_\mu d^{n+1}x. \tag{12.8}$$

In this language, equation $dJ = 0$ is the statement that the corresponding vector field is divergence-free:

$$\boldsymbol{\nabla}\cdot\boldsymbol{J} = \partial_\mu J^\mu = 0. \tag{12.9}$$

In physics, the conservation law J in Noether's theorem is called the *Noether current*, and the integral of J over $M\times\{t\}$ — the *Noether charge*. A field theory version of Noether's theorem is the statement that Noether charge is constant on equations of motion.

Example 12.1. Consider a theory in which the fields are functions with values in a finite-dimensional real or complex vector space V:

$$\mathscr{F} = C^\infty(\mathcal{M}, V)$$

and the Lagrangian density has the form $\mathcal{L} = \mathscr{L}(\varphi, \partial_\mu\varphi)d^{n+1}x$. Choosing a basis of V we represent the fields by $\varphi = (\varphi^1, \dots, \varphi^r)$.

Given an action of a Lie group G in V, let $\rho\colon \mathfrak{g} \to \operatorname{End}(V)$ be the corresponding action of the Lie algebra $\mathfrak{g}$ of G. For $u \in \mathfrak{g}$, denote by $\rho(u)^a_b$ the matrix of $\rho(u)$ in the basis of V: $(\rho(u)\varphi)^a = \rho(u)^a_b\varphi^b$.

Then the vector field ξ_u on $\mathscr{F}$, which corresponds to the endomorphism $\rho(u)$ of a trivial vector bundle $\mathcal{M} \times V \to \mathcal{M}$, at a point $\varphi \in \mathscr{F}$ is given by

$$\xi_u = \int_{\mathcal{M}} (\rho(u)\varphi(x))^a \frac{\delta}{\delta\varphi^a(x)} d^{n+1}x.$$

This simply means that ξ_u at $\varphi \in \mathscr{F}$ is the tangent vector $(\rho(u)^1_b\varphi^b, \dots, \rho(u)^r_b\varphi^b) \in T_\varphi\mathscr{F} \cong \mathscr{F}$.

Assume that the Lagrangian $\mathscr{L}$ is invariant under the action of G. In this case, applying theorem Theorem 12.4 with X and α equal to zero, we see that the corresponding conservation law has the form

$$J_u = \imath_{\xi_u}\gamma,$$

and using formula (11.27) for the variational 1-form γ, we obtain at $(\varphi, x) \in \mathscr{F} \times \mathcal{M}$,

$$J_u = (\rho(u)\varphi(x))^a \frac{\partial\mathscr{L}(\varphi, \partial\varphi)(x)}{\partial(\partial_\mu\varphi^a)} \imath_\mu(d^{n+1}x). \tag{12.10}$$

Equivalently, in this case the vector field $\boldsymbol{J}$ defined by (12.8) is given by

$$\boldsymbol{J} = J^\mu\partial_\mu, \quad J^\mu = (\rho(u)\varphi(x))^a \frac{\partial\mathscr{L}(\varphi, \partial\varphi)(x)}{\partial(\partial_\mu\varphi^a)}. \tag{12.11}$$

As we will see below, in case of the charged scalar field, where $V = \mathbb{C}$, the conservation law (12.10) gives *conservation of charge*.

12.2. Stress-energy tensor

Consider now another special case of symmetries of field theories. As before, the field space is $\mathscr{F} = C^\infty(\mathcal{M}, V)$, where V is a real or complex vector space and the Lagrangian density has the form $\mathcal{L} = \mathscr{L}(\varphi, \partial_\mu\varphi)d^{n+1}x$.

Let G be a Lie group acting on $\mathcal{M}$. We extend this action to $\mathscr{F}$ by letting it act trivially on V, so

$$(g \cdot \varphi)(x) = \varphi(g^{-1}x) \tag{12.12}$$

(later we will also allow non-trivial action on V by incorporating results of Example 12.1).

This also gives, for every $u \in \mathfrak{g}$, a vector field X_u on $\mathcal{M}$ and a vector field ξ_u on $\mathscr{F}$, which extends it. Introducing local coordinates, we will write

$$\begin{aligned} X_u &= X_u^\mu\partial_\mu, \\ \xi_u &= -\int X_u^\mu(x)\partial_\mu\varphi^a(x)\frac{\delta}{\delta\varphi^a(x)}d^{n+1}x. \end{aligned} \tag{12.13}$$

Now suppose that local Lagrangian density $\mathcal{L}$ is invariant under the action of G on $\mathscr{F} \times \mathcal{M}$. Applying Theorem 12.4 with α equal to zero, we see that the corresponding conservation law has the form

$$J_u = \imath_{\xi_u}\gamma + \imath_{X_u}\mathcal{L}. \tag{12.14}$$

Using formula (11.27), we obtain at a point $(\varphi, x) \in \mathscr{F} \times \mathcal{M}$,

$$J_u = \left(-X_u^\nu\partial_\nu\varphi^a\frac{\partial\mathscr{L}}{\partial(\partial_\mu\varphi^a)} + X_u^\mu\mathscr{L}\right)\imath_\mu(d^{n+1}x). \tag{12.15}$$

This motivates the following definition.

Definition 12.5. The stress-energy tensor (or energy-momentum tensor) T^μ_ν is defined by

$$T^\mu_\nu = \partial_\nu \varphi^a \frac{\partial \mathscr{L}}{\partial(\partial_\mu \varphi^a)} - \delta^\mu_\nu \mathscr{L}, \tag{12.16}$$

Note that this tensor only depends on the Lagrangian and is independent of any group actions.

Using T^μ_ν, we can rewrite the conservation law (12.15) as

$$J_u = -X^\nu_u T^\mu_\nu \imath_\mu(d^{n+1}x). \tag{12.17}$$

or, equivalently, that the corresponding vector field is

$$\boldsymbol{J_u} = J^\mu_u \partial_\mu, \qquad J^\mu_u = -X^\nu_u T^\mu_\nu.$$

Results of this section can be easily generalized to the case when G acts both on $\mathcal{M}$ and on V (see Exercise 12.3), or, more generally, when we are given an action of G on $\mathcal{M}$ and a G-equivariant vector bundle on $\mathcal{M}$.

Remark 12.6. Stress–energy tensor can also be defined in another way, in terms of variation of spacetime metric. We will discuss it later when talking about general relativity, see Lemma 22.4.

12.3. Examples

Here we continue to consider the situation where fields are sections of a vector bundle on spacetime $\mathcal{M}$ of dimension $n+1$ and the Lagrangian density has the form $\mathcal{L} = \mathscr{L}(\varphi, \partial_\mu \varphi) d^{n+1}x$. Other examples — gauge theories and the theory of gravity — will be considered later.

Example 12.2. (Real scalar field) Here $\mathcal{M} = \mathbb{R}^n \times \mathbb{R}$ and the fields are real-valued functions on $\mathcal{M}$. The most general local Lagrangian density is

$$\mathcal{L} = \mathscr{L}(\varphi, \partial_\mu \varphi) d^{n+1}x.$$

The abelian group $G = \mathbb{R}^{n+1}$ acts on $\mathcal{M}$ by translations and on $\mathscr{F} \times \mathcal{M}$ by the formula (12.12). If $\mathscr{L}$ does not explicitly depend on $x \in \mathcal{M}$, the Lagrangian density $\mathcal{L}$ is obviously G-invariant, and according to (12.17), corresponding conservation law is

$$J_u = -u^\nu T^\mu_\nu \imath_\mu(d^{n+1}x), \quad u \in \mathbb{R}^{n+1}. \tag{12.18}$$

Specifying u to be standard basis vectors in $\mathbb{R}^{n+1}$, we obtain $n+1$ conserved currents

$$J_\nu = -T^\mu_\nu \imath_\mu(d^{n+1}x).$$

The conservation laws

$$dJ_\nu = 0, \quad n = 0, 1, \ldots, n,$$

imply that

$$T^\mu_\nu = \partial_\nu \varphi \frac{\partial \mathscr{L}}{\partial(\partial_\mu \varphi)} - \delta^\mu_\nu \mathscr{L} \tag{12.19}$$

is divergence free on the solutions of equations of motion,

$$\partial_\mu T^\mu_\nu = 0. \tag{12.20}$$

Thus we see that the stress-energy tensor is the conserved Noether current associated with the spacetime translations.

Remark 12.7. In physics textbooks, (12.20) is also proved as follows. Since $\mathscr{L}$ does not depend explicitly on x^μ, we have by the chain rule

$$\begin{aligned}\partial_\nu \mathscr{L} &= \frac{\partial \mathscr{L}}{\partial \varphi}\partial_\nu\varphi + \frac{\partial \mathscr{L}}{\partial(\partial_\mu\varphi)}\partial_\nu\partial_\mu\varphi \\ &= \left(\frac{\partial \mathscr{L}}{\partial \varphi} - \partial_\mu \frac{\partial \mathscr{L}}{\partial(\partial_\mu\varphi)}\right)\partial_\nu\varphi + \partial_\mu\left(\frac{\partial \mathscr{L}}{\partial(\partial_\mu\varphi)}\partial_\nu\varphi\right).\end{aligned}$$

Thus on the solutions of the Euler-Lagrange equations

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

which is (12.20).

It is very instructive to verify directly the infinitesimal symmetry conditon

$$L_{\hat{\xi}}(\mathcal{L}) = \imath_\xi \delta\mathcal{L} + d(\imath_X\mathcal{L}) = 0,$$

where X and ξ are given by (12.13). We have

$$d(\imath_X\mathcal{L}) = (u^\mu\partial_\mu\mathscr{L})d^{n+1}x,$$

while using the formulas

$$\begin{aligned}\imath_\xi(\delta\varphi(x)) &= -u^\mu\partial_\mu\varphi(x), \\ \imath_\xi(\partial_\mu\delta\varphi(x)) &= -u^\nu\partial_\nu\partial_\mu\varphi(x)\end{aligned}$$

and

$$\delta\mathcal{L} = \left(\frac{\partial \mathscr{L}}{\partial \varphi}\delta\varphi + \frac{\partial \mathscr{L}}{\partial(\partial_\mu\varphi)}\partial_\mu\delta\varphi\right)d^{n+1}x,$$

we obtain

$$\imath_\xi\delta\mathcal{L} = -u^\mu\left(\frac{\partial \mathscr{L}}{\partial \varphi}\partial_\mu\varphi(x) + \frac{\partial \mathscr{L}}{\partial(\partial_\nu\varphi)}\partial_\mu\partial_\nu\varphi\right)d^{n+1}x.$$

Thus $L_{\hat{\xi}}(\mathcal{L}) = 0$ by the same chain rule, used in the 'physics computation' above!

It follows from (12.20) and Stokes' theorem that the corresponding charges

$$Q_\nu = \int_{\mathbb{R}^n} T^0_\nu d^n x \tag{12.21}$$

are conserved on the solutions of the Euler-Lagrange equations,

$$\partial_0 Q_\nu = 0, \quad \nu = 0, 1, \dots, n.$$

Of course, here we are tacitly assuming that for all t the solution φ decays sufficiently fast as $x \to \infty$, so integrals in (12.21) converge. This is so in the next example, where the Euler-Lagrange equations are hyperbolic and the decay conditions are uniform in time.

Example 12.3. (Complex scalar field) As in Example 11.4, consider Lagrangian $\mathscr{L}(\varphi, \partial_\mu\varphi)$ for the complex-valued field $\varphi\colon \mathcal{M} \to \mathbb{C}$, invariant under the U(1)-action, the phase rotations: $\varphi(x) \to e^{i\alpha}\varphi(x)$, $\bar{\varphi}(x) \to e^{-i\alpha}\bar{\varphi}(x)$, where $\alpha \in \mathbb{R}$. According to (12.10), the corresponding conservation law is

$$J = i\left(\varphi\frac{\partial\mathscr{L}}{\partial(\partial_\mu\varphi)} - \bar{\varphi}\frac{\partial\mathscr{L}}{\partial(\partial_\mu\bar{\varphi})}\right)\imath_\mu(d^{n+1}x). \tag{12.22}$$

In particular, for the Lagrangian (11.34) we obtain

$$J = J^\mu\imath_\mu(d^{n+1}x), \quad \text{where} \quad J^\mu = i(\varphi\partial^\mu\bar{\varphi} - \bar{\varphi}\partial^\mu\varphi).$$

Conservation law

$$\partial_\mu J^\mu = 0$$

leads to the conservation of charge: if $\mathcal{M} = M \times \mathbb{R}$, then on the solutions of the equations of motion, the total charge

$$Q = \int_M J^0 d^n x = i\int_M (\varphi\partial^0\bar{\varphi} - \bar{\varphi}\partial^0\varphi)d^n x$$

is conserved,

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

Example 12.4. (Relativistic scalar field) Specializing the discussion in Section 11.7, here we take $\mathcal{M} = \mathbb{R}^{1,3}$ be the Minkowski spacetime and consider the relativistic scalar field $\varphi\colon \mathbb{R}^{1,3} \to \mathbb{R}$, where $\mathbb{R}$ is taken with the trivial action of the Lorentz group $\mathfrak{L}$. Corresponding Lagrangian $\mathscr{L}(\varphi, \partial_\mu\varphi)$ is relativistic invariant — invariant under the action of Poincaré group $\mathfrak{P}$ in the sense of Definition 11.10,

$$\mathscr{L}(\varphi(g^{-1}x), \partial_\mu\varphi(g^{-1}x)) = \mathscr{L}(\varphi(x), \partial_\mu\varphi(x)), \quad g \in \mathfrak{P}.$$

In Example 12.2 it was shown that invariance under the spacetime translations leads to the conservation of the stress-energy tensor, equations (12.20). Putting $T^{\mu\nu} = \eta^{\lambda\nu}T^\mu_\lambda$, we get

$$T^{\mu\nu} = \partial^\nu\varphi\frac{\partial\mathscr{L}}{\partial(\partial_\mu\varphi)} - \eta^{\mu\nu}\mathscr{L}. \tag{12.23}$$

We introduce the energy-momentum vector $(p^0, \boldsymbol{p})$ by letting

$$p^0 = T^{00} = \partial^0\varphi\frac{\partial\mathscr{L}}{\partial(\partial_0\varphi)} - \mathscr{L}$$

and

$$\boldsymbol{p} = (T^{01}, T^{03}, T^{03}), \quad T^{0i} = \partial^0\varphi\frac{\partial\mathscr{L}}{\partial(\partial_i\varphi)}.$$

The conservation law (12.20) reads

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

Putting in it $\nu = 0$, we get

$$\frac{\partial p^0}{\partial t} + \nabla\cdot\boldsymbol{p} = 0, \tag{12.24}$$

which leads to the conservation of the energy of the field,

$$\frac{dE}{dt}=0,\quad \text{where}\quad E=\int_{\mathbb{R}^3} p^0\,d^3x.$$

Correspondingly, the conservation laws (12.20) with $\nu\neq 0$ lead to the conservation of the total momentum of the field,

$$\frac{d\boldsymbol{P}}{dt}=0,\quad \text{where}\quad \boldsymbol{P}=\int_{\mathbb{R}^3}\boldsymbol{p}\,d^3x.$$

In particular, if $\mathscr{L}$ is as in Example 11.3, we have

$$T^{00}=\frac{1}{2}\left((\partial_0\varphi)^2+(\nabla\varphi)^2+m^2\varphi^2+V_{\text{int}}(\varphi)\right)\quad \text{and}\quad T^{0i}=\partial^0\varphi\partial^i\varphi. \tag{12.25}$$

Now consider invariance under the Lorentz transformations. Namely, let $u=M^{\alpha\beta}\in\mathfrak{so}(1,3)$ be the generator of the Lorentz Lie algebra, defined by formula (7.26) in Section 7.3. We have

$$X_u^\mu=(M^{\alpha\beta})^\mu_\nu x^\nu=\eta^{\alpha\mu}x^\beta-\eta^{\beta\mu}x^\alpha,$$

so for the conserved current (12.17), which for $u=M^{\alpha\beta}$ we denoted by $J^{\alpha\beta}=J^{\alpha\beta\mu}\imath_\mu(d^{n+1}x)$, we obtain

$$J^{\alpha\beta\mu}=x^\alpha T^{\mu\beta}-x^\beta T^{\mu\alpha} \tag{12.26}$$

which should satisfy the conservation laws

$$\partial_\mu J^{\alpha\beta\mu}=0.$$

Combining this with the conservation law $\partial_\mu T^{\mu\nu}=0$, we get $0=\partial_\mu J^{\alpha\beta\mu}=T^{\alpha\beta}-T^{\beta\alpha}$. In other words, relativistic invariance implies that the stress-energy tensor $T^{\mu\nu}$ is symmetric:

$$T^{\mu\nu}=T^{\nu\mu}. \tag{12.27}$$

The conservation laws $\partial_\mu J^{\alpha\beta\mu}=0$ lead to the conservation of the components of relativistic total angular momentum

$$L^{\mu\nu}=\int_{\mathbb{R}^3}(T^{0\mu}x^\nu-T^{0\nu}x^\mu)d^3x,\qquad \mu,\nu=0,1,2,3. \tag{12.28}$$

Namely, on the solutions of the equations of motion

$$\frac{dL^{\mu\nu}}{dt}=0.$$

The integrals of motion L^{ij} correspond to the space rotations, while L^{0i} correspond to the Lorentz boosts and depend explicitly on t. From (12.28) we have

$$L^{0i}=\int_{\mathbb{R}^3}(x^iT^{00}-x^0T^{0i})d^3\boldsymbol{x}=\int_{\mathbb{R}^3}x^ip^0-tP_i,$$

so by (12.24) and integration by parts

$$\frac{dL^{0i}}{dt}=\int_{\mathbb{R}^3}x^i\frac{\partial p^0}{\partial t}d^3\boldsymbol{x}-P_i=-\int_{\mathbb{R}^3}x^i\frac{\partial p^i}{\partial x^i}d^3\boldsymbol{x}-P_i=0.$$

It is trivial to generalize these results to the general Minkowski spacetime $\mathcal{M}=\mathbb{R}^{1,n}$ with the Minkowski metric $\eta^{\mu\nu}=\text{diag}(1,-1,\dots,-1)$.

Example 12.5 (Spinor field)**.** As in Example 11.6, consider Lagrangian

$$\mathscr{L}(\psi(x), \overline{\psi}(x)) = i\overline{\psi}(x)\gamma^\mu \partial_\mu \psi(x) - m\overline{\psi}(x)\psi(x).$$

Its stress-energy tensor is obtained as in (12.23) and is given by

$$T^{\mu\nu} = \partial^\nu \psi \frac{\partial \mathscr{L}}{\partial(\partial_\mu \psi)} - \eta^{\mu\nu}\mathscr{L} = i\bar{\psi}\gamma^\mu \partial^\nu \psi - \eta^{\mu\nu}\mathscr{L}.$$

In particular, the energy-momentum vector $(p^0, \boldsymbol{p}) = (T^{00}, T^{01}, T^{02}, T^{03})$ of the spinor field is

$$p^0 = -i\bar{\psi}\gamma^k \partial_k \psi + m\bar{\psi}\psi, \quad p^j = i\bar{\psi}\gamma^0 \partial^j \psi.$$

Conservation law $\partial_\mu T^{\mu\nu} = 0$ gives the conservation of energy of the spinor field.

As in case of the complex scalar field (see Example 12.3), we see that the Lagrangian of the spinor is invariant under the U(1)-action $\psi(x) \mapsto e^{i\alpha}\psi(x)$, $\bar{\psi}(x) \mapsto e^{-i\alpha}\bar{\psi}(x)$. The corresponding conservation law is[1]

$$\partial_\mu J^\mu = 0, \quad \text{where} \quad J^\mu = -i\bar{\psi}\frac{\partial \mathscr{L}}{\partial_\mu \psi} = \bar{\psi}\gamma^\mu \psi.$$

Corresponding conserved quantity

$$Q = \int_{\mathbb{R}^3} J^0 d^3\boldsymbol{x} = \int_{\mathbb{R}^3} \bar{\psi}\gamma^0 \psi \, d^3\boldsymbol{x}$$

plays the role of electric charge.

12.4. Exercises

Exercise 12.1. Extend results in Section 12.1 to the Lagrangians densities that depend on a k-jet of the field φ.

Exercise 12.2. Prove conservation laws in Example 12.4 by direct computation.

Exercise 12.3. Extend the results of Section 12.2, deriving conservation laws for the case when the group G acts both on spacetime $\mathcal{M}$ and on space V.

[1] Here in comparison with (12.22) we put a negative sign in the formula for J^μ in order to get a conserved current $\bar{\psi}\gamma^\mu\psi$. From the quantum theory point of view, it is more natural to use anti-commuting variables when no negative sign is needed.

Chapter 13

Hamiltonian Formulation of Classical Field Theory

In this chapter, we present Hamiltonian formulation of classical field theory, similar to the Hamiltonian formalism in classical mechanics, presented in Chapter 4. For simplicity, we only consider the case of a free real scalar field in Minkowski space-time $\mathbb{R}^{1,3}$, and use a system of units in which $c = 1$, so $x^0 = t$. Hamiltonian formalism for gauge theories will be considered later.

13.1. The phase space and algebra of classical observables

We define the configuration space of a scalar field theory as the Schwartz space

$$\mathscr{M} = \mathscr{S}(\mathbb{R}^3, \mathbb{R})$$

of real-valued functions on $\mathbb{R}^3$, rapidly decaying as $|\boldsymbol{x}| \to \infty$ with all derivatives. It is a Fréchet vector space with the topology defined by the system of the semi-norms

$$\|f\|_{\alpha,\beta} = \sup_{\boldsymbol{x}\in\mathbb{R}^3} |\boldsymbol{x}^\alpha D^\beta f(\boldsymbol{x})|.$$

Since $\mathscr{M}$ is a vector space, it is natural to identify the tangent space $T_\varphi\mathscr{M}$ to $\mathscr{M}$ at every point $\varphi \in \mathscr{M}$ with $\mathscr{M}$, so the tangent bundle $T\mathscr{M}$ is isomorphic to $\mathscr{M} \times \mathscr{M}$.

In classical mechanics, the phase space is the total space of the cotangent bundle to the configuration space. However, in our infinite-dimensional case a topological dual vector space to $\mathscr{M}$ is the space of tempered distributions and is 'too large', so one needs to choose an appropriate subspace. For our purposes it will be sufficient to mimic the finite-dimensional case and consider the space of Schwartz class functions as the subspace in the space of tempered distributions. In other words, we define the phase space $\mathscr{X}$ is to be

$$\mathscr{X} = \mathscr{S}(\mathbb{R}^3, \mathbb{R}) \times \mathscr{S}(\mathbb{R}^3, \mathbb{R}),$$

so points in $\mathscr{X}$ are represented by a pair $(\pi(\boldsymbol{x}),\varphi(\boldsymbol{x}))$ of Schwartz class functions.

The advantage of using the phase space $\mathscr{X}$ is that we can easily define an algebra $\mathscr{A}$ of classical observables on it (cf. Section 4.6).

For this purpose, we will extensively use calculus of variations, developed in Section 11.3. Namely, let $F\colon\mathscr{X}\to\mathbb{R}$ be a smooth functional in the sense of Definition 11.3. Since $\mathscr{X}$ is a linear space, we have $T_{(\pi,\varphi)}\mathscr{X}\simeq\mathscr{X}$ and use notation $\xi=(u,v)$ for a tangent vector $\xi\in T_{(\pi,\varphi)}\mathscr{X}$. Specializing formula (11.6) for multiple directional derivatives to our case, we have

$$(\partial^{m+n}_{u_1\cdots u_m v_1\cdots v_n}F)(\pi,\varphi)=\int_{\mathbb{R}^3}\cdots\int_{\mathbb{R}^3}\frac{\delta^{m+n}F(\pi,\varphi)}{\delta\pi(\boldsymbol{x}_1)\cdots\delta\pi(\boldsymbol{x}_m)\delta\varphi(\boldsymbol{y}_1)\cdots\delta\varphi(\boldsymbol{y}_n)}$$
$$\times u(\boldsymbol{x}_1)\cdots u(\boldsymbol{x}_m)v(\boldsymbol{y}_1)\cdots v(\boldsymbol{y}_m)d^3\boldsymbol{x}_1\cdots d^3\boldsymbol{x}_m d^3\boldsymbol{y}_1\cdots d^3\boldsymbol{y}_n,$$

which introduces multiple variational derivatives of F as symmetric in variables $\boldsymbol{x}_1,\dots\boldsymbol{x}_m$ and $\boldsymbol{y}_1,\dots,\boldsymbol{y}_m$ distributions in $\mathscr{S}(\mathbb{R}^{3(m+n)},\mathbb{R})'$.

Definition 13.1. A smooth functional F is called *admissible* if for all $m,n\geq 0$ and all $(\pi,\varphi)\in\mathscr{X}$, variational derivatives

$$\frac{\delta^{m+n}F(\pi,\varphi)}{\delta\pi(\boldsymbol{x}_1)\cdots\delta\pi(\boldsymbol{x}_m)\delta\varphi(\boldsymbol{y}_1)\cdots\delta\varphi(\boldsymbol{y}_n)}$$

belong to the Schwartz space $\mathscr{S}(\mathbb{R}^{3(m+n)},\mathbb{R})$.

Example 13.1. Let $K(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k)$ be a symmetric Schwartz class function on $\mathbb{R}^{3k}$. Then the functional

$$F(\pi,\varphi)=\int_{\mathbb{R}^3}\cdots\int_{\mathbb{R}^3}K(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k)\varphi(\boldsymbol{x}_1)\dots\varphi(\boldsymbol{x}_k)d^3\boldsymbol{x}_1\dots d^3\boldsymbol{x}_k$$

is admissible.

However, the functional

$$F(\pi,\varphi)=\frac{1}{2}\int\varphi(\boldsymbol{x})^2d^3\boldsymbol{x}$$

is not admissible:

$$\frac{\delta^2F}{\delta\varphi(\boldsymbol{x}_1)\delta\varphi(\boldsymbol{x}_2)}=\delta(\boldsymbol{x}_1-\boldsymbol{x}_2)$$

is not of Schwartz class.

Clearly the product of admissible functionals is an admissible functional.

Remark 13.2. For a smooth functional $F\colon\mathscr{X}\to\mathbb{R}$ its differential dF at a point $(\pi,\varphi)\in\mathscr{X}$ is a continuous linear map $dF\colon\mathscr{X}\to\mathbb{R}$, so that $dF\in\mathscr{S}(\mathbb{R}^3\times\mathbb{R}^3)'$. For an admissible functional F, its differential at (π,φ) is represented by a Schwartz class function, so

$$dF(u,v)=\int_{\mathbb{R}^3}\left(\frac{\delta F(\pi,\varphi)}{\delta\pi(\boldsymbol{x})}u(\boldsymbol{x})+\frac{\delta F(\pi,\varphi)}{\delta\varphi(\boldsymbol{x})}v(\boldsymbol{x})\right)d^3\boldsymbol{x}$$

for all $(u,v)\in\mathscr{X}$, and

$$\frac{\delta F(\pi,\varphi)}{\delta\pi(\boldsymbol{x})}\in\mathscr{S}(\mathbb{R}^3\times\mathbb{R}^3),\qquad\frac{\delta F(\pi,\varphi)}{\delta\varphi(\boldsymbol{x})}\in\mathscr{S}(\mathbb{R}^3\times\mathbb{R}^3).$$

Definition 13.3. The algebra $\mathscr{A}$ of classical observables on $\mathscr{X}$ is the algebra of all admissible functionals on $\mathscr{X}$.

13.2. Symplectic form and Poisson bracket

We formally define the symplectic form on the phase space $\mathscr{X}$ by

$$\Omega = \int_{\mathbb{R}^3} (\delta\pi(\boldsymbol{x}) \wedge \delta\varphi(\boldsymbol{x}))\, d^3\boldsymbol{x}, \tag{13.1}$$

so that $\pi(\boldsymbol{x}), \varphi(\boldsymbol{x})$ play the role of infinite-dimensional Darboux coordinates. This means that the symplectic form Ω is a continuous skew-symmetric bilinear form on $T_{(\pi,\varphi)}\mathscr{X} = \mathscr{X}$, defined by

$$\Omega\,(\xi_1, \xi_2) = \int_{\mathbb{R}^3} (u_1(\boldsymbol{x})v_2(\boldsymbol{x}) - u_2(\boldsymbol{x})v_1(\boldsymbol{x}))\, d^3\boldsymbol{x}, \tag{13.2}$$

where $\xi_1 = (u_1, v_1)$, $\xi_2 = (u_2, v_2) \in \mathscr{X}$. The symplectic form Ω is (weakly) non-degenerate: $\Omega\,(\xi_1, \xi_2) = 0$ for all $\xi_2 \in \mathscr{X}$ implies $\xi_1 = 0$.

As in case of finite-dimensional symplectic manifolds (see formula (4.11) in Section 4.4), for each $F \in \mathscr{A}$ we define a Hamiltonian vector field X_F on $\mathscr{X}$ by

$$dF = -\imath_{X_F}\Omega,$$

so

$$dF(\xi) = \Omega(\xi, X_F) \quad \text{for all} \quad \xi = (u, v) \in \mathscr{X}. \tag{13.3}$$

Similarly to the finite-dimensional case (see formula (4.12) in Section 4.4), from (13.3) we obtain that at a point $(\pi, \varphi) \in \mathscr{X}$,

$$X_F = \left(-\frac{\delta F(\pi,\varphi)}{\delta\varphi(\boldsymbol{x})}, \frac{\delta F(\pi,\varphi)}{\delta\pi(\boldsymbol{x})}\right) \in \mathscr{X}.$$

As in Section 4.6, the Poisson bracket on algebra of observables is defined by the same formula (4.16):

$$\{F, G\} = \Omega(X_F, X_G), \quad \text{where} \quad F, G \in \mathscr{A}.$$

Thus

$$\{F, G\}(\pi, \varphi) = \int_{\mathbb{R}^3} \left(\frac{\delta F}{\delta\pi(\boldsymbol{x})}\frac{\delta G}{\delta\varphi(\boldsymbol{x})} - \frac{\delta F}{\delta\varphi(\boldsymbol{x})}\frac{\delta G}{\delta\pi(\boldsymbol{x})}\right) d^3\boldsymbol{x}, \tag{13.4}$$

where variational derivatives are evaluated at $(\pi, \varphi) \in \mathscr{X}$.

The following result succinctly summarizes symplectic and Poisson geometry of the infinite-dimensional phase space $\mathscr{X}$.

Lemma 13.4. *The symplectic form Ω endows the algebra of observables $\mathscr{A}$ with the Poisson algebra structure given by the Poisson bracket* (13.4).

Proof. It follows from the definition of admissible functionals that $\{F, G\} \in \mathscr{A}$ for $F, G \in \mathscr{A}$. It is also clear that the bracket given by (13.4) satisfies the Leibniz rule. As in case of the canonical Poisson bracket on $\mathbb{R}^{2n}$ (see Section 4.6), the Jacobi identity for the bracket (13.4) is proved by a direct computation. □

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 (13.4) that

$$\{F, \pi(\boldsymbol{x})\} = -\frac{\delta F}{\delta\varphi(\boldsymbol{x})} \quad \text{and} \quad \{F, \varphi(\boldsymbol{x})\} = \frac{\delta F}{\delta\pi(\boldsymbol{x})}.$$

Remark 13.5. In physics textbooks, Poisson structure (13.4) on $\mathscr{A}$ is defined by the following Poisson brackets

$$\{\pi(\boldsymbol{x}), \pi(\boldsymbol{y})\} = \{\varphi(\boldsymbol{x}), \varphi(\boldsymbol{y})\} = 0 \quad \text{and} \quad \{\pi(\boldsymbol{x}), \varphi(\boldsymbol{y})\} = \delta(\boldsymbol{x}-\boldsymbol{y}), \tag{13.5}$$

understood in the distributional sense.

13.3. Legendre transform and Hamilton's equations

We can now introduce the Hamiltonian formalism for classical field theory, similar to the presentation in Section 4.1 for classical mechanics. Namely, we consider a Hamiltonian function H on $\mathscr{X}$ which has the form

$$H(\pi, \varphi) = \int_{\mathbb{R}^3} \mathscr{H}(\pi(\boldsymbol{x}), \varphi(\boldsymbol{x})) d^3\boldsymbol{x}$$

for some Hamiltonian density $\mathscr{H}(\pi(\boldsymbol{x}), \varphi(\boldsymbol{x}))$; as before, we assume that $\mathscr{H}$ only depends on 1-jets of $\pi(\boldsymbol{x}), \varphi(\boldsymbol{x})$ at $\boldsymbol{x}$. Then the time evolution in $\mathscr{X}$ is defined by canonical Hamilton's equations

$$\frac{\partial}{\partial t}\pi(\boldsymbol{x}) = \{H, \pi(\boldsymbol{x})\} = -\frac{\delta H}{\delta\varphi(\boldsymbol{x})}, \tag{13.6}$$

$$\frac{\partial}{\partial t}\varphi(\boldsymbol{x}) = \{H, \varphi(\boldsymbol{x})\} = \frac{\delta H}{\delta\pi(\boldsymbol{x})} \tag{13.7}$$

It is easy to see that these equations imply that for any observable F on $\mathscr{X}$, we have

$$\partial_0 F = \{H, F\}, \quad \text{where} \quad x^0 = t.$$

As in classical mechanics, a Lagrangian system can be rewritten using the Hamiltonian formalism. We do it for a scalar field described in Example 11.3; recall that for this theory, the Lagrangian density is given by

$$\mathscr{L}(\varphi(x), \partial_\mu\varphi(x)) = \frac{1}{2}\left(\partial_\mu\varphi(x)\partial^\mu\varphi(x) - m^2\varphi^2(x)\right) - V_{\text{int}}(\varphi(x))$$

with the Euler-Lagrange equation

$$(\Box + m^2)\varphi + V'_{\text{int}}(\varphi) = 0. \tag{13.8}$$

In this case, the Legendre transformation is given by

$$\pi(x) = \frac{\partial\mathscr{L}}{\partial(\partial_0\varphi(x))} = \partial_0\varphi(x).$$

By analogy with finite-dimensional case (see (4.2)), we define the Hamiltonian functional density $\mathscr{H}(\pi,\varphi)$ by

$$\begin{aligned}\mathscr{H}(\pi(x),\varphi(x)) &= \pi(x)\cdot\partial_0\varphi(x) - \mathscr{L}(\varphi(x),\partial_\mu\varphi(x))|_{\partial_0\varphi=\pi}\\ &= \frac{1}{2}\left(\pi^2(x) + (\boldsymbol{\nabla}\varphi(x))^2 + m^2\varphi^2(x)\right) + V_{\text{int}}(\varphi(x))\end{aligned}$$

so that the Hamiltonian is given by

$$H = \int_{\mathbb{R}^3}\mathscr{H}(\boldsymbol{x})d^3\boldsymbol{x} = \int_{\mathbb{R}^3}\left(\frac{1}{2}\left(\pi^2(\boldsymbol{x}) + (\boldsymbol{\nabla}\varphi(\boldsymbol{x}))^2 + m^2\varphi^2(\boldsymbol{x})\right) + V_{\text{int}}(\varphi(\boldsymbol{x}))\right)d^3\boldsymbol{x}. \tag{13.9}$$

In this case, Hamilton's canonical equations (13.6)–(13.7) coincide with Euler–Lagrange equations (13.8). Indeed, using the calculus of variations, we obtain

$$\begin{aligned}\partial_0\varphi(\boldsymbol{x}) &= \frac{\delta H}{\delta\pi(\boldsymbol{x})}(\pi,\varphi) = \pi(\boldsymbol{x})\\ \partial_0\pi(\boldsymbol{x}) &= -\frac{\delta H}{\delta\varphi(\boldsymbol{x})}(\pi,\varphi) = \Delta\varphi(\boldsymbol{x}) - m^2\varphi(\boldsymbol{x}) - V'_{\text{int}}(\varphi(\boldsymbol{x})),\end{aligned}$$

where as usual $\Delta = \sum_{i=1}^3\partial_i^2$, so that (13.6)–(13.7) yield

$$\partial_0^2\varphi(x) = \Delta\varphi(x) - m^2\varphi(x) - V'_{\text{int}}(\varphi(x)),$$

where $x = (t,\boldsymbol{x})$, which is equivalent to (13.8).

Note that though Hamiltonian H is a smooth functional, it is not admissible in the strict sense of the Definition 13.1 (see Example 13.1). However, its first variational derivatives $\delta H/\delta\pi(\boldsymbol{x})$ and $\delta H/\delta\varphi(\boldsymbol{x})$ belong to the Schwartz class, so canonical Hamilton's equations (13.6)–(13.7) make perfect sense. Same applies to the smooth functionals which are integrals over $\mathbb{R}^3$ of local densities, polynomials of $\varphi(\boldsymbol{x})$ and $\pi(\boldsymbol{x})$ and their partial derivatives at $\boldsymbol{x}$.

13.4. Representation of the Poincaré group

In this section we describe the relativistic invariance in the Hamiltonian formalism (again, only for the free scalar field).

Recall that in the Lagrangian formulation relativistic invariance, introduced in the Definition 11.10, means that the Lagrangian of the theory is invariant under the action of the Poincaré group $\mathfrak{P}$ on the space of classical fields $\mathscr{F}$, defined by $(g\cdot\varphi)(x) = \varphi(g^{-1}x)$. This implies that the set of solutions of Euler–Lagrange equations is also preserved by the action of $\mathfrak{P}$.

Defining the action of the Poincaré group on the phase space in the Hamiltonian formalism is much less obvious. Indeed, definition of the phase space starts with choosing a time slice $t = t_0$; such a time slice is obviously not preserved by the action of the Poincaré group, so it is not even clear if one can define an action of $\mathfrak{P}$ on $\mathscr{X}$.

In this section, we show that for the free scalar field one can indeed define an action of $\mathfrak{P}$ on the phase space, and this action preserves the symplectic form (13.1). Our approach will be similar to the one used in the case of a free relativistic particle in Section 8.5. We will work at a formal level, leaving functional analysis details for the interested reader.

Namely, consider the set $\mathfrak{X}$ of solutions of Euler–Lagrange equations for the free scalar field. Since the Lagrangian of this theory is invaraint under the action of the Poincaré group, we have an obvious action of $\mathfrak{P}$ on $\mathfrak{X}$ defined by

$$\varphi(x) \mapsto (g \cdot \varphi)(x) = \varphi(g^{-1}x), \quad x \in \mathbb{R}^{1,3}, \quad g \in \mathfrak{P}. \tag{13.10}$$

On the other hand, once we fixed the time slice $t = t_0$, each solution is uniquely determined by the initial conditions $\varphi(\boldsymbol{x}) = \varphi(t_0, \boldsymbol{x})$ and $\pi(\boldsymbol{x}) = \partial_t \varphi(t, \boldsymbol{x})|_{t=t_0}$. Conversely, for every $\varphi(\boldsymbol{x})$, $\pi(\boldsymbol{x})$ there is a unique solution $\varphi(x)$ of equations of motion with these initial conditions. Thus, we have an embedding

$$\mathscr{X} \hookrightarrow \mathfrak{X}$$

of the phase space into the space of solutions. Therefore, we use the action of $\mathfrak{P}$ on $\mathfrak{X}$ to define the action of $\mathfrak{P}$ on $\mathscr{X}$:

$$g \cdot (\pi, \varphi)(\boldsymbol{x}) = \left(\frac{\partial}{\partial t}(g \cdot \varphi)(t, \boldsymbol{x})\bigg|_{t=t_0}, (g \cdot \varphi)(t_0, \boldsymbol{x}) \right). \tag{13.11}$$

We assume that the action of $\mathfrak{P}$ preserves the phase space $\mathscr{X}$; in other words, if a solution $\varphi(x)$ of equations of motion is such that $\varphi(t_0, \boldsymbol{x})$ and $(\partial_t \varphi(t, \boldsymbol{x}))|_{t=t_0}$ are of Schwartz class, then the same is true for $\varphi(g^{-1}x)$. This is not automatic; for the free field it follows from Lemma 13.9 which we will prove in the next section.

As usual, action of $\mathfrak{P}$ on $\mathscr{X}$ gives rise to its action on the algebra of observables $\mathscr{A}$:

$$(g \cdot F)(\pi, \varphi) = F(g^{-1}(\pi, \varphi)).$$

For example, if $g \in \mathrm{O}(3) \ltimes \mathbb{R}^3$ is an affine transformation of $\mathbb{R}^3$, then

$$(g \cdot F)(\pi, \varphi) = F(g^{-1}(\pi, \varphi)) = F(\pi(g\boldsymbol{x}), \varphi(g\boldsymbol{x}))$$

In particular, if we take F to be the evaluation functional: $F(\varphi) = \varphi(\boldsymbol{x}_0)$, which we had previously (somewhat confusingly) denoted simply by $\varphi(\boldsymbol{x}_0)$, then we get

$$g\varphi(\boldsymbol{x}_0) = \varphi(g\boldsymbol{x}_0).$$

Once again, $\varphi(\boldsymbol{x}_0)$ here is an observable, a function on $\mathscr{X}$.

As in Section 5.1, the Lie algebra $\mathfrak{p}$ of the Poincare group acts on $\mathscr{X}$ by vector fields. Namely, to each $a \in \mathfrak{p}$ one associates a vector field ξ_a on $\mathscr{X}$ by

$$\xi_a|_{(\pi,\varphi)} = \frac{d}{ds}\bigg|_{s=0} (e^{sa} \cdot (\pi, \varphi)) \in T_{(\pi,\varphi)}\mathscr{X} \quad \text{for all} \quad (\pi, \varphi) \in \mathscr{X}.$$

Correspondingly, the Lie algebra $\mathfrak{p}$ acts on the algebra of observables: for $F \in \mathscr{A}$, we have

$$-\xi_a(F)(\pi, \varphi) = \frac{d}{ds}\bigg|_{s=0} F(e^{-sa} \cdot (\pi, \varphi)),$$

and the mapping $\mathfrak{g} \to \mathrm{Vect}(\mathscr{X})$ given by $a \to -\xi_a$ is a morphism of Lie algebras (cf. Section 5.1).

As in the case of a free relativistic particle, we claim that this action is Hamiltonian: for every $a \in \mathfrak{p}$, we have a Hamiltonian functional H_a on $\mathscr{X}$ such that

$$-\xi_a(F) = \{H_a, F\}$$

and $\{H_a, H_b\} = H_{[a,b]}$ (cf. with (5.2)).

The construction of these Hamiltonians is given in the theorem below (compare with Theorem 8.5 in Section 8.5). As expected, these Hamiltonians are exactly the conserved quantities corresponding to the one-parameter subgroups in $\mathfrak{P}$, except for Hamiltonians of the Lorentz boosts (since they do preserve that Hamiltonian H of the field, cf. Section 8.5). We have already discussed these conserved quantities in the Lagrangian formalism in Section 12.3; here we present them in the Hamiltonian picture.

Theorem 13.6. *The Hamiltonian functions corresponding to spacetime translations, space rotations and Lorentz boosts are*

$$
\begin{aligned}
\hat{P}_0 &= H = \int_{\mathbb{R}^3} \mathscr{H}(\boldsymbol{x}) d^3\boldsymbol{x} \\
\hat{P}_i &= \int_{\mathbb{R}^3} \pi(\boldsymbol{x}) \partial_i \varphi(\boldsymbol{x}) d^3\boldsymbol{x} \\
\hat{J}_i &= \epsilon_{ijk} \int_{\mathbb{R}^3} x^j \pi(\boldsymbol{x}) \partial_k \varphi(\boldsymbol{x}) d^3\boldsymbol{x}, \\
\hat{K}_i &= \int_{\mathbb{R}^3} x^i h(\boldsymbol{x}) d^3\boldsymbol{x} + t_0 \hat{P}_i, \quad i = 1, 2, 3,
\end{aligned}
\tag{13.12}
$$

where

$$\mathscr{H}(\boldsymbol{x}) = \frac{1}{2}\left(\pi^2(\boldsymbol{x}) + (\boldsymbol{\nabla}\varphi)^2(\boldsymbol{x}) + m^2\varphi^2(\boldsymbol{x})\right)$$

is the Hamiltonian density of a free relativistic real scalar field.

The Poisson brackets of these functionals are the same as in Theorem 8.5:

$$\{\hat{P}_i, \hat{P}_j\} = \{\hat{P}_i, \hat{P}_0\} = \{\hat{J}_i, \hat{P}_0\} = 0, \quad \{\hat{J}_i, \hat{J}_j\} = \epsilon_{ijk}\hat{J}_k, \tag{13.13}$$

$$\{\hat{K}_i, \hat{K}_j\} = -\epsilon_{ijk}\hat{J}_k, \quad \{\hat{J}_i, \hat{K}_j\} = \epsilon_{ijk}\hat{K}_k, \tag{13.14}$$

$$\{\hat{K}_i, \hat{P}_0\} = \hat{P}_i, \quad \{\hat{K}_i, \hat{P}_j\} = \delta_{ij}\hat{P}_0, \quad \{\hat{J}_i, \hat{P}_j\} = \epsilon_{ijk}\hat{P}_k. \tag{13.15}$$

Proof. First we show that indeed $\xi_a(F) = \{F, H_a\}$ for the generators P_0, P_i, J_i and K_i of the the Poincaré algebra, defined in Section 7.3.

For the generators P_i of space translations it is easy to see, using formulas (13.10)–(13.11), that vector fields ξ_{P_i} at a point (π, φ) in $\mathscr{X}$ are the vectors $(-\partial_i\pi, -\partial_i\varphi) \in \mathscr{X}$, so

$$\xi_{P_i}(F)(\pi, \varphi) = -\int_{\mathbb{R}^3} \left(\frac{\delta F}{\delta\pi(\boldsymbol{x})}\partial_i\pi(\boldsymbol{x}) + \frac{\delta F}{\delta\varphi(\boldsymbol{x})}\partial_i\varphi(\boldsymbol{x})\right) d^3\boldsymbol{x}.$$

Now a very simple computation using (13.4) shows that the right hand side of this formula is indeed the Poisson bracket $\{F, \hat{P}_i\}$. Similarly, for the vector fields ξ_{J_i} corresponding to generators J_i of the space rotations we have

$$\xi_{J_i}(F)(\pi, \varphi) = -\epsilon_{ijk} \int_{\mathbb{R}^3} \left(\frac{\delta F}{\delta\pi(\boldsymbol{x})}x^i\partial_k\pi(\boldsymbol{x}) + \frac{\delta F}{\delta\varphi(\boldsymbol{x})}x^j\partial_k\varphi(\boldsymbol{x})\right) d^3\boldsymbol{x},$$

and the right hand side is again the Poisson bracket $\{F, \hat{J}_i\}$.

For the generator P_0 of the time translation, using (13.6)–(13.7), we get

$$\xi_{P_0} = (-\partial_t\pi(\boldsymbol{x}), -\partial_t\varphi(\boldsymbol{x})) = \left(\frac{\delta H}{\delta\varphi(\boldsymbol{x})}, -\frac{\delta H}{\delta\pi(\boldsymbol{x})}\right)$$

so that

$$\xi_{P_0}(F) = \int_{\mathbb{R}^3} \left(\frac{\delta F}{\delta\pi(\boldsymbol{x})} \frac{\delta H}{\delta\varphi(\boldsymbol{x})} - \frac{\delta F}{\delta\varphi(\boldsymbol{x})} \frac{\delta H}{\delta\pi(\boldsymbol{x})} \right) d^3\boldsymbol{x} = \{F, H\}.$$

Finally we consider the Lorentz boosts. From formulas (7.6) and (13.10)–(13.11) we obtain

$$\xi_{K_i}(F) =$$

$$-\int_{\mathbb{R}^3} \left(\frac{\delta F}{\delta\pi(\boldsymbol{x})} (\partial_i\varphi(\boldsymbol{x}) + t_0\partial_i\pi(\boldsymbol{x}) + x^i(\Delta\varphi(\boldsymbol{x}) - m^2\varphi(\boldsymbol{x})) + \frac{\delta F}{\delta\varphi(\boldsymbol{x})} (t_0\partial_i\varphi(\boldsymbol{x}) + x^i\pi(\boldsymbol{x}) \right) d^3\boldsymbol{x},$$

and we again see that the right hand side is equal to $\{F, \hat{K}_i\}$.

The computation of Poisson brackets (13.13)–(13.15) between these Hamiltonians is left as an exercise for the reader. □

It is quite remarkable that these Hamiltonian functionals are exactly the conserved quantities, obtained in Example 12.4 by using Noether theorem in the Lagrangian formalism. Namely, in case of a free field ($V_{\text{int}}(\varphi) = 0$) we obtain from (12.25) that components $T^{0\mu}$ of the stress-energy tensor at $x = (t_0, \boldsymbol{x})$ for fixed t_0 are

$$T^{00}(\boldsymbol{x}) = \mathscr{H}(\boldsymbol{x}) \quad \text{and} \quad T^{0i}(\boldsymbol{x}) = -\pi(\boldsymbol{x})\partial_i\varphi(\boldsymbol{x}), \quad \text{where} \quad \pi(\boldsymbol{x}) = \partial_0\varphi(t_0, \boldsymbol{x}).$$

(Remember that $\partial^i\varphi = -\partial_i\varphi$). Thus the Hamiltonian functionals $\hat{P}_0$ and $\hat{\boldsymbol{P}}$ are the Hamiltonian H and total momentum $-\boldsymbol{P}$ of the field. Similarly, we have from (12.28)

$$L^{ij} = \epsilon_{ijk}\hat{J}_k,$$

so Hamiltonian functionals $\hat{\boldsymbol{J}}$ correspond to the conserved total angular momentum. Finally, for the integrals of motion L^{0i} corresponding to the Lorentz boosts we have that at fixed $t = t_0$,

$$L^{0i}(t_0) = \hat{K}_i.$$

Note that while L^{0i} are integrals of motion, the Hamiltonian functionals $\hat{K}_i$ are not (cf. Remark 8.6). Namely, according to (13.15) they satisfy the Hamilton's equations

$$\frac{d\hat{K}_i}{dt}(t) = \{H, \hat{K}_i\} = -\hat{P}_i, \quad \hat{K}_i(t)\Big|_{t=t_0} = \hat{K}_i,$$

which can be easily solved

$$\hat{K}_i(t) = \int_{\mathbb{R}^3} x^i h(t, \boldsymbol{x}) d^3\boldsymbol{x} + t_0\hat{P}_i.$$

Here $h(t, \boldsymbol{x})$ is the Hamiltonian density, evaluated on the solution $(\pi(t, \boldsymbol{x}), \varphi(t, \boldsymbol{x}))$ of canonical Hamilton equations (13.6)–(13.7). Indeed, using (12.24) and integration by parts, we obtain

$$\frac{d\hat{K}_i}{dt}(t) = \int_{\mathbb{R}^3} x^i \nabla \cdot (\pi\nabla\varphi) d^3\boldsymbol{x} = -\hat{P}_i.$$

On the other hand, the integrals of motion L^{0i} have the form

$$L^{0i} = \int_{\mathbb{R}^3} x^i h(t, \boldsymbol{x}) d^3\boldsymbol{x} + t\hat{P}_i,$$

and coincide with $\hat{K}_i(t)$ only at $t = t_0$.

Remark 13.7. One can also define the Hamiltonian action of the Poincaré group on the phase space of a real field with the interaction potential $V_{\text{int}}(\varphi)$ in Example 11.3, if the following conditions are met.

I. The Cauchy problem for the Euler-Lagrange equation
$$(\square + m^2)\varphi + V'_{\text{int}}(\varphi) = 0,$$
$$\varphi(x)|_{t=t_0} = \varphi(\boldsymbol{x}), \quad \partial_0\varphi(x)|_{t=t_0} = \pi(\boldsymbol{x})$$
for $\varphi(\boldsymbol{x}), \pi(\boldsymbol{x}) \in \mathscr{S}$ for all t has a unique solution $\varphi(x)$ such that for each t one has $\varphi(t, \boldsymbol{x}), \partial_0\varphi(t, \boldsymbol{x}) \in \mathscr{S}$.

II. If $\varphi(x)$ is a solution with the Schwarz class initial data at $t = t_0$, than for each $g \in \mathfrak{P}$ the solution $\varphi(g^{-1}x)$ also has a Schwarz class initial data.

We leave it to the interested reader to provide all necessary details.

13.5. Free scalar field as a completely integrable Hamiltonian system

In this section, we show that the free scalar field can be treated as a (continuous) collection of non-interacting harmonic oscillators. We start with the Euler-Lagrange equation for a free real scalar field, which is the Klein-Gordon equation

$$(\square + m^2)\varphi(x) = 0, \tag{13.16}$$

with $m > 0$.

To solve it, we use the Fourier transform,
$$\hat{\varphi}(k) = \frac{1}{(2\pi)^2}\int_{\mathbb{R}^4} e^{ik\cdot x}\varphi(x)d^4x, \quad \text{where} \quad k\cdot x = k^\mu x_\mu = k^0x^0 - \boldsymbol{k}\boldsymbol{x},$$
so that (13.16) is equivalent to the following equation on $\hat{\varphi}$:

$$(k^2 - m^2)\hat{\varphi}(k) = 0. \tag{13.17}$$

Its general solution is a distribution supported on the "mass shell" – two-sheeted hyperboloid $\mathcal{O}_m$, defined by equation $k^2 = (k^0)^2 - \boldsymbol{k}^2 = m^2$, or equivalently,

$$k^0 = \pm\omega_{\boldsymbol{k}}, \quad \omega_{\boldsymbol{k}} = \sqrt{\boldsymbol{k}^2 + m^2} > 0. \tag{13.18}$$

The 'positive-energy part' $k_0 > 0$ of $\mathcal{O}_m$ is denoted by $\mathcal{O}_m^+$.

Obviously, the general solution of (13.17) is given by
$$\hat{\varphi}(k) = \rho_1(\boldsymbol{k})\delta(k^0 - \omega_{\boldsymbol{k}}) + \rho_2(\boldsymbol{k})\delta(k^0 + \omega_{\boldsymbol{k}}),$$
and reality condition $\overline{\hat{\varphi}(k)} = \hat{\varphi}(-k)$ gives $\rho_2(\boldsymbol{k}) = \overline{\rho_1(-\boldsymbol{k})}$. For future convenience[1], we rewrite the previous formula in the following form

$$\hat{\varphi}(k) = \frac{\sqrt{2\pi}}{\sqrt{2\omega_{\boldsymbol{k}}}}\left(a(\boldsymbol{k})\delta(k^0 - \omega_{\boldsymbol{k}}) + \overline{a(-\boldsymbol{k})}\delta(k^0 + \omega_{\boldsymbol{k}})\right), \tag{13.19}$$

where $a(\boldsymbol{k})$ is an arbitrary complex-valued distribution on $\mathbb{R}^3$. Substituting this $\hat{\varphi}(k)$ into the inverse Fourier transform
$$\varphi(x) = \frac{1}{(2\pi)^2}\int_{\mathbb{R}^4} e^{-ik\cdot x}\hat{\varphi}(k)d^4k,$$

[1] To have simple Poisson brackets in Theorem 13.10.

and changing in the second integral $\boldsymbol{k}$ by $-\boldsymbol{k}$ we obtain

$$\varphi(x) = \frac{1}{(2\pi)^{\frac{3}{2}}} \int_{\mathbb{R}^3} \left(a(\boldsymbol{k})e^{-ik\cdot x} + \bar{a}(\boldsymbol{k})e^{ik\cdot x}\right) \frac{d^3\boldsymbol{k}}{\sqrt{2\omega_{\boldsymbol{k}}}}, \quad \text{where} \quad k = (\omega_{\boldsymbol{k}}, \boldsymbol{k}). \tag{13.20}$$

Formula (13.20) gives a general solution of the Klein–Gordon equation; it depends on an arbitary complex-valued distribution $a(\boldsymbol{k})$.

Remark 13.8. The Minkowski metric induces a pseudo-Riemannian metric of signature $(-,-,-)$ on the hyperboloid $\mathcal{O}_m$. A simple computation shows that the Lorentz invariant volume form of this metric is proportional to $\dfrac{d^3\boldsymbol{k}}{2\omega_{\boldsymbol{k}}}$. In physics textbooks this volume form is introduced by

$$\delta(k_0^2 - \omega_{\boldsymbol{k}}^2) = \frac{1}{2\omega_{\boldsymbol{k}}}\left(\delta(k_0 - \omega_{\boldsymbol{k}}) + \delta(k_0 + \omega_{\boldsymbol{k}})\right),$$

which follows from the elementary change of variables formula in the theory of distributions.

As discussed in the previous section, this also allows us to give a description of the phase space $\mathscr{X}$. Namely, for $(\pi(\boldsymbol{x}), \varphi(\boldsymbol{x})) \in \mathscr{X} = \mathscr{S}(\mathbb{R}^3, \mathbb{R}) \times \mathscr{S}(\mathbb{R}^3, \mathbb{R})$ consider the solution $\varphi(x)$ of the Klein–Gordon equation with the initial conditions

$$\varphi(0, \boldsymbol{x}) = \varphi(\boldsymbol{x}) \quad \text{and} \quad \partial_0\varphi(0, \boldsymbol{x}) = \pi(\boldsymbol{x}).$$

From

$$\begin{aligned}
\varphi(\boldsymbol{x}) &= \frac{1}{(2\pi)^{\frac{3}{2}}} \int_{\mathbb{R}^3} \left(a(\boldsymbol{k})e^{i\boldsymbol{k}\boldsymbol{x}} + \bar{a}(\boldsymbol{k})e^{-i\boldsymbol{k}\boldsymbol{x}}\right) \frac{d^3\boldsymbol{k}}{\sqrt{2\omega_{\boldsymbol{k}}}} = \frac{1}{(2\pi)^{\frac{3}{2}}} \int_{\mathbb{R}^3} \left(a(\boldsymbol{k}) + \bar{a}(-\boldsymbol{k})\right) e^{i\boldsymbol{k}\boldsymbol{x}} \frac{d^3\boldsymbol{k}}{\sqrt{2\omega_{\boldsymbol{k}}}}, \\
\pi(\boldsymbol{x}) &= \frac{-i}{(2\pi)^{\frac{3}{2}}} \int_{\mathbb{R}^3} \omega_{\boldsymbol{k}} \left(a(\boldsymbol{k})e^{i\boldsymbol{k}\boldsymbol{x}} - \bar{a}(\boldsymbol{k})e^{-i\boldsymbol{k}\boldsymbol{x}}\right) \frac{d^3\boldsymbol{k}}{\sqrt{2\omega_{\boldsymbol{k}}}} \\
&= \frac{-i}{(2\pi)^{\frac{3}{2}}} \int_{\mathbb{R}^3} \omega_{\boldsymbol{k}} \left(a(\boldsymbol{k}) - \bar{a}(-\boldsymbol{k})\right) e^{i\boldsymbol{k}\boldsymbol{x}} \frac{d^3\boldsymbol{k}}{\sqrt{2\omega_{\boldsymbol{k}}}}
\end{aligned}$$

it is easy to express $a(\boldsymbol{k})$ in terms of Fourier transforms $\hat{\varphi}(\boldsymbol{k})$, $\hat{\pi}(\boldsymbol{k})$ (see Exercise 13.5):

$$a(\boldsymbol{k}) = \frac{1}{\sqrt{2}}\left(\sqrt{\omega_{\boldsymbol{k}}}\,\hat{\varphi}(\boldsymbol{k}) + \frac{i\hat{\pi}(\boldsymbol{k})}{\sqrt{\omega_{\boldsymbol{k}}}}\right), \quad \bar{a}(\boldsymbol{k}) = \frac{1}{\sqrt{2}}\left(\sqrt{\omega_{\boldsymbol{k}}}\,\hat{\varphi}(-\boldsymbol{k}) - \frac{i\hat{\pi}(-\boldsymbol{k})}{\sqrt{\omega_{\boldsymbol{k}}}}\right), \tag{13.21}$$

where $a(\boldsymbol{k}) \in \mathscr{S}(\mathbb{R}^3, \mathbb{C})$, the Schwartz space of *complex-valued* functions on $\mathbb{R}^3$. Using this representation, we can prove the following result.

Lemma 13.9. *Let $\varphi(x)$ be a solution of the Klein-Gordon equation with the Schwarz class initial data at $t = 0$. Then for each $g \in \mathfrak{P}$ solution $\varphi(gx)$ also has Schwarz class initial data.*

Proof. It follows from (13.20) that the statement is trivial for the subgroup $\mathbb{R}^4$ of spacetime translations. To prove it for the Lorentz group $\mathfrak{L}$, we rewrite (13.20) as

$$\varphi(x) = \frac{1}{(2\pi)^{\frac{3}{2}}} \int_{\mathbb{R}^3} \left(\tilde{a}(\boldsymbol{k})e^{-ik\cdot x} + \overline{\tilde{a}(\boldsymbol{k})}e^{ik\cdot x}\right) \frac{d^3\boldsymbol{k}}{2\omega_{\boldsymbol{k}}},$$

where $k = (\omega_{\boldsymbol{k}}, \boldsymbol{k}) \in \mathcal{O}_m$ and $\tilde{a}(\boldsymbol{k}) = \sqrt{2\omega_{\boldsymbol{k}}}a(\boldsymbol{k}) \in \mathscr{S}(\mathbb{R}^3, \mathbb{C})$. Using invariance of the volume form on $\mathcal{O}_m$ under $\mathfrak{L}$, it is sufficient to verify the statement for the restricted Lorentz group $\mathfrak{L}_+^\uparrow$ (see Section 7.2). We have for $g \in \mathfrak{L}_+^\uparrow$,

$$\begin{aligned}\varphi(gx) &= \frac{1}{(2\pi)^{\frac{3}{2}}}\int_{\mathbb{R}^3}\left(\tilde{a}(\boldsymbol{k})e^{-ik\cdot gx} + \overline{\tilde{a}(\boldsymbol{k})}e^{ik\cdot gx}\right)\frac{d^3\boldsymbol{k}}{2\omega_{\boldsymbol{k}}} \\ &= \frac{1}{(2\pi)^{\frac{3}{2}}}\int_{\mathbb{R}^3}\left(\tilde{a}(\boldsymbol{k})e^{-ig^{-1}k\cdot x} + \overline{\tilde{a}(\boldsymbol{k})}e^{ig^{-1}k\cdot x}\right)\frac{d^3\boldsymbol{k}}{2\omega_{\boldsymbol{k}}} \\ &= \frac{1}{(2\pi)^{\frac{3}{2}}}\int_{\mathbb{R}^3}\left(\tilde{a}(g\cdot\boldsymbol{k})e^{-ik\cdot x} + \overline{\tilde{a}(g\cdot\boldsymbol{k})}e^{ik\cdot x}\right)\frac{d^3\boldsymbol{k}}{2\omega_{\boldsymbol{k}}}.\end{aligned}$$

where $g\cdot\boldsymbol{k}$ stands for the action of $g \in \mathfrak{L}_+^\uparrow$ on $\mathbb{R}^3$ induced by the isomorphism $\mathcal{O}_m^+ \simeq \mathbb{R}^3$. Clearly $\tilde{a}(g\cdot\boldsymbol{k})$ is a complex-valued Schwarz function, so the statement follows. □

The formulas (13.21) give an isomorphism of the phase space $\mathscr{X}$ and the Schwartz space $\mathscr{S}(\mathbb{R}^3, \mathbb{C})$, which is a complex manifold with complex coordinates $a(\boldsymbol{k})$. It is very beneficial to use a complex geometry language in field theory (especially for the purposes of quantization), and to treat $a(\boldsymbol{k})$ and their complex conjugates $\bar{a}(\boldsymbol{k})$ as coordinates on $\mathscr{X}$. We have the following simple result.

Theorem 13.10. *Under the identification $\mathscr{X} \simeq \mathscr{S}(\mathbb{R}^3, \mathbb{C})$, we have the following formulas for the Poisson bracket for the complex coordinates*

$$\{a(\boldsymbol{k}), a(\boldsymbol{l})\} = \{\bar{a}(\boldsymbol{k}), \bar{a}(\boldsymbol{l})\} = 0 \quad \text{and} \quad \{a(\boldsymbol{k}), \bar{a}(\boldsymbol{l})\} = i\delta(\boldsymbol{k}-\boldsymbol{l}). \tag{13.22}$$

The Hamiltonian is given by

$$H = \int_{\mathbb{R}^3}\omega_{\boldsymbol{k}}\,\bar{a}(\boldsymbol{k})a(\boldsymbol{k})\,d^3\boldsymbol{k}. \tag{13.23}$$

Proof. Formulas for the Poisson bracket immediately follow from the formulas for Poisson brackets of $\hat{\varphi}$, $\hat{\pi}$ given in Exercise 13.5:

$$\{\hat{\pi}(\boldsymbol{k}), \hat{\pi}(\boldsymbol{l})\} = \{\hat{\varphi}(\boldsymbol{k}), \hat{\varphi}(\boldsymbol{l})\} = 0 \quad \text{and} \quad \{\hat{\pi}(\boldsymbol{k}), \hat{\varphi}(\boldsymbol{l})\} = \delta(\boldsymbol{k}+\boldsymbol{l}).$$

For the Hamiltonian, we use Plancherel's theorem, which implies

$$\begin{aligned}H &= \frac{1}{2}\int_{\mathbb{R}^3}\left(\pi^2(\boldsymbol{x}) + (\boldsymbol{\nabla}\varphi)^2(\boldsymbol{x}) + m^2\varphi^2(\boldsymbol{x})\right)d^3\boldsymbol{x} \\ &= \frac{1}{2}\int_{\mathbb{R}^3}\left(|\hat{\pi}(\boldsymbol{k})|^2 + \omega_{\boldsymbol{k}}^2|\hat{\varphi}(\boldsymbol{k})|^2\right)d^3\boldsymbol{k} \\ &= \int_{\mathbb{R}^3}\omega_{\boldsymbol{k}}\,\bar{a}(\boldsymbol{k})a(\boldsymbol{k})\,\mathrm{d}^3\,\boldsymbol{k}.\end{aligned}$$ □

An important corollary of these computations is that when rewritten in terms of Fourier modes, Hamilton's equations (13.6)–(13.7) decouple:

$$\begin{aligned}\dot{a}(\boldsymbol{k}) &= \{H, a(\boldsymbol{k})\} = -i\omega_{\boldsymbol{k}}a(\boldsymbol{k}), \\ \dot{\bar{a}}(\boldsymbol{k}) &= \{H, \bar{a}(\boldsymbol{k})\} = \quad i\omega_{\boldsymbol{k}}\bar{a}(\boldsymbol{k}),\end{aligned}$$

so we can easily solve them:

$$a(t,\boldsymbol{k}) = e^{-i\omega_{\boldsymbol{k}}t}a(\boldsymbol{k}), \quad \bar{a}(t,\boldsymbol{k}) = e^{i\omega_{\boldsymbol{k}}t}\bar{a}(\boldsymbol{k}).$$

One can also introduce real coordinates in the Fourier transformed phase space:

$$P(\boldsymbol{k}) = \frac{\sqrt{\omega_k}}{\sqrt{2}}(a(\boldsymbol{k})+\bar{a}(\boldsymbol{k})), \quad Q(\boldsymbol{k}) = \frac{i}{\sqrt{2\omega_{\boldsymbol{k}}}}(a(\boldsymbol{k})-\bar{a}(\boldsymbol{k})).$$

We leave it to the reader to check that in these coordinates, the Poisson brackets are given by

$$\{P(\boldsymbol{k}), P(\boldsymbol{l})\} = \{Q(\boldsymbol{k}), Q(\boldsymbol{l})\} = 0, \qquad \{P(\boldsymbol{k}), Q(\boldsymbol{l})\} = \delta(\boldsymbol{k}-\boldsymbol{l})$$

so that the symplectic form Ω is given by

$$\Omega = \int_{\mathbb{R}^3} (\delta P(\boldsymbol{k}) \wedge \delta Q(\boldsymbol{k}))\, d^3\boldsymbol{k},$$

and the Hamiltonian of the Klein-Gordon model takes the form

$$H = \frac{1}{2}\int_{\mathbb{R}^3} \left(P^2(\boldsymbol{k}) + \omega_{\boldsymbol{k}}^2 Q^2(\boldsymbol{k})\right) d^3\boldsymbol{k}.$$

Comparing this with the Hamlitonian description of a system of harmonic oscillators given in Section 4.3, we see that a free scalar field is a completely integrable Hamiltonian system, which in terms of Fourier modes is described by infinitely many non-interacting harmonic oscillators with the frequencies $\omega_{\boldsymbol{k}} = \sqrt{\boldsymbol{k}^2+m^2}$, parametrized by $\boldsymbol{k} \in \mathbb{R}^3$.

13.6. Exercises

Exercise 13.1. Find the Hamiltonian H for the Lagrangian (11.36) in Example 11.5, and show that the term $\pi_0^2(\boldsymbol{x})$, where $\pi_\mu(\boldsymbol{x})$ are canonical momenta, enters H with a negative sign.

Exercise 13.2. Prove that for each t solutions of the Hamilton equations of motion (13.6)–(13.7) satisfy the same Poisson brackets (13.5).

Exercise 13.3.

(1) Let $\mathscr{H}(\boldsymbol{x})$ be the Hamiltonian density defined in Theorem 13.6; we consider it as a functional on the phase space $\mathscr{X}$. Prove that then

$$\{\mathscr{H}(\boldsymbol{x}), \varphi(\boldsymbol{y})\} = \pi(\boldsymbol{x})\delta(\boldsymbol{x}-\boldsymbol{y})$$
$$\{\mathscr{H}(\boldsymbol{x}), \pi(\boldsymbol{y})\} = -m^2\varphi(\boldsymbol{x})\delta(\boldsymbol{x}-\boldsymbol{y}) - \boldsymbol{\nabla}\varphi(\boldsymbol{x})\cdot\boldsymbol{\nabla}_{\boldsymbol{x}}\delta(\boldsymbol{x}-\boldsymbol{y})$$
$$\{\mathscr{H}(\boldsymbol{x}), \mathscr{H}(\boldsymbol{y})\} = \pi(\boldsymbol{x})\boldsymbol{\nabla}\varphi(y)\cdot\boldsymbol{\nabla}_{\boldsymbol{y}}\delta(\boldsymbol{x}-\boldsymbol{y}) - \pi(\boldsymbol{y})\boldsymbol{\nabla}\varphi(\boldsymbol{x})\boldsymbol{\nabla}_{\boldsymbol{x}}\delta(\boldsymbol{x}-\boldsymbol{y}).$$

(2) Derive the Poisson brackets (13.13)–(13.15).

Exercise 13.4. Prove the formula for generators K_i of Lorentz boosts, given in Theorem 13.6, and prove the formulas for Poisson brackets (13.13)–(13.15).

Exercise 13.5. Define Fourier transform of coordinates $\pi(\boldsymbol{x})$, $\varphi(\boldsymbol{x})$ on the phase space $\mathscr{X}$ of a free scalar field by

$$\hat{\varphi}(\boldsymbol{k}) = \frac{1}{(2\pi)^{3/2}}\int \varphi(\boldsymbol{x})e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}\, d^3\boldsymbol{x}$$
$$\hat{\pi}(\boldsymbol{k}) = \frac{1}{(2\pi)^{3/2}}\int \pi(\boldsymbol{x})e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}\, d^3\boldsymbol{x} \tag{13.24}$$

(note the sign in the exponent).

Prove that then the Poisson brackets of these observables are given by

$$\{\hat{\pi}(\boldsymbol{k}),\hat{\pi}(\boldsymbol{l})\}=\{\hat{\varphi}(\boldsymbol{k}),\hat{\varphi}(\boldsymbol{l})\}=0 \quad\text{and}\quad \{\hat{\pi}(\boldsymbol{k}),\hat{\varphi}(\boldsymbol{l})\}=\delta(\boldsymbol{k}+\boldsymbol{l}).$$

Exercise 13.6. Show that the total momentum $\boldsymbol{P}=(P^1,P^2,P^3)$ of a free scalar field in terms of $a(\boldsymbol{k})$ is given by

$$\boldsymbol{P}=\int_{\mathbb{R}^3}\boldsymbol{k}\bar{a}(\boldsymbol{k})a(\boldsymbol{k})\,d^3\boldsymbol{k}. \tag{13.25}$$

Exercise 13.7. Show that the generators $L_{jl}=\epsilon_{jlk}\hat{J}_k$ of the angular momentum and generators $L_{0j}=\hat{K}_j$ of the Lorentz boosts of a free scalar field in terms of $a(\boldsymbol{k})$ are given by

$$L_{jl}=i\int_{\mathbb{R}^3}\bar{a}(\boldsymbol{k})\left(k_j\frac{\partial}{\partial k^l}-k_l\frac{\partial}{\partial k^j}\right)\big(\sqrt{2\omega_{\boldsymbol{k}}}a(\boldsymbol{k})\big)\,\frac{d^3\boldsymbol{k}}{\sqrt{2\omega_{\boldsymbol{k}}}},$$

$$L_{0j}=i\int_{\mathbb{R}^3}\omega_{\boldsymbol{k}}\bar{a}(\boldsymbol{k})\frac{\partial}{\partial k^j}\big(\sqrt{2\omega_{\boldsymbol{k}}}a(\boldsymbol{k})\big)\,\frac{d^3\boldsymbol{k}}{\sqrt{2\omega_{\boldsymbol{k}}}}.$$

Exercise 13.8. Verify that for the coordinates on $\mathscr{X}$ given by

$$\rho(\boldsymbol{k})=|a(\boldsymbol{k})|^2 \quad\text{and}\quad \theta(\boldsymbol{k})=-\arg a(\boldsymbol{k}),$$

the symplectic form and the free scalar field Hamiltonian take the form

$$\Omega=\int_{\mathbb{R}^3}(\delta\rho(\boldsymbol{k})\wedge\delta\theta(\boldsymbol{k}))d^3\boldsymbol{k} \quad\text{and}\quad H=\int_{\mathbb{R}^3}\omega_{\boldsymbol{k}}\rho(\boldsymbol{k})\,d^3\boldsymbol{k}$$

(such canonical coordinates are called the *action-angle coordinates*, see Theorem 4.23).

Notes and References

Every physics textbook starts with the Lagrangian formulation of classical field theory, see e.g. [**BS1983**], [**Ryd1996**], and the monograph [**DFN1984**] for the discussion from the mathematics perspective. Inspired by the exposition in [**DF1999**], in Chapter 11 we present Lagrangian formalism in a more invariant form. We define classical fields in Section 11.2 as smooth sections of some vector bundle over the space-time, so this definition does not include gauge fields and gravitational fields, to be considered later in Parts III and IV.

In Section 11.3 we start with the invariant formulation of the classical multi-dimensional calculus of variations (see, e.g. [**GF1963**]), give an invariant definition of the Lagrangian density in Section 11.4, and in Section 11.5 derive general field-theoretic Euler-Lagrange equations. As in [**DF1999**], we introduce the notion of local Lagrangian densities and in Theorem 11.7 derive the Euler-Lagrange equations in a standard form as in physics textbooks, and in Theorem 11.9 give their invariant formulation. Following [**DF1999**], in Chapter **??** we introduce the notion of infinitesimal symmetries and give an invariant formulation of the Noether theorem in classical field theory. In Sections 12.2 and 12.3 we define the stress-energy tensor, and for simple relativistic field theories present explicit conservation laws: the conservation of charge, of total energy-momentum and of angular momentum.

In Chapter 13 we develop the Hamiltonian formalism for the scalar field. Specifically, in Section 13.1 we define admissible functionals on the phase space of a scalar field and show that they form the algebra of observables. In Sections 13.2 and 13.3 we introduce, rather at a formal level, the symplectic form and a Poisson bracket for the scalar field observables, and show how the Hamiltonian formalism extends to a classical field theory. In Section 13.4 we explicitly describe the Hamiltonian action of the Poincaré group on the phase space of a free relativistic real scalar field, which is analogous to the action of Poincaré group on the phase space of a free relativistic particle, discussed in Section 8.5. This shows how relativistic invariance is formulated in Hamiltonian formalism. We leave it to the interested reader to turn these computations into rigorous statements in some well-defined version of symplectic geometry on an infinite-dimensional manifold. Finally, in Section 13.5, using the Fourier transfrom, we explicitly solved the Klein-Gordon equation with Schwarz class initial data and show that it is completely integrable Hamiltonian system. The phase space of the theory has a natural structure of a complex manifold, and in Theorem 13.10 and Exercises 13.6 and 13.7 we express Hamiltonian functionals for the Poincaré group action in terms of the complex coordinates.

Part 3

Classical Gauge Theories

Chapter 15

Maxwell's Equations

In this chapter, we describe what is probably the best known classical field theory: Maxwell's theory of electromagnetism. We will concentrate on the mathematical formalism, referring the reader to numerous physics textbooks for practical applications and computations.

Throughout this chapter, we work in Minkowski spacetime $M \simeq \mathbb{R}^{1,3}$. As before, we use coordinates $x^0 = ct$, x^1, x^2, x^3 and the metric $\eta = \mathrm{diag}(1, -1, -1, -1)$. We will also commonly write points in $\mathbb{R}^{1,3}$ as pairs $(\boldsymbol{r}, t)$, where $\boldsymbol{r} = (x^1, x^2, x^3)$.

15.1. Physics formulation

Maxwell's electromagnetic theory describes interactions between electrically charged particles and fields created by such particles. From mathematical standpoint, the objects described by this theory are

- Electric field $\boldsymbol{E}(\boldsymbol{r}, t)$ and magnetic field $\boldsymbol{B}(\boldsymbol{r}, t)$, both of which are time-dependent vector fields on $\mathbb{R}^3$.
- Electric charges, described by a (signed) real-valued charge density function $\rho(\boldsymbol{r}, t)$ so that the total charge of a volume $V \subset \mathbb{R}^3$ (at time t) is

  $$Q(t) = \int_V \rho(\boldsymbol{r}, t)\, d^3\boldsymbol{r}.$$

  It is also common to allow the density function to be a distribution; in particular, physicists commonly consider density functions of the form

  $$\rho(\boldsymbol{r}) = \sum_{a=1}^{N} e_a \delta(\boldsymbol{r} - \boldsymbol{r}_a)$$

  which describe collection of N point particles, each with charge e_a and position $\boldsymbol{r}_a$.
- Currents, which describe motion of charges. A current is described by a time-dependent vector field $\boldsymbol{j}(\boldsymbol{r}, t)$. In the simplest case of point particles, charge e_0 at

a moving point $\boldsymbol{r}_0(t)$ produces a current

$$\boldsymbol{j}(\boldsymbol{r},t) = e_0\boldsymbol{v}(t)\delta(\boldsymbol{r}-\boldsymbol{r}_0(t)), \quad \text{where} \quad \boldsymbol{v}(t) = \frac{d\boldsymbol{r}_0(t)}{dt}$$

(as before, coefficients of this vector field are distributions, not smooth functions).

In general, relation between the current and charge density is given by the *charge conservation law*: if we denote by $Q(t)$ the total charge contained in (time-independent) region $V \subset \mathbb{R}^3$, then the time derivative $\frac{dQ}{dt}$ is equal to the total flux of vector field $\boldsymbol{j}$ through the boundary of V with the minus sign:

$$\frac{\partial}{\partial t}\int_V \rho(t,\boldsymbol{r})d^3\boldsymbol{r} = -\int_{\partial V} \boldsymbol{j}\cdot dS.$$

This gives the *continuity equation*:

$$\nabla\cdot\boldsymbol{j} = -\frac{\partial\rho}{\partial t} \tag{15.1}$$

where, as usual in the multivariable calculus,

$$\nabla\cdot\boldsymbol{j} = \sum\frac{\partial j_i}{\partial x^i}$$

is the divergence of vector field $\boldsymbol{j} = (j_1, j_2, j_3)$.

From physical standpoint, $\boldsymbol{E}$ is generated by electric charges, and $\boldsymbol{B}$ by moving charges. Exact meaning of this is captured by Maxwell equations, which summarize the basic laws of electromagnetism. In a free space these equations have the following beautiful form[1]

$$\nabla\cdot\boldsymbol{E} = \frac{1}{\varepsilon_0}\rho \qquad \text{(Gauss law)} \tag{15.2}$$

— the electric flux leaving a volume is proportional to the charge inside;

$$\nabla\cdot\boldsymbol{B} = 0 \qquad \text{(Gauss law for magnetism)} \tag{15.3}$$

— there are no magnetic charges, the total magnetic flux through a closed surface is zero;

$$\nabla\times\boldsymbol{E} = -\frac{\partial\boldsymbol{B}}{\partial t} \qquad \text{(Faraday's induction law)} \tag{15.4}$$

— the voltage induced in a closed circuit is proportional to the rate of change of the magnetic flux it encloses;

$$\nabla\times\boldsymbol{B} = \mu_0\boldsymbol{j} + \mu_0\varepsilon_0\frac{\partial\boldsymbol{E}}{\partial t} \qquad \text{(Ampère's circular law with Maxwell's addition)} \tag{15.5}$$

— 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 the *permitivity of the free space* and the constant μ_0 is called the *permeability of the free space* or *magnetic constant.* They satisfy

$$\mu_0\varepsilon_0 = \frac{1}{c^2}, \tag{15.6}$$

[1] We are using standard notations for the divergence and curl from the multivariable calculus.

where c is the speed of light in vacuum.[2]

Maxwell equations imply all laws of the electromagnetism: Coulomb law, Bio-Laplace-Savart law, etc.

15.2. Maxwell's equations in the language of differential forms

In order to better understand the mathematical structure of Maxwell equations, let us rewrite them using the language of differential forms. To do that, we first remind some basic facts about Hodge $\star$ operator.

Let V be an oriented n-dimensional real vector space with a non-degenerate symmetric bilinear form $(\ ,\ )$ (not necessarily positive-definite). We extend this form to exterior power space $\Lambda^k V$ by

$$(u_1 \wedge \cdots \wedge u_k, v_1 \wedge \cdots \wedge v_k) = \det((u_i, v_j)).$$

Let $\omega \in \Lambda^n V$ be the volume form on V — an element in $\Lambda^n V$ with positive orientation and such that $(\omega, \omega) = 1$.

Then for $v \in \Lambda^k V$ its Hodge dual is a vector $\star v \in \Lambda^{n-k} V$, satisfying

$$u \wedge \star v = (u, v)\omega \quad \text{for all} \quad u \in \Lambda^k V \tag{15.7}$$

(it is easy to see that this uniquely defines $\star v$).

In particular, given an oriented pseudo-Riemannian manifold X of dimension n, we can apply the Hodge $\star$ operator in each $V = T_x^* X$, getting an isomorphism

$$\star\colon \Omega^k(X) \cong \Omega^{n-k}(X) \tag{15.8}$$

which is linear over $C^\infty(X)$.

It is easy to give explicit formulas for $\star$. Namely, assume that in local coordinates x^i the volume form ω is given by

$$\omega = dx^1 \wedge \cdots \wedge dx^n = \frac{1}{n!}\epsilon_I dx^I \tag{15.9}$$

where we assume summation over all multiindices $I = (i_1, \dots, i_n)$ (we do not assume that $i_1 < \cdots < i_n$), $dx^I = dx^{i_1} \wedge \cdots \wedge dx^{i_n}$, and ϵ_I is the fully antisymmetric tensor:

$$\epsilon_I = \begin{cases} \operatorname{sgn}(\sigma) & \text{if } i_k = \sigma(k) \text{ for some } \sigma \in S_n \\ 0 & \text{otherwise} \end{cases} \tag{15.10}$$

In this case, consider a k-form

$$F = \frac{1}{k!} F_I dx^I,$$

[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 $\boldsymbol{E}_{\mathrm{CGS}} = c^{-1}\boldsymbol{E}_{\mathrm{SI}}$.

where as before we assume summation over all I with $|I| = k$, and F_I is antisymmetric. Then one can show (see Exercise 15.2) that

$$\begin{aligned} \star F &= \frac{1}{(n-k)!}(\star F)_J\, dx^J, \qquad |J| = n-k, \\ (\star F)_J &= \frac{1}{k!} F^I \epsilon_{IJ}, \end{aligned} \tag{15.11}$$

where $\epsilon_{IJ} = \epsilon_{i_1 \dots i_k j_1 \dots j_{n-k}}$ is the fully antisymmetric tensor (15.10), and F^I are the components of the polyvector field $\tilde{F} \in \Lambda^k T(X)$ corresponding to F under the isomorphsim $\Omega^k(X) \simeq \Lambda^k T(X)$ given by η:

$$F^I = \eta^{IJ} F_J$$

where $\eta^{IJ} = \eta^{i_1 j_1} \dots \eta^{i_k j_k}$ and $|J| = k$. Physicists usually say that F^I is obtained from F_J by "rasing the indices".

In particular, this implies

$$\star dx^I = \frac{1}{k!} \sum_{\sigma \in S_k} \operatorname{sgn}(\sigma) \eta^{I\sigma(J)} \epsilon_{JJ'} dx^{J'}, \tag{15.12}$$

where $|J| = k$ and J' is the complement of J: $J' = \{1, \dots, n\} - J$, written in increasing order: $j'_1 < j'_2 < \dots < j'_{n-k}$ (thus, there is summation over J but no summation over J' in the last formula, as J' is fully determined by J).

This and other properties of Hodge $\star$ operator are given in exercises at the end of this chapter. Another useful property discussed in the exercises is the following:

$$\begin{aligned} \star^2 &= 1 \qquad \text{on } \Omega^p(\mathbb{R}^3), \\ \star^2 &= (-1)^{p+1} \text{ on } \Omega^p(\mathbb{R}^{1,3}). \end{aligned} \tag{15.13}$$

For example, for $X = \mathbb{R}^3$ with the usual (positive definite) inner product

$$ds^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2,$$

and orientation defined by $\omega = dx^1 \wedge dx^2 \wedge dx^3$, the Hodge $\star$ operator is given by

$$\begin{aligned} \star dx^i &= \tfrac{1}{2} \epsilon_{ijk} dx^j \wedge dx^k, \\ \star(dx^i \wedge dx^j) &= \epsilon_{ijk} dx^k. \end{aligned} \tag{15.14}$$

E.g. we have $\star dx^1 = dx^2 \wedge dx^3$ and $\star(dx^2 \wedge dx^3) = dx^1$, etc.

Moreover, in this case one can identify the space of one-forms and space of vector fields using the inner product: $\mathrm{Vect}(\mathbb{R}^3) \simeq \Omega^1(\mathbb{R}^3)$. Combining it with the Hodge $\star$ operator, we can also identify $\Omega^2(\mathbb{R}^3) \simeq \Omega^1(\mathbb{R}^3) \simeq \mathrm{Vect}(\mathbb{R}^3)$ and $\Omega^3(\mathbb{R}^3) \simeq \Omega^0(\mathbb{R}^3) = C^\infty(\mathbb{R}^3)$. Under these identifications, the exterior derivative operator $\Omega^1(\mathbb{R}^3) \to \Omega^2(\mathbb{R}^3)$ becomes the curl operator

$$\mathrm{Vect}(\mathbb{R}^3) \to \mathrm{Vect}(\mathbb{R}^3) \colon \xi \mapsto \boldsymbol{\nabla} \times \xi,$$

and $d \colon \Omega^2(\mathbb{R}^3) \to \Omega^3(\mathbb{R}^3)$ becomes the divergence operator

$$\mathrm{Vect}(\mathbb{R}^3) \to C^\infty(\mathbb{R}^3) \colon \xi \mapsto \boldsymbol{\nabla} \cdot \xi.$$

Thus, the de Rham complex of differential forms on $\mathbb{R}^3$ becomes

$$\begin{array}{ccccccccc} 0 & \longrightarrow & C^\infty(\mathbb{R}^3) & \xrightarrow{\nabla} & \mathrm{Vect}(\mathbb{R}^3) & \xrightarrow{\nabla\times} & \mathrm{Vect}(\mathbb{R}^3) & \xrightarrow{\nabla\cdot} & C^\infty(\mathbb{R}^3) & \longrightarrow & 0 \\ & & \| & & \downarrow{\scriptstyle \xi\mapsto\tilde\xi} & & \downarrow{\scriptstyle \xi\mapsto\star\tilde\xi} & & \downarrow{\scriptstyle f\mapsto f\,d^3\boldsymbol{x}} & & \\ 0 & \longrightarrow & \Omega^0(\mathbb{R}^3) & \xrightarrow{d} & \Omega^1(\mathbb{R}^3) & \xrightarrow{d} & \Omega^2(\mathbb{R}^3) & \xrightarrow{d} & \Omega^3(\mathbb{R}^3) & \longrightarrow & 0 \end{array}$$

where for a vector field $\xi = \xi^i\partial_i$, we define one form $\tilde\xi$ by

$$\tilde\xi = \xi_i dx^i, \quad \xi_i = g_{ij}\xi^j$$

where g_{ij} is metric in $\mathbb{R}^3$. Since in the standard coordinates $g_{ij} = \delta_{ij}$, in this case we simply get $\xi_i = \xi^i$, so $\tilde\xi = \xi_i dx^i$.

Going back to Maxwell's equations, define the time-dependent one-form $\boldsymbol{E}^{(1)} \in \Omega^1(\mathbb{R}^3)$ and the 2-form $\boldsymbol{B}^{(2)} = \star\boldsymbol{B}^{(1)} \in \Omega^2(\mathbb{R}^3)$, corresponding to vector fields $\boldsymbol{E}$, $\boldsymbol{B}$ respectively under isomorphisms $\mathrm{Vect}(\mathbb{R}^3) \simeq \Omega^1(\mathbb{R}^3)$, $\mathrm{Vect}(\mathbb{R}^3) \simeq \Omega^2(\mathbb{R}^3)$ described above:

$$\begin{aligned} \boldsymbol{E}^{(1)} &= E_1 dx^1 + E_2 dx^2 + E_3 dx^3, \\ \boldsymbol{B}^{(2)} &= B_1 dx^2 \wedge dx^3 + B_2 dx^3 \wedge dx^1 + B_3 dx^1 \wedge dx^2. \end{aligned}$$

Then Maxwell equations (15.3) and (15.4) become

$$\begin{aligned} d\boldsymbol{B}^{(2)} &= 0, \\ d\boldsymbol{E}^{(1)} &= -\frac{\partial \boldsymbol{B}^{(2)}}{\partial t}. \end{aligned} \tag{15.15}$$

making them easy to understand. (Here d is the exterior derivative for forms in $\mathbb{R}^3$.) The other two equations will be discussed shortly.

Moreover, equations (15.15) can in fact be united into a single equation. Namely, let us define the following 2-form in 4-dimensional spacetime $\mathbb{R}^{1,3}$, called the *electromagnetic field tensor*,

$$\begin{aligned} F &= \frac{1}{c} dx^0 \wedge \boldsymbol{E}^{(1)} - \boldsymbol{B}^{(2)} \\ &= \tfrac{1}{c} E_1\, dx^0 \wedge dx^1 + \tfrac{1}{c} E_2\, dx^0 \wedge dx^2 + \tfrac{1}{c} E_3\, dx^0 \wedge dx^3 \\ &\quad - B_1\, dx^2 \wedge dx^3 - B_2\, dx^3 \wedge dx^1 - B_3\, dx^1 \wedge dx^2. \end{aligned} \tag{15.16}$$

Then it is immediate that equations (15.15) are equivalent to the following equation on 2-form F:

$$dF = 0, \tag{15.17}$$

where d is now the full exterior derivative operator in $\mathbb{R}^{1,3}$.

It is very convenient to write F in coordinates:

$$F = \tfrac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu \tag{15.18}$$

(as before, indices μ, ν run over $0, 1, 2, 3$), where $F_{\mu\nu}$ is represented by the following skew-symmetric 4×4 matrix

$$F_{\mu\nu} = \begin{pmatrix} 0 & \frac{1}{c}E_1 & \frac{1}{c}E_2 & \frac{1}{c}E_3 \\ -\frac{1}{c}E_1 & 0 & -B_3 & B_2 \\ -\frac{1}{c}E_2 & B_3 & 0 & -B_1 \\ -\frac{1}{c}E_3 & -B_2 & B_1 & 0 \end{pmatrix}. \tag{15.19}$$

Then an explicit computation shows that (15.17) is equivalent to the following equation, called the *Bianchi identity*

$$\partial_\nu F_{\lambda\mu} + \partial_\lambda F_{\mu\nu} + \partial_\mu F_{\nu\lambda} = 0, \quad \lambda, \mu, \nu = 0, 1, 2, 3. \tag{15.20}$$

Note that there is no summation in this identity: it has to be satisfied for every triple λ, μ, ν.

This gives an easy way to write two of the four Maxwell equations, namely (15.3) and (15.4).

To rewrite the second pair of Maxwell equations, equations (15.2) and (15.5), we observe that in the absence of the sources (charges and currents), these equations can be obtained from the first pair (15.3)–(15.4) by the *electro-magnetic duality*

$$\frac{1}{c}\boldsymbol{E} \mapsto -\boldsymbol{B} \quad \text{and} \quad \boldsymbol{B} \mapsto \frac{1}{c}\boldsymbol{E}. \tag{15.21}$$

This duality can be easily understood in terms of the 2-form F: it sends F to $\star F$, where $\star$ is the Hodge dual operator in the 4-dimensional space $\mathbb{R}^{1,3}$ with the Minkowski metric $(+, -, -, -)$ (see Exercise 15.4):

$$\begin{aligned} \star F &= -dx^0 \wedge \boldsymbol{B}^{(1)} - \tfrac{1}{c}\boldsymbol{E}^{(2)} \\ &= -B_1\, dx^0 \wedge dx^1 - B_2\, dx^0 \wedge dx^2 - B_3\, dx^0 \wedge dx^3 \\ &\quad - \tfrac{1}{c}E_1\, dx^2 \wedge dx^3 - \tfrac{1}{c}E_2\, dx^3 \wedge dx^1 - \tfrac{1}{c}E_3\, dx^1 \wedge dx^2, \end{aligned} \tag{15.22}$$

so that equations (15.2) and (15.5) can be written as

$$\begin{aligned} d\boldsymbol{E}^{(2)} &= 0, \\ d\boldsymbol{B}^{(1)} &= \frac{1}{c^2}\frac{\partial \boldsymbol{E}^{(2)}}{\partial t}. \end{aligned} \tag{15.23}$$

where d is exterior derivative for forms in $\mathbb{R}^3$, or, equivalently, as a single equation in $\mathbb{R}^{1,3}$:

$$d \star F = 0. \tag{15.24}$$

To summarize, Maxwell equations in an empty space (without sources) can be written succinctly as

$$\begin{aligned} dF &= 0 \\ d \star F &= 0, \end{aligned} \tag{15.25}$$

where $\star$ is the Hodge star operator in the Minkowski spacetime $\mathbb{R}^{1,3}$.

As an immediate corollary, we obtain that Maxwell equations are relativistic invariant.

Theorem 15.1. *Maxwell equations in empty space are invariant under the action of Poincare group on $\mathbb{R}^{1,3}$ if $\boldsymbol{E}$, $\boldsymbol{B}$ transform as components of 2-form F as in* (15.16).

Explicit formulas for transformations of $\boldsymbol{E}, \boldsymbol{B}$ under the action of Poincare group are given in Exercise 15.6.

Remark 15.2. The different signs in Maxwell equations (15.4) and (15.5) play a fundamental role, reflected in the electro-magnetic duality. It forces the use of a pseudo-Riemannian metric and naturally introduces the Minkowski spacetime.

15.3. Maxwell's equations with sources

To write Maxwell's equations with sources (charges and currents) in the language of differential forms, we combine the charge density ρ and current $\boldsymbol{j}$ into a single 1-form $J = J_\mu dx^\mu$ in $\mathbb{R}^{1,3}$, defined by

$$J_0 = c\rho, \quad J_1 = -j_1, J_2 = -j_2, J_3 = -j_3, \tag{15.26}$$

so that

$$J = c\rho\, dx^0 - j_1 dx^1 - j_2 dx^2 - j_3 dx^3.$$

We will also frequently use the 3-form $\star J$; using explicit formulas for Hodge $\star$ operator in $\mathbb{R}^{1,3}$ given in Exercise 15.4, one easily sees that

$$\star J = c\rho\, dx^1 \wedge dx^2 \wedge dx^3 - j_1 dx^0 \wedge dx^2 \wedge dx^3 + j_2 dx^0 \wedge dx^1 \wedge dx^3 - j_3 dx^0 \wedge dx^1 \wedge dx^2$$

In this language, the continuity equation

$$\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla} \cdot \boldsymbol{j} = 0$$

is rewritten as

$$d\star J = 0.$$

Now using relation $\mu_0 \varepsilon_0 = 1/c^2$ and formula (15.22), we obtain the following result.

Theorem 15.3. *Let F, J be defined by* (15.26). *Then Maxwell's equations* (15.2) *and* (15.5) *are equivalent to*

$$\star d \star F = -\mu_0 J. \tag{15.27}$$

Equivalently, since $\star^2 = 1$ on the space of 3-forms on $\mathbb{R}^{1,3}$, (15.27) can be rewritten as

$$d\star F = -\mu_0 \star J, \tag{15.28}$$

Note that this immediately implies the continuity equation $d\star J = 0$.

Yet one more way of rewriting (15.27) is given by "raising indices" of F and J. Namely, define

$$\begin{aligned} F^{\mu\nu} &= \eta^{\mu\alpha}\eta^{\nu\beta}F_{\alpha\beta}, \\ J^\mu &= \eta^{\mu\nu}J_\nu, \end{aligned} \tag{15.29}$$

where $\eta^{\mu\nu}$ is the Minkowski metric in $\mathbb{R}^{1,3}$. It follows from the explicit formula (15.11) for the $\star$ operator, that $F^{\mu\nu}$ is related to the dual field tensor by

$$(\star F)_{\mu\nu} = \tfrac{1}{2}\epsilon_{\mu\nu\alpha\beta}F^{\alpha\beta}.$$

Then the second pair of Maxwell equations can be written in the following form

$$\partial_\mu F^{\mu\nu} = \mu_0 J^\nu, \tag{15.30}$$

often used in physics textbooks, and the continuity equation becomes

$$\partial_\mu J^\mu = 0. \tag{15.31}$$

To summarize, the Maxwell's equations on $\mathbb{R}^4$ have the following form

$$\begin{aligned} dF &= 0 \\ \star d \star F &= -\mu_0 J, \end{aligned} \tag{15.32}$$

where the four-current J satisfies the continuity equation $d \star J = 0$.

15.4. Vector potential and Lagrangian

Our next goal is to show that Maxwell's equations can be obtained as a special case of general theory developed in Chapter 11, i.e. as Euler-Lagrange equations for an appropriate Lagrangian. To do that, we need to decide first what we consider to be the fields of the theory. The natural idea is to consider the electric and magnetic fields $\boldsymbol{E}$ and $\boldsymbol{B}$ (or, equivalently, the 2-form F) as the primary fields of our theory. However, for many reasons it is better to start with something else: the four-potential.

By Poincaré lemma, the first Maxwell equation — equation $dF = 0$ in (15.32) — has a solution

$$F = dA \quad \text{where} \quad A = A_\mu dx^\mu,$$

so that

$$F = \tfrac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu, \qquad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{15.33}$$

Physicists call (A_0, A_1, A_2, A_3) the *electromagnetic four-potential.*

It is also common to introduce scalar potential φ and vector potential $\boldsymbol{A}$ defined by

$$A_0 = \frac{1}{c}\varphi, \qquad \boldsymbol{A} = (A^1, A^2, A^3) = -(A_1, A_2, A_3) \tag{15.34}$$

(this last equality is a special case of raising the indices: $A^i = \eta^{ij} A_j$).

Then explicit computation shows that the equation $F = dA$ is equivalent to

$$\begin{aligned} \boldsymbol{B} &= \boldsymbol{\nabla} \times \boldsymbol{A}, \\ \boldsymbol{E} &= -\boldsymbol{\nabla}\varphi - \frac{\partial \boldsymbol{A}}{\partial t}. \end{aligned} \tag{15.35}$$

The 1-form A, a solution to the first Maxwell equation, is defined up to a *gauge transformation*

$$A \mapsto A + df, \tag{15.36}$$

where f is a smooth real-valued function on $\mathbb{R}^{1,3}$. It is natural to identify gauge equivalent 1-forms and define the space $\mathscr{F}$ of classical fields as equivalence classes of A under transformations (15.36),

$$\mathscr{F} = \Omega^1(\mathbb{R}^{1,3})/d\Omega^0(\mathbb{R}^{1,3}). \tag{15.37}$$

In terms of the four-potential A, Maxwell equations are written as

$$\star d \star dA = -\mu_0 J.$$

Equivalently, if we introduce the operator $d^* = \star d \star$ (see Exercise 15.7 for discussion of this operator and its properties), then Maxwell's equations take the form

$$d^* dA = -\mu_0 J. \tag{15.38}$$

In the special case when $J = 0$, this equation simplifies and becomes

$$d \star F = d \star dA = 0. \tag{15.39}$$

Remark 15.4. In Section 18.2 we will reveal that true geometric meaning of electromagnetic four-potential as unitary connection $d + iA$ in the U(1)-line bundle over the Minkowski spacetime. Correspondingly, transformations (15.36) are indeed the gauge transformations, defining the action of the gauge group on the space of connections. In physics terminology, A is called *abelian gauge field*.

General discussion in Section 11.5 in Chapter 11 can be easily adapted to the space $\mathscr{F}$ of classical fields. The Maxwell equations with sources satisfying the continuity equation can be obtained from the principle of least action. Namely, consider the Lagrangian density

$$\mathcal{L}(A) = \mathscr{L}([A]_1) d^4 x,$$

where

$$\mathscr{L}([A]_1) = -\frac{1}{16\pi} F_{\mu\nu} F^{\mu\nu} - \frac{\mu_0}{4\pi} A_\mu J^\mu \quad \text{and} \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{15.40}$$

Equivalently, in coordinate-free form for the Lagrangian density $\mathcal{L}(A)$ we have the following expression (see Exercise 15.2):

$$\mathcal{L}(A) = -\frac{1}{8\pi}(F \wedge \star F + 2\mu_0 A \wedge \star J), \quad \text{where} \quad F = dA. \tag{15.41}$$

In particular, if $J = 0$, then

$$\mathcal{L}(A) = -\frac{1}{8\pi}(F \wedge \star F) = \frac{1}{8\pi}\left(\frac{1}{c^2}\boldsymbol{E}^2 - \boldsymbol{B}^2\right) d^4 x.$$

Now consider the action functional

$$S(A) = \int_{\mathbb{R}^4} \mathcal{L}(A) = -\frac{1}{16\pi} \int_{\mathbb{R}^4} (F_{\mu\nu} F^{\mu\nu} + 4 A_\mu J^\mu) d^4 x, \tag{15.42}$$

where it is tacitly assumed that 1-form J and 2-form $F = dA$ have compact support (or decay sufficiently fast at infinity).

Note that while the Lagrangian density $\mathcal{L}(A)$ is not invariant under gauge transformations (15.36), it easily follows from Stokes theorem and the continuity equation $d \star J = 0$ that the action is gauge-invariant: $S(A + df) = S(A)$.

Then exactly as in Theorem 11.7, the critical points of the action functional satisfy the Euler-Lagrange equations

$$\frac{\partial \mathscr{L}}{\partial A_\nu} - \partial_\mu \frac{\partial \mathscr{L}}{\partial(\partial_\mu A_\nu)} = 0. \tag{15.43}$$

Using (15.40), we readily obtain that equation (15.43) gives the second Maxwell equation $\partial_\mu F^{\mu\nu} = J^\nu$. One can also use a more abstract Theorem 11.9, where the variational 1-form γ is given by the following simple formula

$$\gamma = \delta A_\nu(x) \wedge \left(\frac{\partial \mathscr{L}}{\partial(\partial_\mu A_\nu(x))}\right) \imath_\mu(d^4x) = -\frac{1}{4\pi}\delta A_\nu \wedge F^{\mu\nu}\imath_\mu(d^4x), \tag{15.44}$$

or

$$\gamma = -\frac{1}{4\pi}\delta A \wedge \star F,$$

which are readily obtained from (11.27).

We summarize this discussion it in the following statement.

Proposition 15.5. *The critical points of the action functional $S(A)$ are given by the Maxwell equations.*

This statement can be also verified directly, without using the general formalism in Section 11.5. Namely, using the symmetry property of the Hodge star operator

$$\alpha \wedge \star\beta = \beta \wedge \star\alpha$$

and Stokes' theorem, we have for compactly supported $a \in \Omega^1(\mathbb{R}^4)$,

$$\begin{aligned}\delta S(A) &= \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} S(A + \varepsilon a) \\ &= -\frac{1}{4\pi}\int\limits_{\mathbb{R}^4} (da \wedge \star F + \mu_0 a \wedge \star J) \\ &= -\frac{1}{4\pi}\int\limits_{\mathbb{R}^4} a \wedge (d\star F + \mu_0 \star J).\end{aligned}$$

So equation $\delta S(A) = 0$ for all a yields

$$d\star F = -\mu_0 \star J.$$

15.5. The stress-energy tensor

It is easy to adapt a general discussion in Section 12.2 to the case of free electromagnetic field. In this case the space of classical fields $\mathscr{F}$ is now a quotient (15.37) of the space of 1-forms over the space of exact forms, so the action of the Poincare group $\mathfrak{P}$ on $\mathscr{F}$ is different from the action (12.12) on the space of scalar fields. This also means that we can not use explicit formula for stress–energy tensor given in Section 12.2; instead, we need to repeat the arguments of that section with necessary changes.

Specifically, suppose that a Lie group G acts on a manifold M. Corresponding G-action on differential forms is given by pullbacks, and the Lie algebra $\mathfrak{g}$ of G acts by Lie derivatives; if $X \in \mathfrak{g}$, the Lie derivative L_X is given by the Cartan formula

$$L_X = \imath_X \circ d + d \circ \imath_X.$$

In particular, for a 1-form A we have

$$L_X A = \imath_X F + d(\imath_X A), \quad F = dA.$$

Here the second term represents the gauge transformation (15.36), and the first term is gauge invariant, so the action on the space $\mathscr{F}$ of gauge equivalence classes of 1-forms A is given by

$$L_X(A) = \imath_X F. \tag{15.45}$$

Applying this to the group of space-time translations $G = \mathbb{R}^{1,3}$, where the corresponding vector fields are $X_\mu = \partial_\mu$, it follows from (15.45) that

$$(L_X A)_\nu = F_{\mu\nu},$$

so the vector fields $\xi_\mu = \xi_{X_\mu}$ on $\mathscr{F}$ take the form

$$\xi_\mu = -\int_{\mathbb{R}^4} F_{\mu\nu}(x)\frac{\delta}{\delta A_\nu(x)} d^4x. \tag{15.46}$$

The corresponding conserved current is given by formula (12.14)

$$J_\mu = \imath_{\xi_\mu}\gamma + \imath_\mu \mathcal{L},$$

where $\mathcal{L}$ is given by (15.40) with $J^\mu = 0$, and γ is given by (15.44). It can be written as

$$J_\mu = -T_\mu^\nu \imath_\nu(d^4x),$$

where

$$T_\mu^\nu = \frac{1}{4\pi}\left(F_{\alpha\mu}F^{\nu\alpha} + \frac{1}{4}\delta_\mu^\nu F_{\alpha\beta}F^{\alpha\beta}\right). \tag{15.47}$$

In addition to the conservation laws

$$\partial_\mu T_\nu^\mu = 0,$$

the tensor T_ν^μ is traceless,

$$T_\mu^\mu = 0,$$

and the tensor

$$T^{\mu\nu} = \eta^{\mu\alpha}T_\alpha^\nu = \frac{1}{4\pi}\left(-\eta_{\alpha\beta}F^{\mu\alpha}F^{\nu\beta} + \frac{1}{4}\eta^{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}\right) \tag{15.48}$$

is symmetric, $T^{\mu\nu} = T^{\nu\mu}$.

The tensor $T^{\mu\nu}$ is the *stress-energy tensor* of the electromagnetic field. Its components contain the *energy density*

$$T^{00} = \frac{1}{8\pi}\left(\frac{1}{c^2}\boldsymbol{E}^2 + \boldsymbol{B}^2\right)$$

and the *momentum density*

$$T^{0i} = \frac{1}{4\pi}F^{0k}F^{ik} = \frac{1}{4\pi c}(\boldsymbol{E}\times\boldsymbol{B})_i, \quad i = 1,2,3.$$

The vector $\boldsymbol{S} = \boldsymbol{E}\times\boldsymbol{B}$ is called the *Umov-Poynting vector.*

The conservation law

$$\partial_\mu T^{0\mu} = 0$$

can be written as

$$\frac{\partial T^{00}}{\partial t} = -\boldsymbol{\nabla}\cdot\boldsymbol{S}$$

and implies that the total energy of the electromagnetic field

$$\mathscr{E} = \int_{\mathbb{R}^3} T^{00} d^3\boldsymbol{r}$$

is conserved. Of course, this can be verified directly using Maxwell's equations and the calculus formula

$$\boldsymbol{\nabla} \cdot (\boldsymbol{a} \times \boldsymbol{b}) = \boldsymbol{b} \cdot (\boldsymbol{\nabla} \times \boldsymbol{a}) - \boldsymbol{a} \cdot (\boldsymbol{\nabla} \times \boldsymbol{b}).$$

Remark 15.6. Instead of the transformation law (15.45), physics textbooks often use a naive formula $\partial_{X_\mu} A_\nu = \partial_\mu A_\nu$. It gives a non-symmetric tensor, that requires an ad hoc addition of a divergence-free term to get the symmetric tensor (15.48). We emphasize that the correct action of the vector fields ∂_μ on the space $\mathscr{F}$ is given by (15.45).

15.6. Exercises

Exercise 15.1. Let V be an n-dimensional real vector space with non-degenerate symmetric bilinear form η and orientation given by volume form ω.

(1) Show that $\star 1 = \omega$, and $\star\omega = (-1)^{n_-}$, where n_- is the number of minuses in the signature of η.

(2) Show that for 1-form $\alpha \in \Lambda^1 V = V$, we have $\star\alpha = \imath_{\tilde{\alpha}}\omega$, where $\tilde{\alpha} \in V^*$ is the covector corresponding to α under isomorphism $V \simeq V^*$ given by η: if $\alpha = \alpha_i e^i$, then $\tilde{\alpha} = \eta^{ij}\alpha_i e_j$, where e^i, e_j are dual bases in V, V^* respectively.

In particular, for a 1-form $\alpha = \alpha_i dx^i$ on a manifold M, we have $\star\alpha = \imath_{\tilde{\alpha}}\omega$, where $\tilde{\alpha} = \eta^{ij}\alpha_i\partial_j$.

Exercise 15.2. Let V be an n-dimensional real vector space with non-degenerate symmetric bilinear form η and orientation given by volume form ω. Let e^i, $i = 1, \dots, n$ be a basis in V; as in Section 15.2, for a multiindex $I = (i_1, \dots, i_k)$ we denote $e^I = e^{i_1} \wedge \dots \wedge e^{i_k} \in \Lambda^k V$. Show the following.

(1) $(e^I, e^J) = \sum_{\sigma \in S_k} \eta^{I\sigma(J)} \operatorname{sgn}(\sigma)$, where $\eta^{IJ} = \eta^{i_1 j_1} \dots \eta^{i_k j_k}$.

(2) Let $A = \frac{1}{k!} A_I e^I \in \Lambda^k V$ (as before, the sum is over all multiindices I with $|I| = k$, and A_I is antisymmetric), and similarly for B. Then

$$(A, B) = \frac{1}{k!} A_I B^I$$

where $B^I = \eta^{IJ} B_J$, so that

$$A \wedge \star B = \frac{1}{k!}(A_I B^I)\omega.$$

(3) Assume that the volume form ω is given by

$$\omega = \frac{1}{n!}\epsilon_I e^I, \qquad |I| = n$$

(as before, we assume that ϵ_I is antisymmetric). Then for $B = \frac{1}{k!} B_I e^I \in \Lambda^k V$, we have

$$\star B = \frac{1}{(n-k)!}(\star B)_J\, e^J, \qquad |J| = n - k,$$

where

$$(\star B)_J = \frac{1}{k!} B^I \epsilon_{IJ}$$

Exercise 15.3. We keep notation and assumptions of the previous problem.

(1) Show that for a multiindex I with $|I| = k$,
$$\star e^I = \frac{1}{k!}(e^I, e^J)\epsilon_{JJ'}e^{J'}$$
where $J' = \{1, \dots, n\} - J$ written in the increasing order (thus, there is no summation over J', as it is fully determined by J).

(2) Assume additionally that e^i are orthogonal: $\eta^{ij} = 0$ for $i \neq j$. Show that in this case,
$$\star e^I = \eta^{II}\epsilon_{II'}e^{I'}.$$
In particular, if e^i are orthonormal, then
$$\star e^I = \epsilon_{II'}e^{I'}.$$

(3) Prove that the operator $\star^2 \colon \Lambda^k V \to \Lambda^k V$ is given by
$$\star^2|_{\Lambda^k V} = (-1)^{n_- + k(n-k)},$$
where n_- is the number of minuses in the signature of the form η.

Exercise 15.4. In this problem, we discuss the relation of Hodge $\star$ operator in $\mathbb{R}^3$ and $\mathbb{R}^{1,3}$.

Let N be an oriented n-dimensional Riemannian manifold with (positive-definite) metric g_N; denote by vol_N the volume form on N determined by the orientation and metric. Let $M = \mathbb{R} \times N$ with metric $(dx^0)^2 - g_N$ and volume form $dx^0 \wedge \mathrm{vol}_N$. Denote by $\star_M$, $\star_N$ Hodge operators in M, N respectively.

Show that if $B \in \Omega^k(N)$ is a time-dependent k-form, then
$$\star_M B = dx^0 \wedge \star_N B$$
$$\star_M(dx^0 \wedge B) = (-1)^k \star_N B.$$

Most common application of this formula is when $N = \mathbb{R}^3$, $M = \mathbb{R}^{1,3}$.

Exercise 15.5. Use the previous problem to deduce the following explicit formulas of Hodge $\star$ operation in $\mathbb{R}^{1,3}$.
$$\begin{aligned}
\star(dx^1 \wedge dx^2 \wedge dx^3) &= dx^0, & \star dx^0 &= dx^1 \wedge dx^2 \wedge dx^3,\\
\star(dx^0 \wedge dx^2 \wedge dx^3) &= dx^1, & \star dx^1 &= dx^0 \wedge dx^2 \wedge dx^3,\\
\star(dx^0 \wedge dx^1 \wedge dx^3) &= -dx^2, & \star dx^2 &= -dx^0 \wedge dx^1 \wedge dx^3,\\
\star(dx^0 \wedge dx^1 \wedge dx^2) &= dx^3, & \star dx^3 &= dx^0 \wedge dx^1 \wedge dx^2.
\end{aligned}$$

Exercise 15.6. Consider change of coordinates in $\mathbb{R}^{1,3}$ described by Section 7.2: new coordinates $(x^\mu)'$ are related to old coordinates x^μ by

$$x = \frac{x' + vt'}{\sqrt{1 - \dfrac{v^2}{c^2}}}, \qquad t = \frac{t' + \dfrac{v}{c^2}x'}{\sqrt{1 - \dfrac{v^2}{c^2}}}. \tag{15.49}$$

where we denote for simplicity $x^0 = ct$, $x^1 = x$ and similarly for x', t'; coordinates x^2, x^3 are unchanged (see (7.7)).

Show that then the components of the four-potential A in coordinates x^μ and $(x^\mu)'$ are related by

$$\varphi = \frac{\varphi' + \dfrac{v}{c}A_1}{\sqrt{1 - \dfrac{v^2}{c^2}}}, \qquad A_1 = \frac{A_1' + \dfrac{v}{c}\varphi'}{\sqrt{1 - \dfrac{v^2}{c^2}}}, \quad A_2 = A_2', \quad A_3 = A_3' \tag{15.50}$$

where $\varphi = cA_0$. Deduce from this that the electric and magnetic fields $\boldsymbol{E}$, $\boldsymbol{B}$ transform by

$$E_1 = E_1', \quad E_2 = \frac{E_2' + \frac{v}{c}B_3'}{\sqrt{1-\frac{v^2}{c^2}}}, \quad E_3 = \frac{E_3' - \frac{v}{c}B_2'}{\sqrt{1-\frac{v^2}{c^2}}} \tag{15.51}$$

$$B_1 = B_1', \quad B_2 = \frac{B_2' - \frac{v}{c}E_3'}{\sqrt{1-\frac{v^2}{c^2}}}, \quad B_3 = \frac{B_3' + \frac{v}{c}E_2'}{\sqrt{1-\frac{v^2}{c^2}}} \tag{15.52}$$

Exercise 15.7. Let M be an n-dimensional oriented pseudo-Riemannian manifold with volume form ω. Define inner product of forms $\alpha, \beta \in \Omega^k(M)$ by

$$((\alpha,\beta)) = \int_M (\alpha,\beta)\omega = \int_M \alpha \wedge \star\beta$$

(assuming that α, β are compactly supported or decay fast enough for the integral to converge).

(1) Prove that then

$$((d\alpha,\beta)) = ((\alpha, d^*\beta)), \qquad \alpha \in \Omega^{k-1}(M), \beta \in \Omega^k(M)$$

where the operator $d^*\colon \Omega^k(M) \to \Omega^{k-1}(M)$ is given by

$$d^* = (-1)^k \star^{-1} d\star.$$

In particular, if n is even, and the metric is positive definite, then

$$d^* = -\star d \star$$

and for $M = \mathbb{R}^{1,3}$, we have $d^* = \star d \star$.

(2) Show that for a 1-form $A = A_\mu dx^\mu$, we have

$$d^*A = -\partial_\mu A^\mu.$$

where as usual, $A^\mu = \eta^{\mu\nu}A_\nu$ and we assume that the coordinate system x^μ is chosen so that the volume form is given by $\omega = dx^1 \wedge \dots dx^n$.

(3) Consider the operator

$$d^*d + dd^*\colon \Omega^k \to \Omega^k.$$

Show that for $k = 0$, this operator, up to a sign, consides with the Laplace–Beltrami operator:

$$(d^*d + dd^*)f = -\partial_\mu(\partial^\mu f) = \partial_\mu(\eta^{\mu\nu}\partial_\nu f)$$

where, as before, we assume that the volume form is given by $\omega = dx^1 \wedge \dots dx^n$.

In particular, for $\mathbb{R}^{1,3}$ coincides with negative of the d'Alembert box operator:

$$(d^*d + dd^*)f = -\Box f = -\partial_\mu \partial^\mu f.$$

(compare with (11.31)).

(4) Show that more generally, for a k-form $F = \frac{1}{k!}\sum F_I dx^I \in \Omega^k(\mathbb{R}^{1,3})$, we have

$$(dd^* + d^*d)F = -\frac{1}{k!}\sum(\Box F_I)dx^I.$$

Forms satisfying condition $(dd^* + d^*d)F = 0$ are called *harmonic*.

Exercise 15.8 (Charged particle in electromagnetic field)**.** Consider a relativistic particle of mass m and electric charge e in the external electromagnetic field with the potential $A = A_\mu dx^\mu$.

(1) Show that the Euler-Lagrange equations for the action functional

$$S(\gamma) = -mc\int_{P_0}^{P_1} ds - \frac{e}{c}\int_{P_0}^{P_1} A_\mu dx^\mu = \int_{t_0}^{t_1}\left(-mc^2\sqrt{1-\frac{v^2}{c^2}} + \frac{e}{c}\boldsymbol{A}\cdot\boldsymbol{v} - e\varphi\right)dt$$

have the form

$$\frac{d\boldsymbol{p}}{dt} = \boldsymbol{F},$$

where $\boldsymbol{p}$ is the relativistic momentum and $\boldsymbol{F}$ is the Lorentz force,

$$\boldsymbol{F} = e\left(\boldsymbol{E} + \frac{\boldsymbol{v}}{c}\times\boldsymbol{B}\right).$$

(2) Show that canonically conjugated momentum $\boldsymbol{P}$ and the energy $\mathscr{E}$ of the particle are given by $\boldsymbol{P} = \boldsymbol{p} + \frac{e}{c}\boldsymbol{A}$ and $\mathscr{E} = c\sqrt{m^2c^2 + \boldsymbol{p}^2} + e\varphi$.

(3) Show that the Euler-Lagrange equations can be written in the Hamiltonian form

$$\dot{\boldsymbol{P}} = -\frac{\partial\mathscr{H}}{\partial\boldsymbol{r}}, \quad \dot{\boldsymbol{r}} = \frac{\partial\mathscr{H}}{\partial\boldsymbol{P}},$$

where

$$\boldsymbol{E} = -\boldsymbol{\nabla}\varphi - \frac{\partial\boldsymbol{A}}{\partial t}, \quad \boldsymbol{B} = \boldsymbol{\nabla}\times\boldsymbol{A},$$

and the Hamiltonian function $\mathscr{H}$ is obtained from $\mathscr{E}$ by replacing $\boldsymbol{p}$ by $\boldsymbol{P} - \frac{e}{c}\boldsymbol{A}$.

Chapter 16

Gauge Fixing and Hamiltonian Formalism in Electromagnetism

16.1. Gauge fixing

As discussed in the previous chapter, Maxwell's equations in the empty space take the form

$$dF = 0 \quad \text{and} \quad d{\star}F = 0 \tag{16.1}$$

where $F = \frac{1}{2}F_{\mu\nu}dx^\mu \wedge dx^\nu$ is a 2-form on the Minkowski space $\mathbb{R}^{1,3}$. The general solution of equations (16.1) is $F = dA$, where $A = A_\mu dx^\mu$ is a one-form satisfying

$$\star d{\star}dA = 0. \tag{16.2}$$

For given F, the choice of such 1-form A is not unique: if A is a solution, then for any smooth function f on $\mathbb{R}^{1,3}$, we have $F = dA = d(A + df)$, so $A + df$ is also a solution (this is closely related to the fact that the equation (16.2) is not hyperbolic). Solutions A and $A + df$ are commonly called "gauge equivalent"; as we will disucss later in a more general setting, these solutions are obtained from each other by the action of a so-called "gauge group". Physically, these solutions are considered to be equvalent.

It is common in physics to impose additional constraints on the form A to eliminate or at least reduce this ambiguity. It is commonly referred to as "gauge fixing". There are several possible approaches; the most popular is using the *Lorenz*[1] *gauge* condition:

$$d{\star}A = 0, \tag{16.3}$$

or, equivalently, $d^*A = \partial^\mu A_\mu = 0$, where $d^* = \star d \star$ (see Exercise 15.7). We will show below that for any A that satisfies Maxwell's equations, there is a gauge equivalent form which also satisfies the Lorenz gauge condition.

[1]Named after Danish physicist and mathematician Ludvig Lorenz, not to be confused with the Dutch physicist Hedrick Lorentz!

Together with Maxwell's equations $d^*dA = 0$, Lorenz gauge condtion implies that A satisfies the equation

$$(d^*d + dd^*)A = 0$$

which, by results of Exercise 15.7, is equivalent to

$$\Box A_\mu = 0 \text{ for all } \mu \tag{16.4}$$

where

$$\Box = \partial_\mu \partial^\mu = \partial_0^2 - \partial_1^2 - \partial_2^2 - \partial_3^2$$

is the D'Alembert box operator. Note that equation (16.4) is hyperbolic.

Lorenz gauge condition does not fully fix the choice of A: if A is a solution of Maxwell's equations satisfying (16.3), then so does $A + df$ for any function f satisfying $\Box f = 0$. To further reduce ambiguity, we can impose an additional gauge condition, namely the *Hamilton gauge* condition $A_0 = 0$. Together with Lorenz gauge, this gives the *Coulomb gauge conditions*

$$\begin{aligned} d\star A &= 0, \\ A_0 &= 0. \end{aligned} \tag{16.5}$$

Coulomb gauge conditions can be rewritten in terms of the scalar potential $\varphi = cA_0$ and vector potential $\boldsymbol{A} = (A^1, A^2, A^3)$, where $A^i = -A_i$ (see (15.34)). Namely, they are equivalent to requiring that $\varphi = 0$ and the vector potential $\boldsymbol{A}$ satisfies

$$\boldsymbol{\nabla} \cdot \boldsymbol{A} = \partial_i A^i = 0.$$

Lemma 16.1. *Let $A \in \Omega^1(\mathbb{R}^{1,3})$ be a solution of Maxwell's equations in free space ($\rho = 0$, $\boldsymbol{j} = 0$) such that for each time slice $t = t_0$, $A(t_0, \boldsymbol{x})$ is of Schwartz class. Then one can find a unique gauge equivalent form $\tilde{A} = A + df$ which satisfies Coulomb gauge conditions (16.5) and is again of Schwartz class on each time slice.*

Proof. First, gauge transformation $A \mapsto A + df$, where f is chosen so that $\partial_0 f = -A_0$ allows us to replace A by a gauge equivalent form satisfying $A_0 = 0$. It is easy to check that Schwartz class condition is preserved.

Next, if $A_0 = 0$, then $\boldsymbol{E} = -\dfrac{\partial \boldsymbol{A}}{\partial t}$ (see (15.35)) which together with Maxwell's equation $\boldsymbol{\nabla} \cdot \boldsymbol{E} = 0$ implies that

$$\frac{\partial}{\partial t}\Big(\boldsymbol{\nabla} \cdot \boldsymbol{A}\Big) = 0.$$

In other words, $\boldsymbol{\nabla} \cdot \boldsymbol{A}$ is time-independent.

Let χ be a function on $\mathbb{R}^3$ which satisfies the equation

$$\Delta \chi = -\boldsymbol{\nabla} \cdot \boldsymbol{A}$$

where $\Delta = \boldsymbol{\nabla}^2 = \partial_1^2 + \partial_2^2 + \partial_3^2$ is the Laplace operator in $\mathbb{R}^3$. As is well-known, this equation has a solution, which can be written in terms of Green's function. Then it is easy to see that $\boldsymbol{\nabla} \cdot (\boldsymbol{A} + \boldsymbol{\nabla}\chi) = 0$, so the gauged transformed $\tilde{A} = A + d\chi$ satisfies the Lorenz gauge condition together with $\tilde{A}_0 = 0$.

Uniqueness follows from the fact that a solution of $\Delta\chi = -\boldsymbol{\nabla} \cdot \boldsymbol{A}$ is unique if we require it to be of Schwartz class: indeed, if χ_1, χ_2 are two solutions, then $\chi_1 - \chi_2$ would be a harmonic function of Schwartz class. But the ony such function is zero. □

Remark 16.2. Note that Lorenz gauge conditions are relativistic invariant, i.e. invariant under the action of Lorenz group $\mathfrak{L} = \mathrm{SO}(1,3)$. Condition $A_0 = 0$ is not relativistic invariant.

Remark 16.3. Maxwell equations with sources in the Lorenz gauge have the form

$$\Box A^\mu = J^\mu.$$

To summarize, in the Coulomb gauge Maxwell equations in free and empty space take the form

$$\nabla \cdot \boldsymbol{A} = 0, \qquad \Box \boldsymbol{A} = 0 \quad \text{and} \quad \boldsymbol{E} = -\frac{\partial \boldsymbol{A}}{\partial t}, \quad \boldsymbol{B} = \nabla \times \boldsymbol{A}, \tag{16.6}$$

which implies

$$\Box \boldsymbol{E} = 0 \quad \text{and} \quad \Box \boldsymbol{B} = 0. \tag{16.7}$$

16.2. Plane waves

Consider the simple case when vector potential $\boldsymbol{A}$ depends only on coordinates $x = x^1$ and t. In the Coulomb gauge Maxwell equations (16.6) reduce to

$$\frac{\partial^2 \boldsymbol{A}}{\partial t^2} - c^2 \frac{\partial^2 \boldsymbol{A}}{\partial x^2} = 0$$

and have a general solution

$$\boldsymbol{A}(t,x) = \boldsymbol{A}_+\left(t - \frac{x}{c}\right) + \boldsymbol{A}_-\left(t + \frac{x}{c}\right),$$

where $\boldsymbol{A}_+$, $\boldsymbol{A}_-$ are arbitrary vector-valued functions of single variable. They describe plane waves moving in positive and negative directions of the x-axis. We begin by analyzing the wave moving in a positive direction of the x-axis:

$$\boldsymbol{A} = \boldsymbol{A}\left(t - \frac{x}{c}\right).$$

The Coulomb gauge condition $\nabla \cdot \boldsymbol{A} = 0$ gives

$$\frac{\partial A^1}{\partial x} = 0,$$

so $A^1 = at$, where a is a constant. It is sufficient to consider a special case $A^1 = 0$, since according to (15.35) the general case is obtained by adding a constant electric field in the x-direction.

Introducing the direction of propagation of the wave — the unit vector $\boldsymbol{n} = \boldsymbol{e}_1$ — we have $\boldsymbol{A} \perp \boldsymbol{n}$. Since $\partial_1 \boldsymbol{A} = -\frac{1}{c}\partial_t \boldsymbol{A}$ and $\partial_2 \boldsymbol{A} = \partial_3 \boldsymbol{A} = 0$, we obtain

$$\nabla \times \boldsymbol{A} = -\frac{1}{c}\boldsymbol{n} \times \partial_t \boldsymbol{A}.$$

By (15.35), this gives the following formulas for the electric and magnetic fields

$$\boldsymbol{E} = -\partial_t \boldsymbol{A} \quad \text{and} \quad \boldsymbol{B} = -\frac{1}{c}\boldsymbol{n} \times \partial_t \boldsymbol{A} = \frac{1}{c}\boldsymbol{n} \times \boldsymbol{E}. \tag{16.8}$$

Thus we see that electric and magnetic fields $\boldsymbol{E}$ and $\boldsymbol{B}$ are perpendicular to the direction $\boldsymbol{n}$ of propagation of the waves. In physics terminology, electromagnetic plane waves are said to *transverse waves*. Moreover, the electric and magnetic fields are orthogonal with the strengths $E = cB$. Tho summarize, vectors $\boldsymbol{n}, \dfrac{\boldsymbol{E}}{E}, \dfrac{\boldsymbol{B}}{B}$ form an orthonormal positively oriented basis of $\mathbb{R}^3$.

The above discussion can easily be generalized to electromagnetic wave propagating in any direction: for a unit length vector $\boldsymbol{n}$, we consider a solution of Maxwell's equations of the form

$$\boldsymbol{A} = \boldsymbol{A}\left(t - \frac{\boldsymbol{r}\cdot\boldsymbol{n}}{c}\right),$$

and formulas (16.8) hold.

The components of the energy-momentum tensor of a plane wave are easily found to be (see Section 15.5)

$$T^{00} = \frac{E^2}{c^2} \quad \text{and} \quad \boldsymbol{S} = \frac{1}{c^2}\boldsymbol{E}\times\boldsymbol{n}\times\boldsymbol{E} = \frac{E^2}{c^2}\boldsymbol{n},$$

so that $(T^{00})^2 = \boldsymbol{S}^2$.

An important special case of electromagnetic wave is a *monochromatic* plane wave:

$$\boldsymbol{A} = \operatorname{Re}\left\{\boldsymbol{A}_0 e^{-i\omega(t-\frac{\boldsymbol{r}\cdot\boldsymbol{n}}{c})}\right\},$$

where $\boldsymbol{n}$ is a unit vector in the direction of propagation of the wave, and $\boldsymbol{A}_0 \in \mathbb{C}^3$ is a constant complex vector such that $\boldsymbol{A}_0\cdot\boldsymbol{n} = 0$. The number ω is called the *frequency*.

Introducing the *wave vector* $\boldsymbol{k} = \dfrac{\omega}{c}\boldsymbol{n}$, we can rewrite the previous formula as

$$\boldsymbol{A} = \operatorname{Re}\left\{\boldsymbol{A}_0 e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)}\right\}.$$

Using (16.8), we get

$$\boldsymbol{E} = \operatorname{Re}\left\{\boldsymbol{E}_0 e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)}\right\} \quad \text{and} \quad \boldsymbol{B} = \operatorname{Re}\left\{\boldsymbol{B}_0 e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)}\right\},$$

where

$$\boldsymbol{E}_0 = i\omega\boldsymbol{A}_0 \quad \text{and} \quad \boldsymbol{B}_0 = i\boldsymbol{k}\times\boldsymbol{A}_0.$$

It is easy to show that one can choose coordinate axes y and z, perpendicular to $\boldsymbol{n}$ and to each other, such that in the new coordinates we have

$$E_y = b_1\cos(\omega t - \boldsymbol{k}\cdot\boldsymbol{r} - \alpha) \quad \text{and} \quad E_z = \pm b_2\sin(\omega t - \boldsymbol{k}\cdot\boldsymbol{r} - \alpha), \tag{16.9}$$

for some non-negative b_1, b_2 and a phase α. 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 in space the electric field vector $\boldsymbol{E}$ rotates in the plane perpendicular to the direction of propagation filling 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 16.4. Introduce the 4-vector $(k^\mu) = \left(\dfrac{\omega}{c}, \boldsymbol{k}\right)$ and $(k_\mu) = \left(\dfrac{\omega}{c}, -\boldsymbol{k}\right)$ 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*, massless particles with 4-wave vector satisfying $k_0^2 = \boldsymbol{k}^2$.

16.3. The general solution

As in Section 13.5, we show that the electromagnetic field can be treated as a (continuous) collection of non-interacting harmonic oscillators. For this aim, we obtain a general formula for solution of Maxwell equations in Coulomb gauge.

We assume that we are given initial conditions, i.e. the values of vector potential $\boldsymbol{A}(t,\boldsymbol{r})$ and its time derivative $\partial_t \boldsymbol{A}(t,\boldsymbol{r})$ at $t=0$. Thus, we need to solve the following Cauchy problem:

$$\begin{aligned} \Box \boldsymbol{A} &= 0, \\ \boldsymbol{A}(0,\boldsymbol{r}) &= \boldsymbol{A}_0(\boldsymbol{r}), \\ \frac{\partial \boldsymbol{A}}{\partial t}(0,\boldsymbol{r}) &= \boldsymbol{A}_1(\boldsymbol{r}), \end{aligned} \tag{16.10}$$

where $\Box = \partial_\mu \partial^\mu$. Here 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 $|\boldsymbol{r}| \to \infty$.

As in Section 13.5, Cauchy problem for the wave equation in $\mathbb{R}^4$ is solved by the Fourier transform. Namely, let

$$\begin{aligned} \boldsymbol{A}_0(\boldsymbol{r}) &= \frac{1}{(2\pi)^{\frac{3}{2}}} \int_{\mathbb{R}^3} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \boldsymbol{a}_0(\boldsymbol{k}) d^3\boldsymbol{k} \\ \boldsymbol{A}_1(\boldsymbol{r}) &= \frac{1}{(2\pi)^{\frac{3}{2}}} \int_{\mathbb{R}^3} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \boldsymbol{a}_1(\boldsymbol{k}) d^3\boldsymbol{k} \end{aligned} \tag{16.11}$$

Since $\boldsymbol{A}_0(\boldsymbol{r}), \boldsymbol{A}_1(\boldsymbol{r})$ are real-valued, we have $\boldsymbol{a}_0(\boldsymbol{k}) = \bar{\boldsymbol{a}}_0(-\boldsymbol{k})$, $\boldsymbol{a}_1(\boldsymbol{k}) = \bar{\boldsymbol{a}}_1(-\boldsymbol{k})$; similarly, conditions $\boldsymbol{\nabla} \cdot \boldsymbol{A}_0 = 0$, $\boldsymbol{\nabla} \cdot \boldsymbol{A}_1 = 0$ are equivalent to $\boldsymbol{k} \cdot \boldsymbol{a}_0(\boldsymbol{k}) = \boldsymbol{k} \cdot \boldsymbol{a}_1(\boldsymbol{k}) = 0$.

Theorem 16.5. *General solution of Cauchy problem* (16.10) *is*

$$\boldsymbol{A}(t,\boldsymbol{r}) = \frac{1}{(2\pi)^3} \int_{\mathbb{R}^3} \left(e^{-i(\omega_{\boldsymbol{k}} t - \boldsymbol{k}\cdot\boldsymbol{r})} \boldsymbol{a}(\boldsymbol{k}) + e^{i(\omega_{\boldsymbol{k}} t - \boldsymbol{k}\cdot\boldsymbol{r})} \bar{\boldsymbol{a}}(\boldsymbol{k}) \right) d^3\boldsymbol{k}, \tag{16.12}$$

where $\omega_{\boldsymbol{k}} = c|\boldsymbol{k}|$ *and*

$$\boldsymbol{a}(\boldsymbol{k}) = \frac{1}{2}\boldsymbol{a}_0(\boldsymbol{k}) + \frac{1}{2ic|\boldsymbol{k}|} a_1(\boldsymbol{k}),$$

with $\boldsymbol{a}_0$, $\boldsymbol{a}_1$ *defined by* (16.11).

Proof. Each component of $\boldsymbol{A}$ satisfies Klein-Gordon equation with zero mass, so general solution — formula (16.12) — is obtained from the corresponding formulas (13.20)–(13.21) by setting $m=0$ and remembering that $t^0 = ct$. □

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,

$$\begin{aligned}\mathscr{E} &= \frac{1}{8\pi}\int_{\mathbb{R}^3}\left(\frac{1}{c^2}\boldsymbol{E}^2+\boldsymbol{B}^2\right)d^3\boldsymbol{r} \\ &= \frac{1}{4\pi}\int_{\mathbb{R}^3}\left(\frac{1}{c^2}\omega_{\boldsymbol{k}}^2\,\boldsymbol{a}(\boldsymbol{k})\bar{\boldsymbol{a}}(\boldsymbol{k})+(\boldsymbol{k}\times\boldsymbol{a}(\boldsymbol{k}))\cdot(\boldsymbol{k}\times\bar{\boldsymbol{a}}(\boldsymbol{k}))\right)d^3\boldsymbol{k} \\ &= \frac{1}{2\pi c^2}\int_{\mathbb{R}^3}\omega_{\boldsymbol{k}}^2\,\boldsymbol{a}(\boldsymbol{k})\cdot\bar{\boldsymbol{a}}(\boldsymbol{k})d^3\boldsymbol{k},\end{aligned}$$

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

$$\boldsymbol{k}\cdot\boldsymbol{P}(\boldsymbol{k}) = \boldsymbol{k}\cdot\boldsymbol{Q}(\boldsymbol{k}) = 0.$$

16.4. Hamiltonian formalism in classical electrodynamics

In this section, we describe Maxwell's equations in Hamiltonian formalism. For simplicity, we work in the system of units where $c = 1$. In this case, the Lagrangian function of classical electromagnetic theory in the absence of sources is given by the following formula (see Section 15.4):

$$\mathscr{L}(A) = -\frac{1}{16\pi} F_{\mu\nu}F^{\mu\nu} = \frac{1}{8\pi}(\boldsymbol{E}^2 - \boldsymbol{B}^2). \tag{16.13}$$

As in Section 15.4, we consider $\mathscr{L}$ as a function of the 4-potential $A = (A_0, A_1, A_2, A_3) = (A_0, -\boldsymbol{A})$ and its time derivative $\partial_0 A$, expressing $\boldsymbol{E}$, $\boldsymbol{B}$ in terms of A as follows:

$$\boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{A}$$
$$\boldsymbol{E} = -\boldsymbol{\nabla} A_0 - \partial_0 \boldsymbol{A}$$

or, equivalently

$$B_j = -\epsilon_{jkl}\partial_k A_l,$$
$$E_i = \partial_0 A_i - \partial_i A_0.$$

However, we can not just repeat the same steps we did for a free field theory to get a Hamiltonian description of Maxwell's theory of electromagnetism, since the Lagrangian (16.13) is singular, so the Legendre transform is not an isomorphism. This is immediate from observing that $\mathscr{L}(A)$ does not depend on $\partial_0 A_0$; a deeper reason for this degeneracy lies in the fact that $\boldsymbol{E}, \boldsymbol{B}$ depend only on the gauge equivalency class of A — we will discuss this in more detail later.

To get the Hamiltonian description, we will follow the method described in Chapter 6, rewriting (16.13) as a first order Lagrangian. To do this, we observe the trivial identity $\boldsymbol{E}^2 = 2(\partial_0 A_i - \partial_i A_0)E_i - \boldsymbol{E}^2$, which allows us to rewrite the Lagrangian as

$$\begin{aligned} \mathscr{L}(A) &= \frac{1}{4\pi}\left((\partial_0 A_i - \partial_i A_0)E_i - \frac{1}{2}(\boldsymbol{E}^2 + \boldsymbol{B}^2)\right) \\ &= \frac{1}{4\pi}\left(-\boldsymbol{E}\cdot\partial_0\boldsymbol{A} - \boldsymbol{E}\cdot\boldsymbol{\nabla}A_0 - \frac{1}{2}(\boldsymbol{E}^2 + \boldsymbol{B}^2)\right). \end{aligned} \tag{16.14}$$

Since $\boldsymbol{E}\cdot\boldsymbol{\nabla}A_0 = -(\boldsymbol{\nabla}\cdot\boldsymbol{E})A_0 + \boldsymbol{\nabla}(A_0\boldsymbol{E})$ and by Stokes' theorem, an integral of total divergence is zero, we obtain the following formula for the electromagnetic field Lagrangian

$$L = \frac{1}{4\pi}\int_{\mathbb{R}^3}\Big(-\boldsymbol{E}(\boldsymbol{x})\cdot\partial_0\boldsymbol{A}(\boldsymbol{x}) - \tfrac{1}{2}(\boldsymbol{E}^2(\boldsymbol{x}) + \boldsymbol{B}^2(\boldsymbol{x})) + A_0(\boldsymbol{x})\,\boldsymbol{\nabla}\cdot\boldsymbol{E}(\boldsymbol{x})\Big)\,d^3\boldsymbol{x}. \tag{16.15}$$

We will consider $A_0, \boldsymbol{A}, \boldsymbol{E}$ as independent variables, and $\boldsymbol{B} = \boldsymbol{\nabla}\times\boldsymbol{A}$ as a function of $\boldsymbol{A}$.

Let us compare (16.15) with the first order Lagrangian studied in (6.13):

$$\mathcal{L} = \boldsymbol{p}\dot{\boldsymbol{q}} - H(\boldsymbol{p},\boldsymbol{q}) - \sum_{a=1}^{m}\lambda_a\varphi^a(\boldsymbol{p},\boldsymbol{q}).$$

One can see that (16.15) is indeed an infinite-dimensional analog of (6.13) as follows:

- $\boldsymbol{A}(\boldsymbol{x})$ is an analog of $\boldsymbol{q}$, $\boldsymbol{E}(\boldsymbol{x})$ is an analog of $\boldsymbol{p}$, and the term $\dfrac{1}{4\pi}\int E_i(\boldsymbol{x})\partial_0 A_i(\boldsymbol{x})$ plays the role of $\boldsymbol{p}\dot{\boldsymbol{q}}$ term. Thus, the conjugate momenta to $A_i(\boldsymbol{x})$ is $\dfrac{1}{4\pi}E_i(\boldsymbol{x})$.
- The energy $H(\boldsymbol{p},\boldsymbol{q})$ is the total energy of the electromagnetic field:

$$H = \frac{1}{8\pi}\int (\boldsymbol{E}^2(\boldsymbol{x}) + \boldsymbol{B}^2(\boldsymbol{x}))d^3\boldsymbol{x}, \quad \boldsymbol{B} = \boldsymbol{\nabla}\times\boldsymbol{A}. \tag{16.16}$$

- The term $\dfrac{1}{4\pi}\int A_0(\boldsymbol{x})\,\boldsymbol{\nabla}\cdot\boldsymbol{E}(\boldsymbol{x})$ is an analog of the term $\lambda_a\varphi^a(\boldsymbol{p},\boldsymbol{q})$; thus, $A_0(\boldsymbol{x})$ are Lagrange multipliers and $\boldsymbol{\nabla}\cdot\boldsymbol{E}(\boldsymbol{x})$ are the constraints. For future convenience, we introduce the notation

$$C(\boldsymbol{x}) = \boldsymbol{\nabla}\cdot\boldsymbol{E}(\boldsymbol{x}).$$

- The phase space $\mathcal{M} = \{(\boldsymbol{E},\boldsymbol{A})\}$ is the following infinite-dimensional real vector space[2]

$$\mathcal{M} = \mathscr{S}(\mathbb{R}^3,\mathbb{R}^3)\times\mathscr{S}(\mathbb{R}^3,\mathbb{R}^3)$$

with the symplectic form Ω

$$\Omega = \frac{1}{4\pi}\int_{\mathbb{R}^3}(dE_i(\boldsymbol{x})\wedge dA_i(\boldsymbol{x}))\,d^3\boldsymbol{x}, \tag{16.17}$$

so that the pairs $(E_i(\boldsymbol{x}),A_i(\boldsymbol{x}))$ are Darboux coordinates on $\mathcal{M}$ with the canonical Poisson brackets

$$\{E_i(\boldsymbol{x}),A_j(\boldsymbol{y})\} = 4\pi\delta_{ij}\delta(\boldsymbol{x}-\boldsymbol{y}), \quad i,j=1,2,3. \tag{16.18}$$

According to general formalism in Chapter 6, the Maxwell's equations — the Euler-Lagrange equations for the Lagrangian $\mathscr{L}(A)$ — can be written as Euler–Lagrange equations (6.14)–(6.16), which in our case become

$$\begin{aligned} \frac{d}{dt}E_i(\boldsymbol{x}) &= \{H-C,E_i(\boldsymbol{x})\}, \quad C = \frac{1}{4\pi}\int_{\mathbb{R}^3}A_0(\boldsymbol{x})C(\boldsymbol{x})d^3\boldsymbol{x},\\ \frac{d}{dt}A_i(\boldsymbol{x}) &= \{H-C,A_i(\boldsymbol{x})\},\\ C(\boldsymbol{x}) &= 0. \end{aligned} \tag{16.19}$$

It is very instructive to verify it directly. Namely, using canonical Poisson brackets (16.18), it is easy to check that we have the following relations (see Exercise 16.2):

$$\begin{aligned} \{H,A_i(\boldsymbol{x})\} &= E_i(\boldsymbol{x}),\\ \{H,E_i(\boldsymbol{x})\} &= \epsilon_{ijk}\partial_k B_j(\boldsymbol{x}), \qquad B_j = (\boldsymbol{\nabla}\times\boldsymbol{A})_j = -\epsilon_{jkl}\partial_k A_l,\\ \{C(\boldsymbol{x}),E_i(\boldsymbol{y})\} &= 0,\\ \{C(\boldsymbol{x}),A_i(\boldsymbol{y})\} &= 4\pi\frac{\partial}{\partial x^i}\delta(\boldsymbol{x}-\boldsymbol{y}). \end{aligned} \tag{16.20}$$

This readily implies that for C defined in (16.19) we have

$$\{C,A_i(\boldsymbol{x})\} = -\partial_i A_0(\boldsymbol{x}), \tag{16.21}$$

$$\{C,E_i(\boldsymbol{x})\} = 0. \tag{16.22}$$

[2] Here $\mathscr{S}(\mathbb{R}^3,\mathbb{R}^3)$ stands for the space of $\mathbb{R}^3$-valued Schwartz functions on $\mathbb{R}^3$.

Thus, we can rewrite Euler–Lagrange equations (16.19) as

$$\begin{aligned} \frac{d}{dt}E_i(\boldsymbol{x}) &= \epsilon_{ijk}\partial_k B_j(\boldsymbol{x}), \\ \frac{d}{dt}A_i(\boldsymbol{x}) &= E_i(\boldsymbol{x}) - \partial_i A_0(\boldsymbol{x}), \\ C(\boldsymbol{x}) &= 0. \end{aligned} \tag{16.23}$$

or, equivalently,

$$\begin{aligned} &\partial_t \boldsymbol{E} = \boldsymbol{\nabla}\times\boldsymbol{B}, \\ &\boldsymbol{E} = -\boldsymbol{\nabla}A_0 - \partial_t\boldsymbol{A}, \\ &\boldsymbol{\nabla}\cdot\boldsymbol{E} = 0. \end{aligned} \tag{16.24}$$

Since $\boldsymbol{B} = \boldsymbol{\nabla}\times\boldsymbol{A}$, we see that these equations coincide with the Maxwell's equations (15.2)–(15.5) in the absence of sources.

Thus we have shown that the Lagrangian of electromagnetic theory can be written as a first order Lagrangian in the form (6.13), and that the Euler–Lagrange equations for this Lagrangian are exactly Maxwell's equations.

Since the Lagrangian is singular, we can not use the Legendre transform to rewrite this as a Hamiltonian system on $\mathcal{M}$. Instead, we will follow the approach of Section 6.3 and show that this system is equivalent to a Hamiltonian system on a reduced phase space $\mathcal{M}_0^*$.

Lemma 16.6. *The constraints $C(\boldsymbol{x}) = \boldsymbol{\nabla}\cdot\boldsymbol{E}(\boldsymbol{x}) = 0$ for $\boldsymbol{x}\in\mathbb{R}^3$, are a first class constraints as defined in Definition 6.4.*

The proof is immediate since

$$\begin{aligned} \{C(\boldsymbol{x}), C(\boldsymbol{y})\} &= 0, \\ \{H, C(\boldsymbol{x})\} &= 0, \end{aligned} \tag{16.25}$$

which is easy to derive from (16.20).

As in Section 6.3, the construction of the reduced phase space is done in two steps. First, we define

$$\mathcal{M}_0 = \{(\boldsymbol{E},\boldsymbol{A})\in\mathcal{M} \mid \boldsymbol{\nabla}\cdot\boldsymbol{E} = 0\}.$$

Next, we need to take the quotient by the infinitesimal action of the vector fields on $\mathcal{M}_0$ generated by the constraints $C(\boldsymbol{x})$. To do that, define

$$C_\varphi = \frac{1}{4\pi}\int \varphi(\boldsymbol{x})C(\boldsymbol{x})\,d^3\boldsymbol{x}$$

where φ is a Schwartz class function on $\mathbb{R}^3$. Then (16.20) implies

$$\{C_\varphi, E_i(\boldsymbol{x})\} = 0, \qquad \{C_\varphi, A_i(\boldsymbol{x})\} = -\partial_i\varphi(\boldsymbol{x}).$$

Thus, the Hamiltonian vector field on $\mathcal{M}$ generated by C_φ is

$$\delta\boldsymbol{E}(\boldsymbol{x}) = 0, \qquad \delta\boldsymbol{A}(\boldsymbol{x}) = \boldsymbol{\nabla}\varphi(\boldsymbol{x}),$$

and the flow of this vector field is the transformation

$$(\boldsymbol{E},\boldsymbol{A}) \mapsto (\boldsymbol{E}, \boldsymbol{A} + s\boldsymbol{\nabla}\varphi). \tag{16.26}$$

These are exactly the gauge transformation of A, discussed in the previous chapter.

Theorem 16.7. *Define the reduced phase space* $\mathcal{M}_0^*$ *as the set of gauge equivalence classes* $(\boldsymbol{E},\boldsymbol{A})\in\mathcal{M}_0$ *under the gauge transformations* (16.26).

(1) *The symplectic form* Ω *descends to* $\mathcal{M}_0^*$ *and defines on it a structure of an infinite-dimensional symplectic manifold. Similarly, the Hamiltonian function* H *given by* (16.16) *aslo descends to* $\mathcal{M}_0^*$.

(2) *The Euler–Lagrange equations* (16.23) *on* $\mathcal{M}_0^*$ *coincide with equations of motion of the Hamiltonian system* $(\mathcal{M}_0^*,\Omega,H)$.

Note that in particular, this shows that the solutions of equations of motion on $\mathcal{M}_0^*$ do not depend on the choice of $A_0(\boldsymbol{x})$. Thus, if we are only interested in solutions modulo the gauge equivalence, we can choose $A_0(\boldsymbol{x})=0$.

This theorem shows that choosing gauge condition such as Coulomb gauge can be viewed as choosing a slice in $\mathcal{M}_0$ transversal to the foliation $\mathcal{P}$ defined by the vector fields generated by constraints, as in Section 6.3. For example, one can use the Lorenz gauge:

$$D(\boldsymbol{x})=-\boldsymbol{\nabla}\cdot\boldsymbol{A}(\boldsymbol{x})=0,$$

which, together with $A_0(\boldsymbol{x})=0$, gives Coulomb gauge conditions (16.5). By Lemma 16.1, this allows one to choose a unique representative in each gauge equivalence class, which strongly suggests that this slice is transversal to the foliation.

One can show transversality by showing that the integral operator with the kernel $\{C(\boldsymbol{x}),D(\boldsymbol{y})\}$ is non-degenerate in $L^2(\mathbb{R}^3)$ (this is the infinite-dimensional analog of condition $\det\left(\{\chi_a,\varphi^b\}\right)_{a,b=1}^m\neq 0$, see (6.27)). Indeed, it follows from (16.18) 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 $\mathcal{M}_0^*$ of classical electrodynamics can also be defined as a linear subspace in $\mathcal{M}$ defined by

$$\mathcal{M}_0^*=\left\{(\boldsymbol{E}(\boldsymbol{x}),\boldsymbol{A}(\boldsymbol{x}))\in\mathcal{M}\mid C(\boldsymbol{x})=D(\boldsymbol{x})=0,\quad \boldsymbol{x}\in\mathbb{R}^3\right\}.$$

Since

$$\{D(\boldsymbol{x}),D(\boldsymbol{y})\}=0,$$

Darboux coordinates for the symplectic form $\Omega_0=\Omega|_{\mathcal{M}_0^*}$ can be found by the general procedure described in Section 6.3.

Using Exercises 6.4 and 6.5 in Chapter 6, the Poisson bracket $\{\ ,\ \}_0$ on $\mathcal{M}_0^*$, associated with the symplectic form Ω_0, can be written as a restriction of the Dirac bracket on $\mathcal{M}$, associated with the second class constraints $C(\boldsymbol{x}),D(\boldsymbol{x})$ for all $\boldsymbol{x}\in\mathbb{R}^3$. Namely, we have

$$\begin{aligned}\{F,G\}_{\mathrm{DB}}=\{F,G\}+\int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\Big(&\{F,C(\boldsymbol{x})\}G(\boldsymbol{y}-\boldsymbol{x})\{D(\boldsymbol{y}),G\}-\\ &-\{F,D(\boldsymbol{x})\}G(\boldsymbol{x}-\boldsymbol{y})\{C(\boldsymbol{y}),G\}\Big)d^3\boldsymbol{x}d^3\boldsymbol{y},\end{aligned}\tag{16.27}$$

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 (16.18) we readily compute that

$$\{E_i(\boldsymbol{x}), E_j(\boldsymbol{y})\}_{\mathrm{DB}} = \{A_i(\boldsymbol{x}), A_j(\boldsymbol{y})\}_{\mathrm{DB}} = 0 \tag{16.28}$$

and

$$\{E_i(\boldsymbol{x}), A_j(\boldsymbol{y})\}_{\mathrm{DB}} = 4\pi\delta^{\perp}_{ij}(\boldsymbol{x}-\boldsymbol{y}), \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^3, \tag{16.29}$$

where the distribution $\delta^{\perp}_{ij}(\boldsymbol{x})$ is the *transverse δ-function*,

$$\delta^{\perp}_{ij}(\boldsymbol{x}) = \frac{1}{(2\pi)^3}\int_{\mathbb{R}^3}\left(\delta_{ij} - \frac{k_i k_j}{\boldsymbol{k}^2}\right) e^{i\boldsymbol{k}\boldsymbol{x}} d^3\boldsymbol{k}, \quad i, j = 1, 2, 3. \tag{16.30}$$

It satisfies

$$\partial_i \delta^{\perp}_{ij}(\boldsymbol{x}) = 0, \quad j = 1, 2, 3.$$

Thus Dirac bracket (16.27) yields a 'transverse' Poisson structure $\{\ ,\ \}$ on $\mathcal{M}$, determined by (16.28)–(16.29). 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 $\mathcal{M}^*_0$ 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 $\mathcal{M}^*_0$

$$\dot{\boldsymbol{E}}(\boldsymbol{x}) = \{H, \boldsymbol{E}(\boldsymbol{x})\}_0,$$
$$\dot{\boldsymbol{A}}(\boldsymbol{x}) = \{H, \boldsymbol{A}(\boldsymbol{x})\}_0,$$

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.

16.5. Coupling electromagnetiic field with matter

Recall that the Lagrangian density of the electromagnetic field in the presence of sources is given by (15.40),

$$\mathscr{L}(A) = -\frac{1}{16\pi}F_{\mu\nu}F^{\mu\nu} - \frac{\mu_0}{4\pi}A_\mu J^\mu \quad \text{and} \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu.$$

The equations of motion read

$$\partial_\mu F^{\mu\nu} = \mu_0 J^\mu,$$

where J^μ is a conserved current, $\partial_\mu J^\mu = 0$.

The current J^μ can be realized as a conservation law for the spinor field with the Dirac Lagrangian density

$$\mathscr{L}(\bar{\psi}, \psi) = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi,$$

considered in Example 11.6. Namely, we have in Example 12.5 that the current

$$J^\mu = \bar{\psi}\gamma^\mu\psi$$

is a Noether conservation law. Thus it is natural to couple the electromagnetic field with the spinor field and consider Lagrangian density

$$\mathscr{L} = \frac{1}{4\pi}\left(-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - e\bar{\psi}\gamma^\mu A_\mu\psi\right), \tag{16.31}$$

where instead of μ_0 we introduced a constant e. Corresponding equations of motion are

$$\partial_\mu F^{\mu\nu} = e\bar{\psi}\gamma^\mu\psi \quad \text{and} \quad i\gamma^\mu(\partial_\mu + ieA_\mu)\psi - m\psi = 0. \tag{16.32}$$

As we have seen in Section 15.4, Maxwell's equations are invariant under the gauge transformations $A_\mu \mapsto A_\mu + \partial_\mu f$, while Dirac equation is only U(1)-invariant. However, coupling with the electromagnetic field promotes global U(1)-invariance to gauge invariance with the structure group U(1). This means that Lagrangian (16.31) is invariant under the transformations

$$A \mapsto A + df \quad \text{and} \quad \psi(x) \mapsto e^{-ief(x)}\psi(x), \;\; \bar{\psi}(x) \mapsto e^{ief(x)}\bar{\psi}(x)$$

for arbitrary smooth real-valued function $f(x)$ on $\mathbb{R}^4$.

As we will discuss it Section 17.2 nad 18.2. $\partial_\mu + iA_\mu$ are covariant directional derivatives for a connection ∇ in the principal U(1)-bundle (or in the line bundle $\mathfrak{u}(1)_P$) over $\mathbb{R}^4$ with the one-form ieA. As it customary in physics textbooks, here we use $\mathfrak{u}(1) = i\mathbb{R}$ and not identify it with $\mathbb{R}$, as in Section 18.2 (see Remark 17.3). Introducing physicist's notation

$$\not{D} = \gamma^\mu D_\mu = \gamma^\mu(\partial_\mu + ieA_\mu),$$

we can rewrite the Lagrangian (16.31) as

$$\mathscr{L} = \frac{1}{4\pi}\left(-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(i\not{D} - m)\psi\right). \tag{16.33}$$

Remark 16.8. It is remarkable that (16.33) is the Lagrangian of quantum electrodynamics (QED), the relativistic quantum field theory describing the interaction of light and matter. Specifically, photons, a quanta of the electromagnetic field, interact with the electrons and their anti-particles — positrons — a quanta of the spinor fields ψ and $\bar{\psi}$. Photons are massless elementary particles of spin 1 (bosons) and carry no electric charge, while electrons and positrons are massive elementary particles of spin $\frac{1}{2}$ (fermions) and carry electric charges e and $-e$. The total electric charge

$$Q = e\int_{\mathbb{R}^3} \bar{\psi}(\boldsymbol{x})\gamma^0\psi(\boldsymbol{x})d^3\boldsymbol{x}$$

is conserved, and in the quantum theory is an integer multiple of e (i.e. is an operator with the eigenvalues e times the number of particles minus the number of anti-particles).

In a similar way we can couple the electromagnetic field with the complex scalar field, considered in Example 11.4. We know from Example 12.3 that corresponding conserved current is

$$J^\mu = i(\varphi\partial^\mu\bar{\varphi} - \bar{\varphi}\partial^\mu\varphi),$$

and we obtain a U(1)-gauge invariant Lagrangian by promoting partial derivatives to covariant derivatives,

$$\mathscr{L} = \frac{1}{4\pi}\left(-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + D_\mu\varphi D^\mu\bar{\varphi} - m^2|\varphi|^2\right). \tag{16.34}$$

This procedure is known as *minimal coupling*. Correspondingly, the current J^μ is being promoted to a current

$$\hat{J}^\mu = i(\varphi D^\mu\bar{\varphi} - \bar{\varphi}D^\mu\varphi),$$

where partial derivatives are replaced by the covariant derivatives, and the Euler-Lagrange equations for the electromagnetic field are Maxwell's equations with the current $\hat{J}^\mu$.

16.6. Exercises

Exercise 16.1. Derive formulas (16.9).

Exercise 16.2. Prove formulas (16.20) for Poisson brackets on $\mathcal{M}$.

Exercise 16.3. Similarly to what was done in Section 13.4, show that Poincaré group acts on the reduced phase space $\mathcal{M}_0^*$ in a Hamiltonian way and find the corresponding Hamiltonian functionals.

Exercise 16.4. Express the Hamiltonian functionals from the previous exercise in terms of normal modes $a(\boldsymbol{k})$ and $\bar{a}(\boldsymbol{k})$.

Exercise 16.5. Derive equations of motion for the Lagrangian (16.34).

Chapter 17

Connections and Curvature

In the previous chapter, we discussed Maxwell's theory of electromagnetism. It turns out that it is a very special case of a large class of theories, known as Yang–Mills theories. Such a theory depends on a choice of a compact Lie group, called the structure Lie group of the theory; for electromagnetism, the structure group is the group U(1). To formulate the Yang–Mills theory for arbitrary compact Lie group G and explain its relation with Maxwell's theory of electromagnetism, one needs to use differential geometry of principal and vector bundles. It is succinctly summarized below; readers familiar with this material can skip this chapter, returning to it as needed.

17.1. Vector bundles and G-bundles

We begin by reminding the key facts about vector bundles and G-bundles.

Let $\pi\colon E \to M$ be a complex vector bundle of rank r over a manifold M (later, we will take M to be the spacetime of our theory). We denote by E_x the fiber of E at point $x \in M$ and by $\mathcal{E}$ the sheaf of smooth sections of E; thus, for an open $U \subset M$, $\mathcal{E}(U) = \Gamma(\mathcal{E}, U)$ is the vector space of all sections of E over U.

Choose an open cover $M = \bigcup_{\alpha\in A} U_\alpha$ such that the restriction of E to each U_α can be trivialized: we have an isomorphism of vector bundles

$$\varphi_\alpha\colon E|_{U_\alpha} \simeq U_\alpha \times \mathbb{C}^r. \tag{17.1}$$

Then on each $U_{\alpha\beta} = U_\alpha \cap U_\beta$, we have two trivializations $\varphi_\alpha, \varphi_\beta$ which are related by

$$\varphi_\alpha = g_{\alpha\beta}\varphi_\beta \text{ on } U_\alpha \cap U_\beta \tag{17.2}$$

for some functions $g_{\alpha\beta}$ on $U_\alpha \cap U_\beta$ with values in the group $\mathrm{GL}(r, \mathbb{C})$. These functions are called the *transition functions*; they satisfy the cocycle conditions:

$$\begin{aligned} &g_{\alpha\beta}g_{\beta\alpha} = 1 \text{ on } U_\alpha \cap U_\beta,\\ &g_{\alpha\beta}g_{\beta\gamma} = g_{\alpha\gamma} \text{ on } U_\alpha \cap U_\beta \cap U_\gamma \end{aligned} \tag{17.3}$$

and thus define a class in the Čech cohomology group $\check{H}^1(M, \mathrm{GL}(r,\mathbb{C}))$.

In terms of local trivializations, sections of E can be described as collections of functions $s_\alpha\colon U_\alpha \to \mathbb{C}^r$ such that

$$s_\alpha = g_{\alpha\beta} s_\beta \text{ on } U_\alpha \cap U_\beta. \tag{17.4}$$

So far, all our transition functions took values in the group $\mathrm{GL}(r,\mathbb{C})$. However, it frequently happens that one can choose trivializations for which all transition functions take values in some closed subgroup $G \subset \mathrm{GL}(r,\mathbb{C})$. In such a situation, we say that vector bundle E has the structure group G. For example, if E carries a Hermitian structure (i.e. a positive definite Hermitian form), then one can choose local trivializations which identify the Hermitian form with the standard inner product in $\mathbb{C}^r$; thus, in this case all transition functions will take values in the group $G = \mathrm{U}(r)$.

The real vector bundles of rank r over M are introduced in the same way by replacing the vector space $\mathbb{C}^r$ by $\mathbb{R}^r$.

In fact, there is a way to define a G-bundle for arbitrary Lie group G, whether or not it is a subgroup in $\mathrm{GL}(r,\mathbb{C})$. This can be done using the language of principal G-bundles.

Definition 17.1. Given a Lie group G and a smooth manifold M, a *principal G-bundle* over M is a fiber bundle $\pi\colon P \to M$ with a smooth *right* G-action: every $g \in G$ defines a bundle map $P \to P\colon p \mapsto pg$ such that the action of G on each fiber P_x 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, i.e. a diffeomorphism

$$\varphi_\alpha\colon P|_{U_\alpha} \to U_\alpha \times G, \tag{17.5}$$

which commutes with the right action of G: if $\varphi(p) = (x,h)$, then $\varphi(pg) = (x,hg)$. Here we use notation $P|_U = \pi^{-1}(U)$.

As before, this implies that on the intersection $U_{\alpha\beta} = U_\alpha \cap U_\beta$ we have two different trivialization $\varphi_\alpha, \varphi_\beta$. Thus, we can consider composition $\varphi_\alpha \circ \varphi_\beta^{-1}\colon U_{\alpha\beta} \times G \to U_{\alpha\beta} \times G$. Since every diffeomorphism $G \to G$ which commutes with the right action of G on itself must be given by *left* multiplication by some element of G, we see that

$$\varphi_\alpha \circ \varphi_\beta^{-1}(x,g) = (x, g_{\alpha\beta}(x) g) \tag{17.6}$$

for some collection of transition functions $g_{\alpha\beta}\colon U_{\alpha\beta} \to G$. These transition functions are analogs of transition functions for vector bundles defined in (17.2); they satisfy the same cocycle conditions (17.3). Conversely, it is easy to show that every collection of transition functions $g_{\alpha\beta}$ satisfying the cocycle conditions defines a principal G-bundle.

As before, sections of a principal G-bundle can be described locally by collections of maps $\mathcal{S}_\alpha\colon U_\alpha \to G$, satisfying

$$\mathcal{S}_\alpha = g_{\alpha\beta} \mathcal{S}_\beta \quad \text{on} \quad U_{\alpha\beta}. \tag{17.7}$$

The relation between the notions of a principal G-bundle and a vector bundle is given by the following construction. Given a Lie group G, a principal G-bundle P over M, and a representation $R\colon G \to \mathrm{GL}(V)$ of G in a complex or real vector space V, there is a vector bundle $V_P \to M$ of rank $r = \dim V$, associated with P. It can be defined as the quotient

$$V_P = P \times_G V = (P \times V)/G, \tag{17.8}$$

where the right G-action on $P \times V$ is given by

$$(p, v) \cdot g = (p \cdot g, R(g^{-1})v), \quad p \in P,\ v \in V.$$

In particular, given a local section s of P over some open $U \subset M$ and a vector $v \in V$, we can define a section sv of V_P as the composition $U \to P \times V \to V_P$, where the first map is $x \mapsto (s(x), v)$.

If $g^P_{\alpha\beta} \in G$ are the transition functions of P, then the transition functions for the vector bundle V_P are given by $g^V_{\alpha\beta} = \rho(g^P_{\alpha\beta}) \in \rho(G) \subset \mathrm{GL}(V)$. In particular, applying it to the case when $G \subset \mathrm{GL}(V)$ and $\rho\colon G \to \mathrm{GL}(V)$ is the tautological representation, we see that in this case the principal G-bundle and the associated vector bundle have the same transition functions, so conversely, one can recover the principal G-bundle from the vector bundle. Thus, in this case the notions of a principal G-bundle and of a vector bundle with a structure group G are effectively equivalent.

An important example of a vector bundle associated with a principal G-bundle comes from the adjoint representation of G. Denote by $\mathfrak{g}$ the Lie algebra of G; it has a natural adjoint action of G. Thus, for every principal G-bundle P we can define a vector bundle

$$\mathfrak{g}_P = P \times_G \mathfrak{g}, \tag{17.9}$$

called the *adjoint bundle* (notation $\operatorname{ad} P$ is also used). In particular, if $G = \mathrm{GL}(V)$, then one can easily check that so defined vector bundle is given by $\mathfrak{g}_P = \operatorname{End} E$, where E is the vector bunde associated with the tautological representation of $\mathrm{GL}(V)$: $E = P \times_{\mathrm{GL}(V)} V$.

The vector bundle $\mathfrak{g}_P$ has an alternative description. Namely, let us say that a tangent vector $\xi \in T_pP$ is *vertical* if $\pi_*\xi = 0$, where $\pi\colon P \to M$ is the projection of P onto M. Equivalently, ξ is vertical if it is a tangent vector to the fiber P_x containing p. We will denote by $T^VP \subset TP$ the subbundle of vertical vectors:

$$T^V_pP = \{\xi \in T_pP \mid \pi_*\xi = 0\} = T_pP_x, \qquad x = \pi(p). \tag{17.10}$$

This subbundle is preserved by the action of G on TP (induced by the right action of G on P).

One easily sees that the right action of G on P induces, for every point $p \in P$, a vector space isomorphism

$$\begin{aligned} &\mathfrak{g} \simeq T^V_pP,\\ &a \mapsto \left.\frac{d}{dt}\right|_{t=0} (pe^{at}). \end{aligned} \tag{17.11}$$

We will denote the inverse isomorphism by ψ_p:

$$\psi_p\colon T^V_pP \simeq \mathfrak{g}. \tag{17.12}$$

Using this isomorphism, one can show (see Exercise 17.2) that one has natural isomorphism

$$\mathfrak{g}_P = T^VP/G, \tag{17.13}$$

where $T^V P = \{(p, \xi) \mid p \in P,\ \xi \in T_p^V P\}$ is the total space of the vertical tangent bundle of P.

17.2. Connections

Let E be rank r vector bundle over M, whose fibers are isomorphic to a real or complex r-dimensional vector space V ($\mathbb{R}^r$ or $\mathbb{C}^r$). Denote by $\Omega^k(E)$ the sheaf of smooth k-forms on M with values in E, and by $\Omega^k(M, E)$ the vector space of global sections of this bundle, i.e. the space of smooth k-forms on M with values in E.

Definition 17.2. A *connection* in a vector bundle E is a homomorphism of sheaves $\nabla\colon \Omega^0(E) \to \Omega^1(E)$, which satisfies the Leibniz rule

$$\nabla(fs) = df \otimes s + f(\nabla s) \tag{17.14}$$

for any local smooth section s of E and a smooth function f.

If we use local trivializations (17.1), then in each chart U_α the connection is given by

$$\nabla s_\alpha = ds_\alpha + A^\alpha s_\alpha, \tag{17.15}$$

where d is the de Rham differential and $A^\alpha \in \Omega^1(U_\alpha) \otimes \mathfrak{gl}(V)$ is a matrix-valued one-form. (From now on, we use the notation $\mathfrak{gl}(V)$ for the space $\operatorname{End} V$, considered either as an associative algebra or as a Lie algebra.)

It is easy to show that these one-forms satisfy the transformation law

$$A^\alpha = gA^\beta g^{-1} - dg\, g^{-1} \quad \text{on} \quad U_{\alpha\beta}, \tag{17.16}$$

where $g = g_{\alpha\beta}$ are the transition functions, or, equivalently,

$$A^\beta = g^{-1}A^\alpha g + g^{-1}dg \quad \text{on} \quad U_{\alpha\beta}. \tag{17.17}$$

Here, notation $dg\, g^{-1}$ and $g^{-1}dg$ should be understood in the most naive sense: g (and thus g^{-1}) is a matrix-valued function on M, so dg is a matrix whose entries are 1-forms, and $g^{-1}dg$ is the product of matrices:

$$\begin{aligned} (g^{-1}dg)_{ij} &= \sum_k (g^{-1})_{ik} dg_{kj}, \\ (dg\, g^{-1})_{ij} &= \sum_k (g^{-1})_{kj} dg_{ik} \end{aligned} \tag{17.18}$$

(see (17.28) below for a generalization).

Connection can be thought of as a way of differentiating sections of E: given vector field X on M and a section s,

$$\nabla_X(s) = X(\nabla s). \tag{17.19}$$

In terms of local trivialization, the covariant derivative is given by

$$\nabla_X(s_\alpha) = \partial_X s_\alpha + X(A^\alpha s_\alpha).$$

In particular, given local coordinates $x^1, \dots, x^n$ on a chart $U \subseteq M$, we can define covariant partial derivatives $\nabla_\mu = \nabla_{\partial_\mu}$ corresponding to coordinate vector fields; then

$$\nabla s = \nabla_\mu(s) dx^\mu, \quad \text{where} \quad \nabla_\mu = \partial_\mu + A_\mu \quad \text{and} \quad A = A_\mu dx^\mu. \tag{17.20}$$

Note that A_μ depend both on the trivialization of the bundle over U and on the choice of local coordinates x^μ.

Remark 17.3. In physics literature, the 1-form of the connection is usually defined by $\nabla_\mu = \partial_\mu + igA_\mu$, where g is some positive constant (in Maxwell's theory, $g = e$ is the elementary charge). Thus, when comparing our formulas to formulas in a physics texbook, you need to include appropriate power of ie.

Connections can be used to identify fibers of E at different points. Namely, given a path $\gamma\colon [0,1] \to M$ connecting points $x_0 = \gamma(0)$ and $x_1 = \gamma(1)$, for any initial value $s_0 \in E_{x_0}$ there is a unique way to extend s to a section of E over γ so that $\nabla_{\dot\gamma(t)} s(t) = 0$ for any $t \in [0,1]$ (here dot stands for derivative with respect to t, so if we think of $\gamma(t)$ as trajectory of a moving point, then $\dot\gamma(t) \in T_{\gamma(t)}M$ is the velocity vector at time t). This gives rise to holonomy map

$$\mathrm{Hol}_\gamma\colon E_{x_0} \to E_{x_1}, \tag{17.21}$$

which sends a value $s_0 \in E_{x_0}$ to the value $s_1 = s(1)$ where s is the section satisfying $\nabla_{\dot\gamma(t)}(s(t)) = 0$.

Note that in general, holonomy depends not only on the starting and final points but also on the choice of path γ; we will discuss it later when talking about curvature.

If E has structure group $G \subset \mathrm{GL}(V)$, it makes sense to only consider connections compatible with this structure group. The easiest way to define it is by requiring that for any path γ which is completely contained in one of charts U_α, the holonomy map Hol_γ (which, after choosing a local trivialization, can be considered as a map $V \to V$) is in G. It is easy to show that this condition is equivalent to requiring that in the formula $\nabla = d + A^\alpha$, the one-form A^α takes values in $\operatorname{End}\mathfrak{g}$, where $\mathfrak{g}$ is the Lie algebra of G.

17.3. Connections in principal G-bundles

Many of the above constructions can be repeated for principal G-bundles. Informally, a connection in a principal G-bundle $\pi\colon P \to M$ is a way of lifting paths in M to paths in P. More precise definition is as follows.

Note that in any bundle $\pi\colon P \to M$, we have a natural notion of vertical tangent vector, defined by (17.10). However, the bundle structure by itself does not give us a notion of horizontal vectors. Instead, this needs to be given as additional structure.

Definition 17.4. A *connection* in a principal G-bundle is a choice of subbundle $H \subset TP$, which is preserved by the right action of G and such that for every $p \in P$, the map $\pi_*\colon T_pP \to T_xM$, $x = \pi(p)$, induces an isomorphism $H_p \simeq T_xM$.

Note that the condition that $\pi_*\colon T_pP \to T_xM$ is an isomorphism is equivalent to requiring that for every $p \in P$, we have $T_pP = T_p^VP \oplus H_p$, where T_p^VP is the subbundle of vertical tangent vectors. This allows us to define for every tangent vector $\xi \in T_pP$ its "vertical" and "horizontal" components:

$$\xi = \xi^V + \xi^H, \qquad \xi^V \in T_p^VP, \quad \xi^H \in H_p. \tag{17.22}$$

Example 17.1. Let $P = M \times G$ be the trivial bundle. Then at each point $p = (x, g)$, we have a canonical isomorphism $T_pP = T_xM \oplus T_gG$ and thus we can define a connection by $H_p = T_xM$. So defined connection is usually called the trivial connection.

A connection in a principal G-bundle gives rise to equivariant path lifting.

Lemma 17.5. *Let P be a principal G-bundle P with a connection.*

(1) *For every path $\gamma\colon [0,1] \to M$, there exists a horizontal lift: a path $\hat\gamma\colon [0,1] \to P$ such that $\pi(\hat\gamma(t)) = \gamma(t)$ and for every t, the tangent vector $\frac{d}{dt}\hat\gamma(t)$ is horizontal. Any two such lifts can be obtained from each other by the right action of G on P; such a lift is unique if we fix the starting point $\hat\gamma(0) \in \pi^{-1}(\gamma(0))$.*

(2) *For a path γ in M, define the holonomy operator along γ by*

$$\begin{aligned} \mathrm{Hol}_\gamma\colon P_{\gamma(0)} &\to P_{\gamma(1)}, \\ p &\mapsto \hat\gamma(1), \end{aligned} \tag{17.23}$$

if $\hat\gamma$ is the horizontal lift of γ with starting point $\hat\gamma(0) = p \in P_{\gamma(0)}$. Then this operator commutes with the right action of G.

Note that in particular, if $P = M \times G$ is the trivial G-bundle so that each fiber is identified with G, then each holonomy operator is given by *left* multiplication by an element of G, which follows from the trivial observation that every map $G \to G$ which commutes with the right action of G is given by a left action by some element of G. Thus, in this case the holonomy gives a group morphism $\pi_1(M) \to G$. For example, for the trivial connection described in Example 17.1, we have $\mathrm{Hol}_\gamma = e$ for any γ.

Using Lemma 17.5, one can define a relation between connections in a principal bundle and in the associated vector bundle.

Theorem 17.6. *Let P be a principal G-bundle with a connection and let V_P be the associated vector bundle as defined in (17.8). Then there is a unique connection ∇ in V such that for every horizontal path $\hat\gamma(t)$ in P and a vector $v \in V$, the path $(\hat\gamma(t), v)$ in V_P is horizontal with respect to ∇.*

We will discuss another, more algebraic, way of relating connections in principal bundles and in associated vector bundles in the next section (see Lemma 17.12).

17.4. One-form of a connection

Instead of describing a connection in principal G-bundles by specifying a horizontal subbundle, one can also describe it by an appropriate 1-form, whose kernel consists of horizontal vectors.

Let P be a principal G-bundle. Recall that for every $p \in P$ we have an isomorphism $\psi_p\colon T_p^VP \to \mathfrak{g}$, where T_p^V is the subbundle of vertical vectors, see (17.12).

Definition 17.7. A connection 1-form is a form $A \in \Omega^1(P, \mathfrak{g})$ such that

(1) For any vertical $v \in T_p^VP$, we have $A(v) = \psi_p(v)$.

(2) For any $g \in G$, $u \in T_pP$, we have

$$A_{pg}(g_*u) = \mathrm{Ad}(g^{-1}) \cdot A_p(u), \tag{17.24}$$

where $g_*: T_pP \to T_{pg}P$ is the isomorphism of tangent spaces induced by the right action of g.

The last condition is just a statement that form A is equivariant with respect to the right action of G on P (and adjoint action of G on $\mathfrak{g}$).

Lemma 17.8. *Let $H \subset TP$ be a connection as defined in Definition 17.4. Define the form $A \in \Omega^1(P, \mathfrak{g})$ by*

$$A(u) = 0, \quad u \in H,$$
$$A(u) = \psi_p(u), \quad u \in T_p^V P.$$

Then A is a connection 1-form as defined in Definition 17.7.

Conversely, given a connection 1-form A, one can uniquely construct a connection H by $H_p = \mathrm{Ker}(A_p)$.

This lemma shows that instead of describing a connection by the choice of horizontal subbundle, we can as well describe it by the connection 1-form A. Thus, we will frequenlty refer to such a 1-form as the connection. This is usually a much more convenient (even if less intuitive) way of describing connections. As an easy application, we can prove the following result.

Theorem 17.9. *For a principal G-bundle P, the set of all connections is an affine space over the vector space $\Omega^1(M, \mathfrak{g}_P)$. We will denote it by $\mathcal{A}(P)$.*

Proof. Let A, A' be connection 1-forms. Then $\omega = A' - A$ is a $\mathfrak{g}$-valued 1-form on P such that for a vertical vector u, $\omega(u) = 0$. Thus, $\omega_p(v)$ depends only on the projection $\xi = \pi_*(v) \in TM$ and on point p. Moreover, it follows from (17.24) that if p, p' are two points in the same fiber P_x, $p' = pg$, then

$$\omega_{pg}(\xi) = \mathrm{Ad}\, g^{-1}(\omega_p(\xi))$$

and thus the class of pair $(p, \omega_p(\xi))$ in $P_x \times_G \mathfrak{g}$ doesn't depend on the choice of p. Therefore, $\omega(\xi)$ is well-defined as an element of $(\mathfrak{g}_P)_x$. Conversely, it is easy to prove that if A is a connection 1-form, and $\omega \in \Omega^1(M, \mathfrak{g}_P)$, then $A + \omega$ is also a connection 1-form.

To complete the proof, we need to show that the set of all connection 1-forms is non-empty, i.e. that there exists at least one connection. To prove it, note that if we choose a covering of M by open sets U_α such that $P|_{U_\alpha}$ can be trivialized, then on each U_α there exists a connection one-form A^α corresponding to the trivial connection. The general result now follows by the partition of unity argument: if $\varphi_\alpha \in C^\infty(M)$ are such that $\sum \varphi_\alpha = 1$ and support of φ_α is contained in U_α, then $A = \sum \varphi_\alpha A^\alpha$ is a connection 1-form on M. □

Example 17.2. Let $P = M \times G$ be the trivial G-bundle (with a fixed choice of trivialization). This bundle has a distinguished connection, namely the trivial connection, defined in Example 17.1. Thus, Theorem 17.9 shows that in this case, we have an isomorphism $\mathcal{A}(P) \simeq \Omega^1(M, \mathfrak{g})$.

In the above construction, the connection is described by a 1-form on P. However, this information is redundant: since A is equivariant under the right action of G, it suffices to know A_p at one point of the fiber P_x to uniquely recover it everywhere. Thus, if we choose a local section $s\colon U \to P|_U$ of the bundle P, we can uniquely recover A from the 1-form

$$A^s = s^*(A) \in \Omega^1(U, \mathfrak{g}) \tag{17.25}$$

or, equivalently,

$$A^s(\xi) = A_{s(x)} s_*(\xi), \quad x \in M,\ \xi \in T_x M.$$

It depends on the choice of a section s; to stress this dependence, we will use notation A^s. For example, it is immediate from the definition that $A^s = 0$ if and only if $s_*(\xi)$ is horizontal for any $\xi \in TM$. We will call such sections "horizontal". Note, however, that existence of such sections is not guaranteed, see Theorem 17.14.

Remark 17.10. Note that for a principal G-bundle, choosing a local section s is the same as choosing a local trivialization $\varphi\colon P|_U \to U \times G$. Indeed, a choice of a trivialization defines a section

$$s(x) = \varphi^{-1}(x, e),$$

and it is easy to see that conversely, a choice of section determines φ. Thus, we can also think of A^s as determined by a local trivialization φ; for this reason, we will also use the notation A^φ and refer to A^φ defined above as "local trivialization" of A:

$$A^\varphi = s^*(A) \in \Omega^1(M, \mathfrak{g}), \quad s(x) = \varphi^{-1}(x, e). \tag{17.26}$$

It is natural to ask how connection 1-form A^s changes when we change the section s (or, equivalenlty, how the form A^φ changes when we change the trivialization φ). To do that, we first need to introduce some notation.

Recall that for a function g with values in the group $\mathrm{GL}(r, \mathbb{C})$ we had defined matrix-valued 1-forms $g^{-1}dg$ and $dg \cdot g^{-1}$, see (17.18). We now need to define analogs of these 1-forms for arbitrary Lie group G.

For an element $g \in G$, let $L_g\colon G \to G$ be operator of the left multiplication by g. This induces isomorphism of tangent spaces $(Lg)_*\colon T_h G \to T_{gh} G$. In particular, we have the isomorphism

$$\begin{aligned} T_g G &\simeq \mathfrak{g} \\ \xi &\mapsto (L_{g^{-1}})_* \xi. \end{aligned} \tag{17.27}$$

For brevity, we will use the notation $g^{-1} \cdot \xi$ for $(L_{g^{-1}})_* \xi$. In a similar way, one can also define an isomorphism $T_g G \simeq \mathfrak{g}\colon \xi \mapsto (R_{g^{-1}})_* \xi$ induced by the right multiplication by g^{-1}; again, we will use shorter notation $\xi \cdot g^{-1}$.

Consider the $\mathfrak{g}$–valued 1-form θ_{MC} on G defined by

$$\theta_{MC}(\xi) = (g^{-1}) \cdot \xi \in \mathfrak{g}, \quad \xi \in T_g G. \tag{17.28}$$

This form is called the *Maurer–Cartan form* and is traditionally denoted just by $g^{-1}dg$. It is uniquely defined by the conditions that it is invariant under the left action of G on itself and $\theta_{MC}(\xi) = \xi$ for $\xi \in T_e G = \mathfrak{g}$ (see Exercise 17.5).

In a similar way we define the $\mathfrak{g}$-valued form $dg\, g^{-1}$ on G by

$$\langle dg\, g^{-1}, \xi \rangle = \xi \cdot g^{-1} \qquad \xi \in T_g G. \tag{17.29}$$

Theorem 17.11. *Let A be a connection in a principal G-bundle P, and let s_α, s_β be two local sections of P, related by $s_\beta(x) = s_\alpha(x)g(x)$ for some function $g\colon M \to G$. Denote by $A^\alpha = A^{s_\alpha}$, $A^\beta = A^{s_\beta}$ corresponding trivializations of the connection 1-form, as defined in (17.25).*

Then A^α, A^β are related by

$$A^\alpha = gA^\beta g^{-1} - dg\, g^{-1},$$
$$A^\beta = g^{-1}A^\alpha g + g^{-1}dg.$$

Note that instead of using sections s_α, s_β in the statement of the theorem, we could have used local trivializations φ_α, φ_β, related by $\varphi_\alpha = g\varphi_\beta$, see Remark 17.10.

As one might expect, connections in principal G bundles can be used to construct connections in vector bundles with structure group G.

Lemma 17.12. *Let P be a principal G-bundle over M with connection given by a 1-form $A \in \Omega^1(P, \mathfrak{g})$, let $R : G \to \mathrm{GL}(V)$ be a representation of G and let $\rho : \mathfrak{g} \to \operatorname{End} V$ be the corresponding representation of $\mathfrak{g} = Lie(G)$. Then the associated vector bundle V_P has a unique connection $\nabla = d_A$ with the property that for every section s of P and a vector $v \in V$,*

$$\langle d_A(s(x)v), \xi\rangle = \rho(A^s(\xi))v = \rho(A(s_*(\xi)))v, \quad \xi \in T_xM, \tag{17.30}$$

where $s(x)v \in P \times_G V$ is considered as a section of V_P.

Moreover, in terms of a local trivialization $\varphi\colon P|_U \to U \times G$ of P, which also gives a local trivialization of V_P, one has

$$d_A = d + \rho(A^\varphi), \tag{17.31}$$

where $A^\varphi \in \Omega^1(U, \mathfrak{g})$ is defined by (17.26).

We leave the simple proof of this lemma as an exercise to the reader.

In particular, for the adjoint bundle $\mathfrak{g}_P$ formulas (17.30)–(17.31) become

$$d_A s = ds + [A, s], \tag{17.32}$$

where s is a local section of $\mathfrak{g}_P$, and for brevity we denoted A^φ simply by A.

17.5. Gauge transformations

The gauge group $\mathcal{G}(P)$ of a principal G-bundle P consists of bundle isomorphisms $f\colon P \to P$ that commute with the right action of G. For the trivial bundle $P = M \times G$, the gauge group is the group of all G-valued functions on M, which acts on G by left multiplication: $\mathcal{G}(P) = C^\infty(M, G)$. This immediately follows from the fact that every diffeomorphism of G which commutes with the right action is given by left multiplication by some element $x \in G$.

More generally, using local trivializations (17.5), we see that elements of the gauge group $\mathcal{G}(P)$ are collections $\{f_\alpha\}_{\alpha\in A}$ of smooth functions $f_\alpha\colon U_\alpha \to G$ which satisfy the following relation:

$$f_\alpha = g_{\alpha\beta} f_\beta g_{\alpha\beta}^{-1} \qquad \text{on } U_\alpha \cap U_\beta.$$

From this, one immediately sees that the Lie algebra of the gauge group $\mathcal{G}(P)$ is

$$Lie(\mathcal{G}(P)) = \Gamma(M, \mathfrak{g}_P).$$

The gauge group acts on the space $\mathcal{A}(P)$ of all connections in P; this action is usually referred to as "gauge transformations", and its orbits are called "gauge equivalence classes" (of connections).

Since gauge transformation $\{f_\alpha\}_{\alpha\in A}$ change local trivializations by $\varphi_\alpha \mapsto \tilde{\varphi}_\alpha = f_\alpha\varphi_\alpha$, it acts on local connection 1-forms $A = A^\varphi$ by

$$A \mapsto \tilde{A} = gAg^{-1} - dg\, g^{-1}, \tag{17.33}$$

where we put $\tilde{A} = A^{\tilde{\varphi}}$ and $g = f_\alpha$. Thus the gauge transformations are exactly the transformations of Theorem 17.11. In particular, for the trivial bundle a choice of trivialization gives an isomorphism $\mathcal{A}(P) \simeq \Omega^1(M, \mathfrak{g})$ (see Example 17.2), and we have global gauge transformations

$$A \mapsto gAg^{-1} - dg\, g^{-1}, \quad A \in \Omega^1(M, \mathfrak{g}),\ g \in \mathcal{G}(P) = C^\infty(M, G). \tag{17.34}$$

We leave it to the reader to check that in general the action of the Lie algebra $\Gamma(M, \mathfrak{g}_P)$ on $\mathcal{A}(P)$ in locall trivialization is given by

$$u \cdot A = [u, A] - du, \quad u \in \Gamma(U_\alpha, \mathfrak{g}),\ A \in \Omega^1(U_\alpha, \mathfrak{g}) \tag{17.35}$$

("infinitesimal gauge transformations").

17.6. Curvature of a connection

For simplicity, we begin with the discussion of curvature for a connection in a vector bundle E. Recall that in this case, we defined the connection as a linear map $\nabla\colon \Omega^0(E) \to \Omega^1(E)$ satisfying (17.14).

Lemma 17.13. *A connection ∇ in a vector bundle E can be uniquely extended to a map $\Omega^\bullet(E) \to \Omega^{\bullet+1}(E)$ which satisfies the Leibniz condition*

$$\nabla(\eta \wedge \omega) = (d\eta) \wedge \omega + (-1)^k \eta \wedge \nabla\omega, \qquad \eta \in \Omega^k,\ \omega \in \Omega^l(E), \tag{17.36}$$

where Ω^k stands for sheaf of smooth k-forms on M.

Indeed, we can define the extension by

$$\nabla(\psi \otimes s) = d\psi \otimes s + (-1)^k \psi \wedge \nabla s, \tag{17.37}$$

where $s \in \Omega^0(E)$ and $\psi \in \Omega^k$.

If the vector bundle $E = V_P$ and connection ∇ come from a principal G-bundle P with a connection A, as described in Section 17.4, we will also denote the operator $\nabla\colon \Omega^\bullet(E) \to \Omega^{\bullet+1}(E)$ by d_A; in terms of local trivialization of P (and thus V_P), where the connection is described by 1-form A, we have

$$d_A(\psi \otimes v) = d\psi \otimes s + (-1)^k \psi \wedge \rho(A)v, \quad \psi \in \Omega^k, v \in V. \tag{17.38}$$

We will use notations ∇ and d_A interchangeably.

Consider now the map

$$\nabla^2\colon \Omega^0(E) \to \Omega^2(E).$$

Then (17.36) implies for $f \in \Omega^0$,

$$\begin{aligned}\nabla^2(fs) &= \nabla(df \otimes s + f\nabla s)\\ &= -df \wedge \nabla s + df \wedge \nabla s + f\nabla^2 s = f\nabla^2 s.\end{aligned}$$

This means that the map $\nabla^2 \colon \Omega^0(E) \to \Omega^2(E)$ is linear over Ω^0 and is determined by a 2-form $F \in \Omega^2(M, \operatorname{End} E)$:

$$\nabla^2 s = Fs.$$

Using local trivialization φ_α of E on U_α in which $\nabla = d + A^\alpha$, $A^\alpha \in \Omega^1(U_\alpha, \mathfrak{gl}(V))$, we see that

$$\begin{aligned}\nabla^2 s_\alpha &= (d + A^\alpha)(ds_\alpha + A^\alpha s_\alpha)\\ &= dA^\alpha s_\alpha - A^\alpha \wedge ds_\alpha + A^\alpha \wedge ds_\alpha + A^\alpha \wedge A^\alpha s_\alpha\\ &= (dA^\alpha + A^\alpha \wedge A^\alpha)s_\alpha,\end{aligned}$$

where $A_\alpha \wedge A_\alpha$ is understood as a product in the matrix algebra $\mathfrak{gl}(V)$ together with the usual exterior multiplication. Thus, in trivialization φ_α, we have

$$F^\alpha = dA^\alpha + A^\alpha \wedge A^\alpha. \tag{17.39}$$

Using (17.17) and the Maurer-Cartan formula (see Exercise 17.5), one easily sees that under the change of trivialization, F^α transforms as

$$F^\alpha = g_{\alpha\beta} F^\beta g_{\alpha\beta}^{-1} \quad \text{on} \quad U_{\alpha\beta}, \tag{17.40}$$

which once again shows that $F \in \Omega^2(M, \operatorname{End} E)$. It also implies that under gauge transformations (17.33),

$$F \to \tilde{F} = gFg^{-1}. \tag{17.41}$$

For future use, it is convenient to rewrite (17.39) in a slightly different form, namely

$$F^\alpha = dA^\alpha + \tfrac{1}{2}[A^\alpha \wedge A^\alpha], \tag{17.42}$$

where $[A_\alpha \wedge A_\alpha]$ is understood as the commutator in $\mathfrak{gl}(V)$, together with the usual exterior multiplication. We will often use notation

$$F = F_A = dA + A \wedge A = dA + \tfrac{1}{2}[A \wedge A].$$

In local coordinates $x^1, \dots, x^n$ on a chart $U \subseteq M$ the connection is given by $\nabla = d + A_\mu dx^\mu$, where $A_\mu \colon U \to \mathfrak{gl}(V)$, so

$$F = \tfrac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu, \quad \text{where} \quad F_{\mu\nu} = [\nabla_\mu, \nabla_\nu] = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu} + [A_\mu, A_\nu], \tag{17.43}$$

and $[A_\mu, A_\nu]$ is the commutator in $\mathfrak{gl}(V)$. This notation is used in physics textbooks.

It is also possible to define the curvature of a connection for a principal G-bundle. We skip the details of this construction, just mentioning that in this case, the curvature of a connection A is a 2-form $F_A \in \Omega^2(M, \mathfrak{g}_P)$, which in terms of a local trivialization is again given by the formula (17.42), only now $[\ ,\ ]$ is the commutator in the Lie algebra $\mathfrak{g}$. Similarly, in terms of local coordinates, formula (17.43) holds if we understand the commutator to be the commutator in $\mathfrak{g}$.

As discussed in the previous section, a connection A in a pricipal G-bundle P automatically defines a connection in any associated vector bundle V_P. In particular, a connection A in P defines a connection d_A in the adjoint bundle $\mathfrak{g}_P$ by formula (17.32), and thus, by Lemma 17.13, an operator

$$\begin{aligned} d_A\colon \Omega^p(M,\mathfrak{g}_P) &\to \Omega^{p+1}(M,\mathfrak{g}_P),\\ \omega &\mapsto d\omega + [A\wedge\omega] \end{aligned} \tag{17.44}$$

(as before, we are using a local trivialization of A). Correspondingly, $F_A \in \Omega^2(M,\mathfrak{g}_P)$ defines a linear over functions map

$$\begin{aligned} d_A^2\colon \Omega^p(M,\mathfrak{g}_P) &\to \Omega^{p+2}(M,\mathfrak{g}_P),\\ \omega &\mapsto [F_A\wedge\omega]. \end{aligned} \tag{17.45}$$

Applying (17.44) to the curvature F_A we get the Bianchi identity,

$$d_AF_A = dF + [A\wedge F] = [dA\wedge A] + [A\wedge(dA + \tfrac{1}{2}[A\wedge A])] = \frac{1}{2}[A\wedge[A\wedge A]] = 0. \tag{17.46}$$

The last equality follows from the Jacobi identity in $\mathfrak{g}$.

Bianchi identity can also be obtained from the Jacobi identity for covariant derviatives:

$$[[\nabla_\mu,\nabla_\nu],\nabla_\sigma] + [\nabla_\nu,\nabla_\sigma],\nabla_\mu] + [\nabla_\sigma,\nabla_\mu],\nabla_\nu] = 0.$$

In particular, if the group G is commutative, then Bianchi identity becomes $dF = 0$, which trivially follows from $F = dA$.

Curvature of a connection measures the failure of this connection to be flat. Recall that a connection ∇ (either in a vector bundle or a principal G-bundle over M) is called *flat* if its holonomy Hol_γ depends only on the homotopy class of γ and, therefore, determines a representation of the fundamental group of M. In this case, one can choose local trivialization φ_α so that the one-form of the connection vanishes: $A^\alpha = 0$; for a principal bundle, it is equivalent to choosing a section s which is horizontal. However, not every connection is flat.

Theorem 17.14. *A connection is flat if and only if its curvature is zero:* $F = 0$.

As before, we refer the reader to standard textbooks in differential geometry for the proof of this fact.

17.7. Chern–Weil theory

Connections allow one to define invariants of vector bundles (or of principal G-bunldes). This is known as Chern–Weil construction, and the resulting invariants are called the characteristic classes. The most famous example of them are the Chern and Pontryagin classes.

Let P be a principal G-bundle. Let $\Phi\colon \mathfrak{g}\to\mathbb{C}$ be a polynomial function on the Lie algebra $\mathfrak{g}$, homogeneous of degree k, and invariant under the adjoint action of G. Invariance under the adjoint action shows that for every section s of the adjoint bundle $\mathfrak{g}_P$, the function $\Phi(s)\in C^\infty(M)$ is well defined.

Choose a connection A in P and let $F \in \Omega^2(\mathfrak{g}_P)$ be the curvature of this connection. Then the same argument as above shows that we can define

$$\Phi(F) \in \Omega^{2k}(M, \mathbb{C}). \tag{17.47}$$

The following theorem summarizes the main results of Chern–Weil theory.

Theorem 17.15. *Let F_A be the curvature of a connection A in principal G-bundle P and let $\Phi(F)$ be as defined above. Then we have the following results.*

(1) *The $2k$-form $\Phi(F)$ on M is closed:*
$$d\Phi(F) = 0.$$
(2) *Cohomology class*
$$[\Phi(F)] \in H^{2k}(M, \mathbb{C})$$
does not depend on a choice of a connection A in P.
(3) *A map*
$$\Phi \mapsto \Phi(F)$$
is a homomorphism of the commutative algebra of invariant polynomials on $\mathfrak{g}$ into the commutative algebra $H^{\text{even}}(M, \mathbb{C})$ of differential forms of even degree on M.

The map $\Phi \mapsto \Phi(F)$ is called the *Weil homomorphism*, and cohomology classes $[\Phi(F)]$ are called the *characteristic classes* of a bundle P.

The algebra of invariant polynomials is well known for all compact Lie groups. In particular, for $G = \mathrm{U}(r)$ and more generally for $\mathrm{GL}(r, \mathbb{C})$, this algebra is a free algebra generated by elements P^k, $\deg P^k = k$, $k = 1, \dots r$. These generators are called *elementary invariant polynomials* and are defined by

$$\det(X + tI) = \sum_{k=0}^{r} P^k(X) t^{r-k}, \quad \text{where} \quad X \in \mathfrak{gl}(r, \mathbb{C}), \quad P^0 = 1.$$

Thus, if X has eigenvalues λ_i, then

$$\begin{aligned} P^1(X) &= \sum \lambda_i = \operatorname{tr}(X), \\ P^2(X) &= \sum_{i<j} \lambda_i \lambda_j, \\ &\dots \\ P^r(X) &= \prod \lambda_i = \det(X). \end{aligned}$$

Differential forms

$$c_k(F) = P^k\left(\frac{i}{2\pi} F\right) \tag{17.48}$$

are called *Chern forms*, and the corresponding cohomology classes are called *Chern classes.*

If E is a complex vector bundle of rank r, then its Chern classes are in fact integral cohomology classes:

$$c_k(E) = \left[P^k\left(\frac{i}{2\pi} F\right)\right] \in \check{H}^{2k}(M, \mathbb{Z}), \quad k = 1, \dots, r.$$

Here $\check{H}^{2k}(M,\mathbb{Z})$ stands for the Čech cohomology with coefficients in the constant sheaf $\mathbb{Z}$, and $c_k(E)$ is understood as an image of a differential form (17.48) under the Čech-de Rham isomorphism.

If E is a real vector bundle of rank r, then a choice of a Riemannian metric on E allows to reduce its structure group from $\mathrm{GL}(r,\mathbb{R})$ to $\mathrm{O}(r)$. Since its Lie algebra $\mathfrak{o}(r)$ consists of skew-symmetric matrices, we have

$$\det(tI+X)=\det(tI-X), \quad X\in\mathfrak{o}(r),$$

so $P^k(X)=0$ for odd k. Integral characteristic classes

$$p_k(E)=P^{2k}\left(\frac{1}{2\pi}F\right)\in H^{4k}(M,\mathbb{R})$$

are called *Pontryagin classes*. They are related to the Chern classes of the complexified vector bundle $E\otimes_{\mathbb{R}}\mathbb{C}$ by the following simple formula

$$p_k(E)=(-1)^k c_{2k}(E\otimes_{\mathbb{R}}\mathbb{C})$$

(note that $c_{2k+1}(E\otimes_{\mathbb{R}}\mathbb{C})=0$).

Example 17.3. An easy calculation shows that $\operatorname{tr}(X^2)=P^1(X)^2-2P^2(X)$ for any matrix X. Thus, we have the following formula, which is frequently used: if E is a complex vector bundle of rank r, and F is the curvature of some connection in E, then

$$[\operatorname{tr}F\wedge F]=4\pi^2(2c_2(E)-c_1^2(E))\in H^4(M).$$

In particular, if E is a $\mathrm{SU}(r)$-bundle, then $c_1(E)=0$, so $[\operatorname{tr}F\wedge F]=8\pi^2c_2(E)$, and if E is a real vector bundle, then

$$p_1(E)=-\frac{1}{8\pi^2}[\operatorname{tr}F\wedge F].$$

17.8. Exercises

Exercise 17.1. Let $\pi:P\to M$ be a principal G-bundle. Show that the pullback bundle $\pi^*P\to P$ is a trivial principal G-bundle.

Exercise 17.2. Prove formula (17.13): $\mathfrak{g}_P=T^VP/G$.

Exercise 17.3. Find local trivializations for a vector bundle defined by (17.8), and show that in this case definition (17.14) reduces to (17.16).

Exercise 17.4. Prove that for $G=\mathrm{GL}(r,\mathbb{C})$, the definition of Maurer–Cartan form $g^{-1}dg$ given in (17.28) agrees with the definition given in (17.18).

Exercise 17.5. Show that the Maurer–Cartan form (17.28) is left-invariant (if we let G act trivially on $\mathfrak{g}$). Deduce from this the Maurer–Cartan formula

$$d\theta+\tfrac{1}{2}[\theta\wedge\theta]=0$$

where $[\theta\wedge\theta]$ stands for the wedge product of 1-forms combined with commutator in $\mathfrak{g}$.

Exercise 17.6. Show that extension of a connection ∇ given by (17.37) satisfies formula (17.36) in Lemma 17.13.

Exercise 17.7. This problem generalizes results of Exercise 15.7 to forms with values in vector bundles with connection.

Let M be an n-dimensional oriented pseudo-Riemannian manifold with volume form ω; let P be a G-bundle on M, with a connection A. Let V be a representation of G and and let V_P be the associated vector bundle. Assume that we have chosen a G-invariant symmetric bilinear form $\langle\ ,\ \rangle$ on V, and define, for $\alpha, \beta \in \Omega^k(M, V_P)$,

$$((\alpha, \beta)) = \int_M (\alpha, \beta)\omega = \int_M \langle \alpha \wedge \star\beta \rangle$$

where for $\alpha = \alpha_i X_i \in \Omega^k(M, V_P)$, $\beta = \beta_j Y_j \in \Omega^k(M, V_P)$, with α_i, β_j being scalar k-forms and X_i, Y_j local sections of V_P, we define

$$(\alpha, \beta) = \sum \langle X_i, Y_j \rangle (\alpha_i, \beta_j),$$
$$\langle \alpha \wedge \star\beta \rangle = \sum \langle X_i, Y_j \rangle \alpha_i \wedge \star\beta_j.$$

(1) Prove that then

$$((d_A\alpha, \beta)) = ((\alpha, d_A^*\beta)), \qquad \alpha \in \Omega^{k-1}(M, V_P), \beta \in \Omega^k(M, V_P).$$

where $d_A^* \colon \Omega^k(M, V_P) \to \Omega^{k-1}(M, V_P)$ is given by

$$d_A^* = (-1)^k \star^{-1} d_A \star$$

(2) Show that for a 1-form $\alpha = \alpha_\mu dx^\mu$, we have

$$d_A^*\alpha = -\nabla_\mu^A \alpha^\mu = -\nabla_\mu^A(\eta^{\mu\nu}\alpha_\nu)$$

where η is the metric in M, and as in Exercise 15.7 , we assume that the volume form is given by $dx^1 \wedge \cdots \wedge dx^n$.

Exercise 17.8. Show that part (1) of Theorem 17.15 follows from the Bianchi identity (17.46).

Exercise 17.9. Prove part (2) of Theorem 17.15 (*Hint*: the set of connections is an affine space and thus any two connections can be connected by a path).

Exercise 17.10. Prove that for every closed complex-valued 2-form F on a compact manifold M with the property

$$\left[\frac{i}{2\pi}F\right] \in \check{H}^2(M, \mathbb{Z}),$$

there is a complex line bundle $L \to M$ and a connection $\nabla = d + A$ in L such that $F = dA$.

Chapter 18

Yang–Mills Theory

In this chapter, we define one of the most important field theories in modern physics, the Yang–Mills theory. It depends on a choice of compact Lie group G; special cases of Yang–Mills theory include Maxwell's theory of electromagnetism for $G = \mathrm{U}(1)$, theory of weak interactions for $G = \mathrm{SU}(2)$ and theory of strong interactions for $G = \mathrm{SU}(3)$.

Throughout this chapter, G is a compact real Lie group with the Lie algebra $\mathfrak{g}$. We fix a choice of a positive definite invariant symmetric bilinear form

$$\langle\ ,\ \rangle \colon \mathfrak{g} \otimes \mathfrak{g} \to \mathbb{R}$$

on $\mathfrak{g}$. It is known that such form always exists; if $\mathfrak{g}$ is simple, then it is unique up to a scalar factor — for example, one can take the negative of the Killing-Cartan form: $\langle u, v\rangle = -\operatorname{tr}(\operatorname{ad}_u \circ \operatorname{ad}_v)$. Another common choice for $\mathfrak{su}(n)$ is

$$\langle u, v\rangle = -\operatorname{tr}(uv).$$

All results of this chapter hold for any choice of form $\langle\ ,\ \rangle$.

We will frequently combine this metric on $\mathfrak{g}$ with the wedge product of differential forms on M: for $\alpha = \sum_i \alpha_i u_i \in \Omega^p(M, \mathfrak{g})$, $\beta = \sum_j \beta_j v_j \in \Omega^q(M, \mathfrak{g})$ we define

$$\langle \alpha \wedge \beta\rangle = \sum_{i,j} \langle u_i, v_j\rangle \alpha_i \wedge \beta_j, \quad u_i, v_j \in \mathfrak{g}, \quad \alpha_i \in \Omega^p(M), \beta_j \in \Omega^q(M). \tag{18.1}$$

18.1. Yang–Mills theory

Let M be an oriented pseudo-Riemannian manifold of dimension $n+1$; usually we will take $M = \mathbb{R}^{1,3}$. This manifold will play the role of the spacetime in our theory. We denote by $d^{n+1}x$ the corresponding volume form on M and by $\star$ the Hodge star operator as defined in Section 15.2. We also fix a principal G-bundle P on M (see Chapter 17). Note that for $M = \mathbb{R}^{1,3}$, every principal G-bundle on M is necessarily trivial.

The space of fields of Yang–Mills theory is the space $\mathcal{A}(P)$ of all connections in P; recall that by results of Section 17.4, this is an affine space over the vector space $\Omega^1(M, \mathfrak{g}_P)$. For a connection A, we denote by $F_A \in \Omega^2(M, \mathfrak{g}_P)$ the curvature of this connection. We recall that in local coordinates on M,

$$F = \tfrac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu, \quad \text{where} \quad F_{\mu\nu} = [\nabla_\mu, \nabla_\nu] = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu} + [A_\mu, A_\nu]$$

(see formula (17.43)).

We define the Lagrangian density for a connection A by

$$\mathcal{L}(A) = -\frac{1}{2}\langle F_A \wedge \star F_A \rangle = \mathscr{L}(A)\, d^{n+1}x, \tag{18.2}$$

where $\star$ is the Hodge star operator, and

$$\mathscr{L}(A) = -\frac{1}{4}\langle F_{\mu\nu}, F^{\mu\nu} \rangle \tag{18.3}$$

(compare with Exercise 15.2).

Thus the action of Yang–Mills theory is defined by

$$S(A) = \int_M \mathcal{L}(A) = -\frac{1}{2}\int_M \langle F_A \wedge \star F_A \rangle = -\frac{1}{4}\int_M \langle F_{\mu\nu}, F^{\mu\nu} \rangle d^{n+1}x, \tag{18.4}$$

where as before, we assume that A is such that the integral converges.

Lemma 18.1. *The Yang–Mills Lagrangian (18.2) is invariant under gauge transformations $A \to gAg^{-1} - dgg^{-1}$, where $g \in \mathcal{G}(P)$.*

Indeed, according to (17.41), the curvature F_A transforms under gauge transformations by $F_A \mapsto gF_Ag^{-1}$, and the bilinear form $\langle\ ,\ \rangle$ is invariant under the adjoint action of G.

Theorem 18.2. *Critical points of Yang–Mills action (18.4) are given by*

$$d_A \star F_A = 0 \tag{18.5}$$

Note that by Bianchi identities (17.46) we have $d_A\, F_A = 0$.

Proof. Critical points of the action are given by $\delta S = 0$. Since $F_A = dA + \frac{1}{2}[A \wedge A]$, it is immediate that $\delta F_A = d(\delta A) + [\delta A \wedge A] = d_A(\delta A)$. Thus,

$$\begin{aligned}\delta S &= -\frac{1}{2}\int \Big\langle \delta F_A \wedge \star F_A + F_A \wedge \star \delta F_A \Big\rangle = -\int \langle \delta F_A \wedge \star F_A \rangle \\ &= -\int \Big\langle (d\delta A + [\delta A \wedge A]) \wedge \star F_A \Big\rangle,\end{aligned}$$

where we used $\alpha \wedge \star\beta = \beta \wedge \star\alpha$, see (15.7).

Since the form $\langle\ ,\ \rangle$ on $\mathfrak{g}$ is G-invariant, we have $\langle [a,b], c\rangle = \langle a, [b,c]\rangle$. Together with Stokes' theorem, we can use it to rewrite

$$\delta S = -\int \big\langle \delta A \wedge (d\star F_A + [A \wedge \star F_A]) \big\rangle.$$

Thus, $\delta S = 0$ for any variation δA if and only if

$$d\star F_A + [A \wedge \star F_A] = 0,$$

which is exactly the equation $d_A \star F_A = 0$, see (17.44). □

Remark 18.3. The proof can be somewhat simplified if we use Exercise 17.7:

$$\delta S = -((d_A \delta A, F_A)) = -((\delta A, d_A^* F_A))$$

so $\delta S = 0$ for any δA if and only if $d_A^* F_A = \pm \star d_A \star F_A = 0$.

Most of the time we will assume that the spacetime M has the form $M = N \times \mathbb{R}$, with N playing the role of space. As usual, we will use $x^0 = t$ for time coordinate in $\mathbb{R}$, and x^i, $i = 1, \dots, n$ for coordinates in the space N. We assume that the metric in M is given by

$$\eta = (dx^0)^2 - g_{ij} dx^i \, dx^j, \tag{18.6}$$

where $g_{ij} dx^i \, dx^j$ is some (positive definite) Riemannian metric on N.

In this case, we can write

$$\begin{aligned} A &= A_0 dx^0 + \boldsymbol{A}, \\ F &= dx^0 \wedge E(\boldsymbol{x}) - B(\boldsymbol{x}) \end{aligned} \tag{18.7}$$

where $\boldsymbol{A} = A|_N$ is a time-dependent connection on N, and $E \in \Omega^1(N, \mathfrak{g}_P)$, $B \in \Omega^2(N, \mathfrak{g}_P)$ are time-dependent differential forms on N (compare with (15.16)):

$$\begin{aligned} E &= E_i \, dx^i, \quad E_i = F_{0i} \\ B &= \frac{1}{2} B_{ij} dx^i \wedge dx^j, \quad B_{ij} = -F_{ij}. \end{aligned}$$

Then Lagrangian (18.3) can be rewritten in the form

$$\mathscr{L}(A) = \frac{1}{2}\big(|E|^2 - |B|^2\big), \tag{18.8}$$

where the norms $|\cdot|^2$ are defined using the (positive definite) metric g_{ij} in N:

$$\begin{aligned} |E|^2 &= g^{ij} \langle E_i, E_j \rangle = -\langle F_{0i}, F^{0i} \rangle, \\ |B|^2 &= \frac{1}{2} g^{ik} g^{jl} \langle B_{ij}, B_{kl} \rangle = \frac{1}{2} \langle F_{ij}, F^{ij} \rangle, \end{aligned} \tag{18.9}$$

where, as usual, we use the metric η for raising indices: $F^{\mu\nu} = \eta^{\mu\alpha} \eta^{\nu\beta} F_{\alpha\beta}$.

In this case, Bianchi equation $d_A F = 0$ is rewritten in terms of E, B as

$$d_{\boldsymbol{A}} B = 0, \quad d_{\boldsymbol{A}} E = -\nabla_0^A B, \tag{18.10}$$

where $d_{\boldsymbol{A}}$ now stands for the exterior derivative operator $\Omega^k(N, \mathfrak{g}_P) \to \Omega^{k+1}(N, \mathfrak{g}_P)$ on forms on N defined by connection $\boldsymbol{A} = A|_N$, and ∇_0^A is the covariant derivative operator on N defined by A: $\nabla_0^A B = \partial_0 B + [A_0, B]$.

By Exercise 15.4, we have $\star F = -dx^0 \wedge \star_N B - \star_N E$, where $\star_N$ is the Hodge operator on N defined by metric g. Thus, the equations of motion $d_A \star F = 0$ become

$$d_{\boldsymbol{A}} \underset{N}{\star} E = 0, \quad d_A \underset{N}{\star} B = \nabla_0^A (\underset{N}{\star} E). \tag{18.11}$$

These formulas are analogs of Maxwell's equations (15.23).

We note that in local coordinates x^i on N, equation $d_{\boldsymbol{A}} \star_N E = 0$ is rewritten as

$$\nabla_i^{\boldsymbol{A}} E^i = \nabla_i^{\boldsymbol{A}} (g^{ij} E_j) = 0, \tag{18.12}$$

see Exercise 17.7 (compare with the Gauss law — Maxwell's equation $\boldsymbol{\nabla} \cdot \boldsymbol{E} = 0$).

18.2. Line bundles and Maxwell's equations

Let us now consider a special case of Yang–Mills theory, when the group G is the unitary group U(1) with the Lie algebra $\mathfrak{u}(1)$. We choose a generator X of $\mathfrak{u}(1)$ and the bilinear form on $\mathfrak{u}(1)$ such that $\langle X, X\rangle = 1/4\pi$. Thus we can identify

$$\begin{aligned} \mathfrak{u}(1) &\simeq \mathbb{R}, \\ aX &\mapsto a, \end{aligned}$$

and under this identification, the bilinear form becomes $\langle aX, bX\rangle = \dfrac{ab}{4\pi}$.

As before, we assume that we have chosen a principal U(1)-bundle P — or equivalently, an associated line bundle $L = \mathfrak{u}(1)_P$ — and a connection in this bundle; in local trivialization of P, the connection is described by one-form

$$A^\alpha \in \Omega^1(U_\alpha, \mathfrak{u}) \simeq \Omega^1(U_\alpha, \mathbb{R}).$$

Since the group U(1) is commutative, some of the formulas above are simplified. In particular, the curvature of this connection is given by

$$F = dA,$$

and thus is a closed form: $dF = 0$ (which can be considered as a special case of Bianchi identity).

The Yang–Mills action functional (18.4) on the affine space $\mathcal{A}(L)$ of unitary connections on the line bundle L takes the form

$$S(A) = -\frac{1}{8\pi}\int\limits_M F\wedge\star F = -\frac{1}{16\pi}\int_M F_{\mu\nu}F^{\mu\nu}d^nx \tag{18.13}$$

(in case M is non-compact it is assumed that the connection is such that the integral with $F = dA$ is convergent).

Now consider the simplest case when when $M = \mathbb{R}^{1,3}$ is the Minkowski space and P is the trivial U(1)-bundle, and compare it with the results in Section 15.4. We see that the action functional of the U(1) Yang-Mills theory — formula (18.13) — coincides with the action functional of the Maxwell's theory in the absence of sources — formula (15.42)! In this case, the main statement of Theorem 18.2 reduces to what we had previously computed for the electromagnetic theory in Section 15.4: the critical points of the functional $S(A)$ are given by the Maxwell's equations

$$d\star F_A = d\star dA = 0. \tag{18.14}$$

18.3. Stress-energy tensor for Yang–Mills theory

As in Maxwell's theory, we can define the stress-energy tensor for the Yang–Mills theory. It is computed exactly in the same way as for Maxwell's theory, see Section 15.5. The only place that requires a minor change is the proof that

$$L_X(A) = \imath_X F_A \tag{18.15}$$

modulo gauge equivalence, which now requires the use of formula (17.35) for the infinitesimal gauge transformations. The final answer is given by

$$T^{\mu\nu} = -\eta_{\alpha\beta}\langle F^{\mu\alpha}, F^{\nu\beta}\rangle + \frac{1}{4}\eta^{\mu\nu}\langle F_{\alpha\beta}, F^{\alpha\beta}\rangle. \tag{18.16}$$

Note that this tensor is traceless ($T^{\mu}_{\mu} = 0$) only for $n = 3$, i.e. when the spacetime is 4-dimensional.

In particular, in the situation considered at the end of Section 18.1, where $M = N \times \mathbb{R}$ is a general spacetime with the pseudo-Riemannian metric (18.6), we define the energy density of the Yang–Mills field by

$$\begin{aligned}\mathscr{E} = T^{00} &= -\eta_{ij}\langle F^{0i}, F^{0j}\rangle + \frac{1}{2}|F|^2 \\ &= \frac{1}{2}\big(|E|^2 + |B|^2\big), \quad |F|^2 = \frac{1}{2}\langle F_{\alpha\beta}, F^{\alpha\beta}\rangle.\end{aligned}$$

where $|E|^2$, $|B|^2$ are defined by (18.9).

The same arguments as in Maxwell's case (see Section 15.5) give the following result.

Theorem 18.4. *Total energy of the Yang–Mills field:*

$$H = \int_N \mathscr{E}(\boldsymbol{x}, t)\, d^n\boldsymbol{x}$$

is conserved: $\partial_0 E = 0$.

18.4. Coupling Yang-Mills field with matter

Like in Section 16.5, we can couple the Yang-Mills field over Minkowski spacetime $\mathbb{R}^{1,3}$ with other relativistic fields. For instance, consider a spinor field that transforms by the fundamental (defining) representation $G \to \mathrm{GL}(V)$ of the structure group G. In case $G = \mathrm{U}(r)$ it is a spinor-valued vector

$$\boldsymbol{\psi} = \begin{pmatrix}\psi^1 \\ \vdots \\ \psi^r\end{pmatrix},$$

and we have the following generalization of the Dirac Lagrangian density:

$$\mathscr{L}(\bar{\boldsymbol{\psi}}, \boldsymbol{\psi}) = i\bar{\boldsymbol{\psi}}\gamma^\mu\partial_\mu\boldsymbol{\psi} - m\bar{\boldsymbol{\psi}}\boldsymbol{\psi}, \tag{18.17}$$

where $\bar{\boldsymbol{\psi}} = (\bar{\psi}^1, \dots, \bar{\psi}^r)$ is the adjoint spinor-valued vector, considered as independent field. Clearly, Lagrangian (18.17) is invariant under the action $\boldsymbol{\psi} \to g\boldsymbol{\psi}, \bar{\boldsymbol{\psi}} \to \bar{\boldsymbol{\psi}}g^{-1}$, $g \in G$.

Now let P be a principal G-bundle, A be a connection in P and $\mathscr{L}(A)$ be the Yang-Mills Lagrangian density (18.3). It admits a coupling with the spinor Lagrangian (18.17): according to the minimal coupling procedure, discussed in Section 16.5, we replace partial derivatives ∂_μ by the covariant derivatives

$$D_\mu = \partial_\mu + \rho(A_\mu), \tag{18.18}$$

where $\rho : \mathfrak{g} \to V$ is the fundamental representation. The resulting Lagrangian

$$\mathscr{L}(A, \bar{\boldsymbol{\psi}}, \boldsymbol{\psi}) = -\frac{1}{4}\langle F_{\mu\nu}, F^{\mu\nu}\rangle + \bar{\boldsymbol{\psi}}(i\not{D} - m)\boldsymbol{\psi}, \quad \text{where} \quad \not{D} = D_\mu\gamma^\mu \tag{18.19}$$

is invariant under the gauge transformations

$$A_\mu \to gA_\mu g^{-1} - \partial_\mu g g^{-1}, \quad \boldsymbol{\psi} \to g\boldsymbol{\psi}, \quad \bar{\boldsymbol{\psi}} \to \bar{\boldsymbol{\psi}} g^{-1},$$

where $g \in \mathcal{G}(P) = C^\infty(\mathbb{R}^4, G)$.

Remark 18.5. In case $G = \mathrm{SU}(3)$ Lagrangian (18.19) is the Lagrangian of quantum chromodynamics (QCD), describing the interaction of eight gluon fields — components of $A_\mu(x) \in \mathfrak{su}(3)$ with respect to some basis of $\mathfrak{g} = \mathfrak{su}(3)$ — with six quark and anti-quark fields, component of the spinor-valued vectors $\boldsymbol{\psi}$ and $\bar{\boldsymbol{\psi}}$. The quarks and antiquarks are massive elementary particles of spin $\frac{1}{2}$ (fermions) and fractional electric charges $-\frac{1}{3}e, \frac{2}{3}e$ and $\frac{1}{3}e, -\frac{2}{3}e$, while gluons are massless elementary particles of spin 1 (bosons) that mediates the strong interaction between quarks and carry no electric charge. In physics terminology, three components ψ_i of $\boldsymbol{\psi}$ and three components $\bar{\psi}_i$ of $\bar{\boldsymbol{\psi}}$ are distinguished by three "colors" (say green, red and blue) and by three "anti-colors" (anti-green, anti-red and anti-blue); components of $A_\mu(x)$ are also described by color and anti-color indices. The composite subatomic particles like protons and neutrons are made of quarks and antiquarks and carry zero "color charge".

In a similar way we can couple the Yang-Mills field with the complex vector field $\boldsymbol{\varphi}$. Namely, let V be the fundamental representation of the structure group G — a finite-dimensional complex vector space with the Hermitian inner product. The Lagrangian for field $\boldsymbol{\varphi} : \mathbb{R}^4 \to V$ is

$$\mathscr{L}(\boldsymbol{\varphi}, \partial_\mu \boldsymbol{\varphi}) = \partial_\mu \boldsymbol{\varphi}^\dagger \partial^\mu \boldsymbol{\varphi} - V(|\boldsymbol{\varphi}|^2). \tag{18.20}$$

Here $\boldsymbol{\varphi}^\dagger$ is Hermitian conjugate of the vector $\boldsymbol{\varphi}$, so $\boldsymbol{\varphi}^\dagger \boldsymbol{\varphi} = |\boldsymbol{\varphi}|^2$, and $V : \mathbb{R}_{>0} \to \mathbb{R}$ is some potential. According to the minimal coupling procedure, we should replace ∂_μ by D_μ as in (18.18), and consider the Lagrangian

$$\mathscr{L}(A, \boldsymbol{\varphi}, \partial_\mu \boldsymbol{\varphi}) = -\frac{1}{4}\langle F_{\mu\nu}, F^{\mu\nu}\rangle + D_\mu \boldsymbol{\varphi}^\dagger D^\mu \boldsymbol{\varphi} - V(|\boldsymbol{\varphi}|^2). \tag{18.21}$$

It is invariant under gauge transformations

$$A_\mu \to gA_\mu g^{-1} - \partial_\mu g g^{-1}, \quad \boldsymbol{\varphi} \to g\boldsymbol{\varphi}, \quad g \in \mathcal{G}(P) = C^\infty(\mathbb{R}^4, G).$$

Remark 18.6. In case $G = \mathrm{SU}(2)$ and $V(|\boldsymbol{\varphi}|^2) = \mu^2|\boldsymbol{\varphi}|^2 - \lambda|\boldsymbol{\varphi}|^4$, where $\mu^2 > 0$ and $\lambda > 0$, Lagrangian (18.21) describes the famous Higgs field from the Standard Model of elementary particles. This model unifies three out of four fundamental interactions: electromagnetic, weak and strong (the gravity is excluded), and classifies all known elementary particles. Lagrangian of the Standard Model is gauge invariant with the $\mathrm{SU}(3)\times\mathrm{SU}(2)\times\mathrm{U}(1)$ structure group, and is too complicated to be presented here.

18.5. Hamiltonian formalism in Yang–Mills theory

Let us now describe the Hamiltonian formalism in the Yang–Mills theory. We had already done that for the Maxwell's theory in Section 16.4; here we generalize it to arbitrary structure group G. For simplicity, we will only consider the Minkowski spacetime $M = \mathbb{R}^{1,3}$; however, most results can be generalized to other spacetimes as well, see Remark FIXME. Do we need it?

As in Section 18.1, we let A be a connection in a principal G-bundle P, and consider the Yang–Mills Lagrangian density

$$\mathcal{L}(A) = -\frac{1}{2}\langle F_A \wedge \star F_A\rangle = \mathscr{L}(A)\, d^4x,$$

where F_A is the curvature of connection A, and $\mathscr{L}(A)$ is the Lagrangian function, which in local coordinates takes the form

$$\mathscr{L}(A) = -\frac{1}{4}\langle F_{\mu\nu}, F^{\mu\nu}\rangle, \qquad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu].$$

As in case of classical electrodynamics, we can rewrite $\mathscr{L}$ in the first order formalism. Namely, note that

$$|E|^2 = \sum \langle E_i, E_i\rangle = 2\langle \partial_0 A_i - \partial_i A_0 + [A_0, A_i], E_i\rangle - |E|^2.$$

Thus, we have

$$\mathscr{L} = \frac{1}{2}(|E|^2 - |B|^2) = \langle \partial_0 A_i - \partial_i A_0 + [A_0, A_i], E_i\rangle - \frac{1}{2}(|E|^2 + |B|^2), \tag{18.22}$$

Using invariance of the inner product, we can rewrite $\langle [A_0, A_i], E_i\rangle = \langle A_0, [A_i, E_i]\rangle$. Also,

$$\langle \partial_i A_0, E_i\rangle = -\langle A_0, \partial_i E_i\rangle + \partial_i\langle A_0, E_i\rangle.$$

Since adding a total derivative does not change the action

$$S(A) = \int_{\mathbb{R}^3} \mathscr{L} d^3x,$$

and thus does not change the equations of motion, we can replace the original Lagrangian by

$$\begin{aligned}\mathscr{L} &= \langle \partial_0 A_i, E_i\rangle + \langle A_0, \partial_i E_i + [A_i, E_i]\rangle - \frac{1}{2}(|E|^2 + |B|^2) \\ &= \langle E_i, \partial_0 A_i\rangle - \frac{1}{2}(|E|^2 + |B|^2) + \langle A_0, C(\boldsymbol{x})\rangle,\end{aligned} \tag{18.23}$$

where

$$C(\boldsymbol{x}) = \partial_i E_i + [A_i, E_i] = \nabla_i^A E_i = -d_{\boldsymbol{A}}^* E, \tag{18.24}$$

(see Exercise 17.7).

As in Section 16.4, we see that this is a singular first order Lagrangian of the form (6.13):

$$\mathcal{L} = \boldsymbol{p}\dot{\boldsymbol{q}} - H(\boldsymbol{p}, \boldsymbol{q}) - \sum_{a=1}^{m} \lambda_a \varphi^a(\boldsymbol{p}, \boldsymbol{q}),$$

with $A_i(\boldsymbol{x})$ playing the role of generalized coordinates $\boldsymbol{q}$, $E_i(\boldsymbol{x})$ playing the role of generalized momenta $\boldsymbol{p}$, and $H(\boldsymbol{p}, \boldsymbol{q})$ given by the total energy of the Yang–Mills field

$$H = \int_{\mathbb{R}^3} \mathscr{E}(\boldsymbol{x}) d^3\boldsymbol{x} = \frac{1}{2}\int_{\mathbb{R}^3} (|E|^2 + |B|^2)\, d^3\boldsymbol{x}. \tag{18.25}$$

Then $A_0(\boldsymbol{x})$ play the role of Lagrange multipliers λ_a, and $C(\boldsymbol{x})$ are the constraints.

We define the phase space of Yang–Mills theory to be the following infinite-dimensional real vector space

$$\mathscr{M} = \{(\boldsymbol{A}(\boldsymbol{x}), E(\boldsymbol{x}))\} = \Omega_s^1(\mathbb{R}^3, \mathfrak{g}) \times \Omega_s^1(\mathbb{R}^3, \mathfrak{g}),$$

where $\Omega^1_s(\mathbb{R}^3,\mathfrak{g})$ stands for the space of one-forms on $\mathbb{R}^3$ with all components being Schwartz class functions.

To introduce coordinates on the phase space, we choose a basis $X_a \in \mathfrak{g}$, where $a = 1,\dots,r = \dim\mathfrak{g}$, so that every $u \in \mathfrak{g}$ can be wrriten as

$$u = u^a X_a.$$

Denote by K_{ab}, t^c_{ab} the structure constants of the bilinear form $\langle\ ,\ \rangle$ and the Lie bracket in this basis:

$$\begin{aligned}\langle u,v\rangle &= K_{ab}\,u^a v^b,\\ [u,v] &= t^c_{ab}\,u^a v^b X_c.\end{aligned} \tag{18.26}$$

Then the scalar-valued functions $E^a_i(\boldsymbol{x})$, $A^a_i(\boldsymbol{x})$, $\boldsymbol{x}\in\mathbb{R}^3$, $a = 1,\dots,\dim\mathfrak{g}$, play the role of coordinates on the phase space $\mathscr{M}$.

The symplectic form Ω on $\mathscr{M}$ is

$$\Omega = K_{ab}\int_{\mathbb{R}^3}\Big(\delta E^a_i(\boldsymbol{x})\wedge\delta A^b_i(\boldsymbol{x})\Big)\,d^3\boldsymbol{x}, \tag{18.27}$$

and gives rise to the following Poisson brackets on $\mathscr{M}$:

$$\begin{aligned}&\{E^a_i(\boldsymbol{x}),A^b_j(\boldsymbol{y})\} = \delta_{ij}K^{ab}\delta(\boldsymbol{x}-\boldsymbol{y}),\quad i,j=1,2,3 \quad\text{and}\quad a,b=1,\dots,\dim\mathfrak{g},\\ &\{E^a_i(\boldsymbol{x}),E^b_j(\boldsymbol{y})\} = \{A^a_i(\boldsymbol{x}),A^b_j(\boldsymbol{y})\} = 0.\end{aligned} \tag{18.28}$$

The Euler–Lagrange equations (6.14)–(6.16) in this case become

$$\begin{aligned}\partial_0 E_i(\boldsymbol{x}) &= \{H - C, E_i(\boldsymbol{x})\},\quad C = \int\langle A_0(\boldsymbol{x}),C(\boldsymbol{x})\rangle d^3\boldsymbol{x},\\ \partial_0 A_i(\boldsymbol{x}) &= \{H - C, A_i(\boldsymbol{x})\},\\ C(\boldsymbol{x}) &= 0.\end{aligned} \tag{18.29}$$

Note that the equation $C(\boldsymbol{x}) = 0$ is equivalent to $d^*_{\boldsymbol{A}}E = 0$, which is, as discussed in (18.11), one of equations of motion.

To write these equations explicitly, it is convenient to introduce the following notation. Let $f \in \mathscr{S}(\mathbb{R}^3,\mathfrak{g})$ be a test function and define

$$C_f = \int_{\mathbb{R}^3}\langle C(\boldsymbol{x}), f(\boldsymbol{x})\rangle d^3\boldsymbol{x}.$$

This is a scalar-valued function on the phase space. Clearly, conditions $C(\boldsymbol{x}) = 0$ for all $\boldsymbol{x}$ are equivalent to $C_f = 0$ for all f.

Lemma 18.7. *We have the following Poisson brackets on* $\mathscr{M}$*:*

$$\{C_f, E_i(\boldsymbol{x})\} = [f(\boldsymbol{x}), E_i(\boldsymbol{x})], \tag{18.30}$$

$$\{C_f, A_i(\boldsymbol{x})\} = [f(\boldsymbol{x}), A_i(\boldsymbol{x})] - \partial_i f(\boldsymbol{x}) \tag{18.31}$$

$$\{C_f, C_g\} = -C_{[f,g]}, \tag{18.32}$$

$$\{H, A_i(\boldsymbol{x})\} = E_i(\boldsymbol{x}), \tag{18.33}$$

$$\{H, E_i(\boldsymbol{x})\} = (d^*_{\boldsymbol{A}}B)_i, \tag{18.34}$$

$$\{H, C_f\} = 0. \tag{18.35}$$

where the operator $d_A^*\colon \Omega^k(\mathbb{R}^3, \mathfrak{g}) \to \Omega^{k-1}(\mathbb{R}^3, \mathfrak{g})$ is defined in Exercise 17.7. Note that some of these relations involve Poisson brackets of a scalar-valued function on $\mathscr{M}$ such as C_f and a $\mathfrak{g}$-valued function such as $E_i(\boldsymbol{x})$. We define such Poisson brackets in the obvious way: $\{E_i(\boldsymbol{x}), C_f\} = \{E_i^a(\boldsymbol{x}), C_f\}X_a$.

Proof. The straightforward proof of this lemma can be obtained by writing everything in terms of components, as commonly used by physicists. Namely we write $E(\boldsymbol{x}) = E_i^a(\boldsymbol{x})X_a dx^i$, $B(\boldsymbol{x}) = B_{ij}^a X_a dx^i \wedge dx^j$ and use canonical Poisson brackets (18.28) to compute Poisson brackets of B_{ij}^a with E_k^b, A_k^b and then use them to compute Poisson brackets of H with E, A, C_f (see Exercise 18.2). However, the computations are rather tedious and not very illuminating.

A more invariant approach is as follows. Assume that we have a (scalar-valued) observable, i.e. a function F on the phase space $\mathscr{M}$ such that

$$\delta F(\boldsymbol{A}, E) = \int_{\mathbb{R}^3} (\delta \boldsymbol{A}(\boldsymbol{x}), \varphi_1(\boldsymbol{x})) + (\delta E(\boldsymbol{x}), \varphi_2(\boldsymbol{x})) d^3\boldsymbol{x} \tag{18.36}$$

for some Schwartz class one-forms $\varphi_1, \varphi_2 \in \Omega_s^1(\mathbb{R}^3, \mathfrak{g})$. Here, as in Exercise 17.7, we define the pairing of k-forms on $\mathbb{R}^3$ with values in $\mathfrak{g}$ by combining the pairing on forms induced by the (positive definite) metric in $\mathbb{R}^3$ with the pairing $\langle\ ,\ \rangle$ on $\mathfrak{g}$. In such situation, we will say that

$$\frac{\delta F}{\delta \boldsymbol{A}(\boldsymbol{x})} = \varphi_1(\boldsymbol{x}), \quad \frac{\delta F}{\delta E(\boldsymbol{x})} = \varphi_2(\boldsymbol{x}).$$

Then it follows from canonical Poisson relations (18.28) that for observables F, G, we have

$$\{F, G\} = \int_{\mathbb{R}^3} \left(\left(\frac{\delta F}{\delta E(\boldsymbol{x})}, \frac{\delta G}{\delta \boldsymbol{A}(\boldsymbol{x})}\right) - \left(\frac{\delta F}{\delta \boldsymbol{A}(\boldsymbol{x})}, \frac{\delta G}{\delta E(\boldsymbol{x})}\right) \right) d^3\boldsymbol{x}, \tag{18.37}$$

This formula is a straightforward generalization of the formula (13.4) for a scalar field theory:

$$\{F, G\}(\pi, \varphi) = \int_{\mathbb{R}^3} \left(\frac{\delta F}{\delta \pi(\boldsymbol{x})} \frac{\delta G}{\delta \varphi(\boldsymbol{x})} - \frac{\delta F}{\delta \varphi(\boldsymbol{x})} \frac{\delta G}{\delta \pi(\boldsymbol{x})} \right) d^3\boldsymbol{x}.$$

Now it is not difficult to check (we leave it to the reader, see Exercise 18.3) that

$$\begin{aligned}
&\frac{\delta C_f}{\delta \boldsymbol{A}(\boldsymbol{x})} = -[f, E(x)], &&\frac{\delta C_f}{\delta E(\boldsymbol{x})} = -d_A f(\boldsymbol{x}),\\
&\frac{\delta \|E\|^2}{\delta \boldsymbol{A}(\boldsymbol{x})} = 0, &&\frac{\delta \|E\|^2}{\delta E(\boldsymbol{x})} = 2E(\boldsymbol{x}),\\
&\frac{\delta \|B\|^2}{\delta \boldsymbol{A}(\boldsymbol{x})} = -2d_A^* B, &&\frac{\delta \|B\|^2}{\partial E(\boldsymbol{x})} = 0,
\end{aligned} \tag{18.38}$$

where

$$\|E\|^2 = \int_{\mathbb{R}^3} |E(\boldsymbol{x})|^2 d^3\boldsymbol{x}, \qquad \|B\|^2 = \int_{\mathbb{R}^3} |B(\boldsymbol{x})|^2 d^3\boldsymbol{x},$$

and $|E(\boldsymbol{x})|^2$ is given by (18.9), so that $H = \frac{1}{2}(\|E\|^2 + \|B\|^2)$. Combining these formulas with (18.37), we get the statement of the lemma. □

Using Lemma 18.7, one can show that the Euler-Lagrange equations (18.29) are equivalent to

$$\begin{aligned} E &= -dA_0 + \nabla_0^A \boldsymbol{A} = -dA_0 + \partial_0 \boldsymbol{A} + [A_0, \boldsymbol{A}], \\ d^*_{\boldsymbol{A}} E &= -\nabla_i^{\boldsymbol{A}} E_i = 0, \\ d^*_{\boldsymbol{A}} B &= \nabla_0^A \star E, \end{aligned}$$

where d is the exterior derivative in $\mathbb{R}^3$, $\star$ is the Hodge operator in $\mathbb{R}^3$, $\boldsymbol{A} = A_i dx^i$ is the restriction of connection A to $\mathbb{R}^3$, and B is given by $B = -F_{\boldsymbol{A}} = -d\boldsymbol{A} - \frac{1}{2}[\boldsymbol{A} \wedge \boldsymbol{A}]$.

The first equation and the formula for B are equivalent to the equation

$$F = dA + \frac{1}{2}[A \wedge A]$$

in $\mathbb{R}^{1,3}$, where $F = dx^0 \wedge E - B$, while the second pair is equivalent to the equation $d_A \star F = 0$ (see formula (18.11)). Thus, the Euler–Lagrange equations (18.29) for the first order Lagrangian (18.23) are equivalent to the equations of motion $d^*_A F_A = 0$ — which of course was to be expected.

Lemma 18.7 immediately implies that the Yang-Mills theory is a Hamiltonian theory with a first class constraints as defined in Section 6.3. Thus, we can restrict the equations of motion to

$$\mathcal{M}_0 = \{(\boldsymbol{E}, \boldsymbol{A}) \in \mathcal{M} \mid C(\boldsymbol{x}) = -d^*_{\boldsymbol{A}} E = 0\}.$$

As in the case of Maxwell's theory (see Section 16.4), the Poisson brackets with constraints C_f are exactly the infinitesimal gauge transformations (17.35). In other words, the flow of the Hamiltonian vector field on the phase space $\mathcal{M}$ defined by C_f is given by the gauge transformations

$$(\boldsymbol{E}, \boldsymbol{A}) \mapsto (g\boldsymbol{E}g^{-1}, g\boldsymbol{A}g^{-1} - dg\, g^{-1}), \quad g(\boldsymbol{x}) = \exp(sf(\boldsymbol{x})). \tag{18.39}$$

Thus, we get the following result.

Theorem 18.8. *Define the reduced phase space $\mathcal{M}_0^*$ as the set of gauge equivalence classes $(\boldsymbol{E}, \boldsymbol{A}) \in \mathcal{M}_0$ under the gauge transformations (18.39).*

(1) *The symplectic form (18.27) descends to $\mathcal{M}_0^*$ and defines on it a structure of infinite-dimensional symplectic manifold. Similarly, the Hamiltonian function H given by (18.25) also descends to $\mathcal{M}_0^*$.*

(2) *The Euler–Lagrange equations (18.29) on $\mathcal{M}_0^*$ coincide with the equations of motion of the Hamiltonian system $(\mathcal{M}_0^*, \Omega, H)$.*

As in the U(1) case, we can also introduce additional constraints, identifying the reduced phase space $\mathcal{M}_0^* = \mathcal{M}_0/(\text{gauge transformations})$ with a submanifold defined by the equations $C(\boldsymbol{x}) = 0$, $D(\boldsymbol{x}) = 0$. The common choice is using a 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

$$\{C^a(\boldsymbol{x}), D^b(\boldsymbol{y})\} = K^{ab} \frac{\partial^2}{\partial x^k \partial y^k} \delta(\boldsymbol{x} - \boldsymbol{y}) + t_c^{ab} A_k^c(\boldsymbol{x}) \frac{\partial}{\partial y^k} \delta(\boldsymbol{x} - \boldsymbol{y}). \tag{18.40}$$

Thus $M^{ab}(\boldsymbol{x},\boldsymbol{y}) = \{C^a(\boldsymbol{x}), D^b(\boldsymbol{y})\}$ is an integral kernel of the differential operator

$$M = -\Delta + \operatorname{ad} A_k(\boldsymbol{x})\partial_k, \tag{18.41}$$

on $L^2(\mathbb{R}^3, \mathfrak{g})$, where Δ is the Laplace operator of the invariant metric $\langle\ ,\ \rangle$ on $\mathfrak{g}$. 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.

Remark 18.9. The differential operator M plays an important role in the path integral approach to the quantum Yang-Mills theory, and is called *Faddeev–Popov operator*.

18.6. Self-duality equations

Yang–Mills theory and thus, equations of motion (18.14), can be defined for any pseudo-Riemannian manifold M. In particular, we can consider the case when M is Riemannian, though corresponding equations do not have natural physical interpretation. However, in case when M is a compact Riemannian 4-manifold, the absolute minima of the Yang-Mills action functional are described by rather simple first order nonlinear partial differential equations, called *self-dual* and *anti-self-dual* equations. These beautiful equations play a fundamental in the Donaldson's theory, the study of the topology of 4-dimensional manifolds, and are used in the quantization of non-abelian gauge theories.

From now on, we assume that M is a compact oriented Riemannian 4-manifold and P is a principal G-bundle on M. Moreover, we assume that G is a connected simple compact Lie group ($G = \mathrm{SU}(2)$ for Donaldson's theory). For a 2-form $\omega \in \Omega^2(M, \mathfrak{g}_P)$ we define

$$|\omega(\boldsymbol{x})|^2 = \frac{1}{2}\langle \omega_{ij}, \omega^{ij}\rangle,$$
$$\langle \omega \wedge \star\omega\rangle = |\omega(\boldsymbol{x})|^2\, d^4\boldsymbol{x},$$

where $\langle\ ,\ \rangle$ is the negative of the Cartan-Killing form on $\mathfrak{g}$, and $d^4\boldsymbol{x}$ is the volume form defined by the orientation and Riemannian metric. Note that $|\omega(\boldsymbol{x})|^2 \geq 0$, with equality only if $\omega(\boldsymbol{x}) = 0$.

We also define

$$\|\omega\|^2 = \int_M |\omega(\boldsymbol{x})|^2 d^4\boldsymbol{x} = \int_M \langle \omega \wedge \star\omega\rangle.$$

Note that $\|\cdot\|^2$ is a positive definite quadratic form on $\Omega^2(M, \mathfrak{g}_P)$: for any $\omega \in \Omega^2(M, \mathfrak{g}_P)$, we have $\|\omega\|^2 \geq 0$, with equality if and only if ω is zero.

The Lagrangian density (18.2) can be rewritten as

$$\mathscr{L}(A) = -\frac{1}{2}|F_A(\boldsymbol{x})|^2,$$

and it is convenient (as we will see below) to define the Yang-Mills action functional by

$$S(A) = -\frac{1}{2\pi}\int_M \mathscr{L}(A) d^4\boldsymbol{x} = \frac{1}{4\pi}\|F_A\|^2. \tag{18.42}$$

On a 4-dimensional Riemannian manifold M the Hodge operator $\star\colon \Omega^2(M) \to \Omega^2(M)$ satisfies $\star^2 = 1$, so we have the decomposition

$$\Omega^2(M) = \Omega^2_+(M) \oplus \Omega^2_-(M), \tag{18.43}$$

where $\Omega_\pm(M)$ are eigenspaces of Hodge $\star$-operator corresponding to the eigenvalues 1 and -1 respectively. The 2-form F on M is called *self-dual* or *anti-self-dual* if $F \in \Omega^2_+(M)$ or $F \in \Omega^2_-(M)$ respectively:

$$\star F = \pm F.$$

This definition trivially extends to forms with values in a vector bundle, in particular, to $F \in \Omega^2(M, \mathfrak{g}_P)$. We will call a connection A in a principal G-bundle P self-dual or anti-self-dual if its curvature 2-form F_A is self-dual or anti-self-dual respectively. Of course, anti-sef-dual connection can be obtained from the self-dual by changing the orientation of M. It follows from the Bianchi identity that self-dual and anti-self-dual connections automatically satisfy Yang-Mills equations

$$d_A \star F_A = 0.$$

Remark 18.10. In the pseudo-Riemannian case $\star^2 = -1$ on 2-forms, and analog of decomposition (18.43) is valid only for complex-valued 2-forms. Corresponding self-duality equations take the form

$$\star F = \pm i\, F$$

and have no solutions in $\Omega^2(M, \mathfrak{g})$. in other words, these equations have only "non-physical" solutions.

It trivially follows from the relation $\langle \alpha \wedge \star\beta \rangle = \langle \beta \wedge \star\alpha \rangle$ that if $\alpha \in \Omega^2_+(M, \mathfrak{g}_P)$, $\beta \in \Omega^2_-(M, \mathfrak{g}_P)$, then $\langle \alpha \wedge \star\beta \rangle = 0$, so

$$\langle (\alpha + \beta) \wedge \star(\alpha + \beta) \rangle = \langle \alpha \wedge \star\alpha \rangle + \langle \beta \wedge \star\beta \rangle. \tag{18.44}$$

Writing

$$F_A = F_+ + F_-,$$

where $F_\pm$ the self-dual and anti-self-dual components of the curvature 2-form F_A, we readily obtain

$$S(A) = \frac{1}{4\pi}\big(\|F_+\|^2 + \|F_-\|^2\big). \tag{18.45}$$

The vector bundle $\mathfrak{g}_P$ has a topological invariant, the first Pontryagin number (called the *instanton number* by physicists) — an integer k, given by evaluation of the first Pontryagin class $p_1(\mathfrak{g}_P)$ on the fundamental cycle $[M] \in H_4(M, \mathbb{Z})$. Using (17.45), definition (**??**) of the invariant metric on $\mathfrak{g}$, and representing $p_1(\mathfrak{g}_P)$ by a 4-form on M as in Example 17.3, we obtain

$$k = -\frac{1}{8\pi^2}\int_M \operatorname{tr}(F_A \wedge F_A) = \frac{1}{8\pi^2}\int_M \langle F_A \wedge F_A \rangle.$$

We also have

$$\langle (\alpha + \beta) \wedge (\alpha + \beta) \rangle = \langle (\alpha + \beta) \wedge \star(\alpha - \beta) \rangle = \langle \alpha \wedge \star\alpha \rangle - \langle \beta \wedge \star\beta \rangle,$$

so

$$k = \frac{1}{8\pi^2}\big(\|F_+\|^2 - \|F_-\|^2\big). \tag{18.46}$$

Theorem 18.11. *Let P be a principal G-bundle over on a connected compact oriented 4-manifold M. If $k > 0$, then absolute minima of the Yang-Mills action functional are given by the self-dual connections, and if $k < 0$ — by the anti-self-dual connections.*

Proof. From (18.45)–(18.46) we trivially have

$$S(A) - 2\pi k \geq \frac{1}{4\pi}\|F_-\|^2 \quad \text{and} \quad S(A) + 2\pi k \geq \frac{1}{4\pi}\|F_+\|^2.$$

Thus $S(A) \geq 2\pi|k|$ and for $k > 0$ the minimum is attain at the self-dual connections, for $k < 0$ — at the anti-self-dual connections, and for $k = 0$ — at the flat connections. □

The existence of such connections is a highly non-trivial analytic problem, and we refer the reader to a special literature. In case $M = S^4$ and $G = \mathrm{SU}(2)$ it can be shown that self-dual connections exist for $k \geq 1$ and depend on $8k-3$ parameters, forming the so-called moduli space of instantons.

18.7. Hitchin's equations

Let G be a compact real form of a complex Lie group, and denote by $*$ the 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 $\mathfrak{g}_P$ over $\mathbb{R}^4$ with the Euclidean metric. A connection $A = A_\mu d^\mu$ is a $\mathfrak{g}$-valued 1-form 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

$$F_{12} = F_{34}, \quad F_{13} = F_{42}, \quad F_{14} = F_{23}.$$

Now consider a dimensional reduction of the self-duality equations: suppose that A_μ do not depend on 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

$$\begin{aligned} F_{12} &= [\phi_1, \phi_2] = F_{24}, \\ F_{13} &= [\nabla_1, \phi_1] = [\phi_2, \nabla_1] = F_{42}, \\ F_{14} &= [\nabla_1, \phi_2] = [\nabla_2, \phi_1] = F_{23}. \end{aligned}$$

Denote by

$$F = F_{12} = \partial_1 A_2 - \partial_2 A_1 + [A_1, A_2]$$

the curvature form of a connection $A_1 dx^1 + A_2 dx^2$ on a trivial $\mathfrak{g}_P$ bundle over $\mathbb{R}^2$. Introducing the complex Higgs field $\phi = \phi_1 - i\phi_2$, the above equations can be written as

$$F = \frac{i}{2}[\phi, \phi^*] \quad \text{and} \quad [\nabla_1 + i\nabla_2, \phi] = 0. \tag{18.47}$$

Put $z = x^1 + ix^2$ and introduce a 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 $\mathfrak{g}_P^{\mathbb{C}} = \mathfrak{g}_P \otimes_{\mathbb{R}} \mathbb{C}$ over $\mathbb{C} \simeq \mathbb{R}^2$ — a complexification of a real vector bundle $\mathfrak{g}_P$. Denoting

$$\Phi = \tfrac{1}{2}\phi\, dz \in \Omega^{1,0}(\mathbb{C}, \mathfrak{g}_P^{\mathbb{C}}), \quad \Phi^* = \tfrac{1}{2}\phi^*\, d\bar{z} \in \Omega^{0,1}(\mathbb{C}, \mathfrak{g}_P^{\mathbb{C}}),$$

we can rewrite equations (18.47) as

$$\begin{aligned} F + [\Phi, \Phi^*] &= 0, \\ \bar{\partial}_A \Phi &= 0. \end{aligned} \tag{18.48}$$

Here

$$[\Phi, \Phi^*] = \Phi \wedge \Phi^* + \Phi^* \wedge \Phi$$

is a graded Lie bracket on $\mathfrak{g}_P^{\mathbb{C}}$-valued 1-forms, and $\bar{\partial}_A$ is a $(0,1)$-component of

$$d_A = \partial + A^{1,0}dz + \bar{\partial} + A^{0,1}d\bar{z} = \partial_A + \bar{\partial}_A.$$

It is remarkable that equations (18.48) make sense over an arbitrary Riemann surface M! Namely, consider a principal G-bundle P over M, a connection A in the adjoint bundle $\mathfrak{g}_P$ and the Higgs field $\Phi \in \Omega^{1,0}(M, \mathfrak{g}_P^{\mathbb{C}})$. The pair (A, Φ) satisfies Hitchin's equations over a Riemann surface M, if

$$F(A) + [\Phi, \Phi^*] = 0 \quad \text{and} \quad \bar{\partial}_A \Phi = 0. \tag{18.49}$$

The second equation states that Φ is a holomorphic section of the complex vector bundle $\mathfrak{g}_P \otimes \Omega^{1,0}(M, \mathbb{C})$ with respect to the complex structure in $\mathfrak{g}_P^{\mathbb{C}}$, 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 Hitchin's equations (18.49) determines a flat connection $d + A + \Phi + \Phi^*$ on $\mathfrak{g}_P^{\mathbb{C}}$.

18.8. Exercises

Exercise 18.1. Derive equations of motion for the Lagrangians (18.19) and (18.21).

Exercise 18.2. Use the canonical Poisson relations (18.28) to derive the following Poisson brackets:

$$\{E_i^a(\boldsymbol{x}), C^b(\boldsymbol{y})\} = -\delta(\boldsymbol{x} - \boldsymbol{y}) t_c^{ab} E_i^c,$$
$$\{A_i^a(\boldsymbol{x}), C^b(\boldsymbol{y})\} = -\big(t_c^{ab} A_i^c(\boldsymbol{x}) + K^{ab}\partial_{y^i}\big)\delta(\boldsymbol{x} - \boldsymbol{y}).$$

Exercise 18.3. Complete the proof of Lemma 18.7.

Exercise 18.4. The 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 Hamiltonian and find the corresponding moment map.

Exercise 18.5. For $M = S^4$ and $G = \mathrm{SU}(2)$ find a self-dual connection with $k = 1$ (this is the famous instanton solution of the Yang-Mills equations).

Exercise 18.6. Prove formula (18.15).

Chapter 19

Chern–Simons Theory

In this chapter, we discuss one more model of a classical field theory, the Chern–Simons theory in dimension 3. As in Yang–Mills theory, fields in Chern–Simons theory are connections in a G-bundle, where G is a simple compact Lie group such as $\mathrm{SU}(n)$. The most important feature of this theory is that unlike Yang–Mills Lagrangian, the Lagrangian of Chern–Simons theory doesn't depend on the metric of spacetime; such theories are called *topological*.

Throughout this chapter, M is an oriented 3-dimensional manifold which will play the role of spacetime of the theory; unless specified otherwise, we assume that M is connected and compact to avoid problems with convergence of integrals. We also fix a connected simply-connected simple compact Lie group G; our main example is the group $G = \mathrm{SU}(n)$, but most features can already be studied for $G = \mathrm{SU}(2)$. We will denote by $\mathfrak{g}$ the Lie algebra of G.

19.1. Topology of simple Lie groups

Before defining the Chern–Simons theory, we need some information about topology of Lie group G.

Theorem 19.1. *Let G be a connected simply-connected simple Lie group. Then the homotopy groups of G are*

$$\pi_1(G) = \pi_2(G) = \{1\}, \quad \pi_3(G) = \mathbb{Z}.$$

This immediately implies that

$$H_3(G, \mathbb{Z}) = \mathbb{Z} \tag{19.1}$$

and thus, $H^3(G, \mathbb{R}) = \mathbb{R}$. Moreover, we can easily describe the generator of $H^3(G, \mathbb{R})$.

Let $\langle\ ,\ \rangle$ be a positive definite G-invariant symmetric bilinear form on the Lie algebra $\mathfrak{g}$; as discussed in Chapter 18, such a form exists and is unique up to a scalar factor. We will fix the normalization shortly.

Lemma 19.2. *Let $\lambda \in \Omega^3(G)$ be defined by*

(1) λ *is left-invariant*

(2) *For* $a, b, c \in \mathfrak{g} = T_eG$, *we have*

$$\lambda(a, b, c) = \langle a, [b, c]\rangle. \tag{19.2}$$

Then λ *is closed and* $H^3(G, \mathbb{R}) = \mathbb{R}[\lambda]$.

From now on, let us fix the normalization of the bilinear form $\langle\ ,\ \rangle$ by the following condition:

$$\int_C \lambda = 2\pi, \qquad C - \text{generator of } H_3(G, \mathbb{Z}). \tag{19.3}$$

Note that there are two choices of a generator of $H_3(G, \mathbb{Z}) \simeq \mathbb{Z}$; we fix the choice of C by the condition that $\int_C \lambda > 0$ for positive definite form $\langle\ ,\ \rangle$.

Example 19.1. Let $G = \mathrm{SU}(2)$. Then there is an embedding $\imath\colon \mathrm{SU}(2) \to \mathbb{R}^4$ which identifies $\mathrm{SU}(2) \simeq S^3 \subset \mathbb{R}^4$ (see Exercise 19.1). Moreover, an explicit computation given in Exercise 19.2 shows that in order to satisfy normalization condition (19.3), the bilinear form $\langle\ ,\ \rangle$ should be taken to be

$$\langle a, b\rangle = \frac{1}{2\pi}(\imath_* a, \imath_* b)_{\mathbb{R}^4} = -\frac{1}{4\pi}\,\mathrm{tr}(ab), \qquad a, b \in \mathfrak{su}(2), \tag{19.4}$$

and thus,

$$\lambda(a, b, c) = \langle a, [b, c]\rangle = -\frac{1}{4\pi}\,\mathrm{tr}(a[b, c]). \tag{19.5}$$

Moreover, the same formula holds for $G = \mathrm{SU}(n)$ (see Exercise 19.3).

These results can be reformulated in terms of the Maurer–Cartan form. Recall that the Maurer–Cartan form $\theta \in \Omega^1(G, \mathfrak{g})$ is the left-invariant form such that for $a \in T_eG = \mathfrak{g}$, we have $\theta(a) = a$, see (17.28); for matrix groups, we have $\theta = g^{-1}dg$.

More generally, given a manifold M and a smooth function $g\colon M \to G$, we define

$$\theta_g = g^*\theta \in \Omega^1(M, \mathfrak{g}). \tag{19.6}$$

As in (18.1), we use notation $\langle \alpha \wedge \beta\rangle$ for the $\mathfrak{g}$-valued differential forms. It immediately follows from the definition of the Maurer–Cartan form that $\langle \theta \wedge \theta\rangle = 0$ and for $a, b, c \in T_eG = \mathfrak{g}$,

$$\langle \theta \wedge [\theta \wedge \theta]\rangle(a, b, c) = 6\langle a, [b, c]\rangle. \tag{19.7}$$

We have the following simple result.

Lemma 19.3. *Let* $\theta \in \Omega^1(G, \mathfrak{g})$ *be the Maurer–Cartan form and let* $C \in H_3(G, \mathbb{Z})$ *be the generator of* $H_3(G, \mathbb{Z})$ *as in* (19.3). *Then*

$$\int_C \langle \theta \wedge [\theta \wedge \theta]\rangle = 12\pi.$$

For $G = \mathrm{SU}(n)$ *this can also be rewritten as*

$$\int_C \mathrm{tr}(\theta \wedge \theta \wedge \theta) = \frac{1}{2}\int_C \mathrm{tr}(\theta \wedge [\theta \wedge \theta]) = -24\pi^2.$$

Proof. This immediately follows from (19.3), since by (19.7) we have $\langle \theta \wedge [\theta \wedge \theta]\rangle = 6\lambda$. □

19.2. Chern–Simons action

After these preliminaries, we can now define the Chern–Simons theory. Let M be a compact oriented 3-manifold without boundary, and let P be a principal G-bundle on M. As in Section 17.4, we denote by $\mathcal{A}(P)$ the infinite-dimensional affine space. of connections in P. The space of fields of Chern–Simons theory is defined to be the space of gauge equivalence classes of connections:

$$\mathcal{F} = \mathcal{A}(P)/\mathcal{G}(P) \tag{19.8}$$

where $\mathcal{G}(P)$ is the group of gauges transformations as defined in Section 17.5.

Before proceeding, we make the following observation.

Lemma 19.4. *Any G-bundle on a 3-manifold M is topologically trivial.*

Proof. For readers familiar with the theory of classifying spaces, the simplest proof is as follows. As is well known, the isomorphism classes of G-bundles on a manifold M are classified by the homotopy classes of maps $M \to BG$. On the other hand, vanishing of fundamental groups $\pi_1(G)$, $\pi_2(G)$ (see Theorem 19.1) implies that $\pi_k(BG) = \{1\}$ for $k \le 3$. By standard results of algebraic topology, this implies that any map of a 3-manifold to BG is homotopic to a constant map. □

In other words, any G-bundle P on M can be trivialized: we can choose a trivialization $\varphi\colon P \to M \times G$. However, such a trivialization is not unique, which will play important role in what follows. In particular, this implies that the gauge group $\mathcal{G}(P)$ is isomorphic to $C^\infty(M, G)$: any two trivializations are related by

$$\varphi^\alpha(p) = g(\pi(p))\varphi^\beta(p)$$

for some function $g\colon M \to G$.

Let us assume that we have chosen a trivialization φ. Then a connection can be described by a 1-form $A \in \Omega^1(M, \mathfrak{g})$. We define the *Chern–Simons action* by

$$S_{CS}[A] = -\int_M \omega_{CS}(A), \tag{19.9}$$

where

$$\omega_{CS} = \langle A \wedge dA\rangle + \tfrac{1}{3}\langle A \wedge [A \wedge A]\rangle \in \Omega^3(M) \tag{19.10}$$

is the *Chern-Simons form.* For $G = \mathrm{SU}(n)$, by results of Example 19.1, the action can be rewritten as

$$S_{CS}[A] = \frac{1}{4\pi}\int_M \mathrm{tr}\Big(A \wedge dA + \tfrac{2}{3}A \wedge A \wedge A\Big) = \frac{1}{4\pi}\int_M \mathrm{tr}(A \wedge dA + \tfrac{1}{3}A \wedge [A \wedge A]). \tag{19.11}$$

We will provide motivation for this definition later.

The first question we need to answer is whether the Chern–Simons action is independent of the trivialization of the bundle, or, equivalently, whether the Chern–Simons action is gauge equivalent. Recall that by Theorem 17.11, under a gauge transformation $g\colon M \to G$ the one-form of the connection transforms as

$$A \mapsto A^g = g^{-1}Ag + g^{-1}dg$$

(for technical reasons, here we use the right action of the gauge group on connections rather than the left action).

Thus, in order for the Chern–Simons action to be independent of trivialization we need to check whether $S_{CS}[A^g] = S_{CS}[A]$.

Theorem 19.5. *Let M be compact. Then*

$$S_{CS}[A^g] - S_{CS}[A] = 2\pi d,$$

where $d \in \mathbb{Z}$ is defined by $[g(M)] = d[C] \in H_3(G, \mathbb{Z})$, and C is the generator of $H_3(G, \mathbb{Z})$ as in (19.3).

Proof. Let $\omega_{CS}(A) \in \Omega^3(M)$ be defined by (19.9). Then using the invariance of the form $\langle\ ,\ \rangle$ and the Maurer–Cartan equation $d\theta + \frac{1}{2}[\theta \wedge \theta] = 0$ (see Exercise 17.5), one can show by a rather lengthy computation that

$$\omega_{CS}(A^g) - \omega_{CS}(A) = -\tfrac{1}{6}\langle\theta_g \wedge [\theta_g \wedge \theta_g]\rangle - d\langle\theta_g \wedge g^{-1}Ag\rangle. \tag{19.12}$$

Details are left to the reader in Exercise 19.5. Since for closed M the integral of exact form is zero, by Lemma 19.3 we have

$$\begin{aligned} S_{CS}(A^g) - S_{CS}(A) &= \frac{1}{6}\int_M \langle\theta_g \wedge [\theta_g \wedge \theta_g]\rangle \\ &= \frac{1}{6}\int_{g(M)} \langle\theta \wedge [\theta \wedge \theta]\rangle \\ &= 2\pi d. \end{aligned}$$

□

Thus, we see that the Chern–Simons action is not invariant under the group $\mathcal{G}(P)$ of gauge transformations. However, it is invariant under the normal subgroup $\mathcal{G}(P)_0$ of gauge transformations homotopic to identity. Under the general gauge transformations, we can only claim that S_{CS} is well-defined as an element of $\mathbb{R}/2\pi\mathbb{Z}$. In particular, this implies that $\exp(iS_{CS})$ is well-defined.

19.3. Chern–Simons form as a secondary characteristic class

The fact that Chern–Simons action changes by an integer multiple of 2π under the gauge transformations can be described in a different (but closely related) way, in terms of integration over 4-manifolds.

We begin by recalling the following classical result of low-dimensional topology.

Theorem 19.6. *Any oriented compact 3-manifold M without boundary can be written as a boundary of some 4-manifold: $M = \partial X$, for some (not unique) 4-manifold X.*

Let now P be a G bundle on M. By Lemma 19.4, P can be trivialized. Let us choose a trivialization; it identifies P with the trivial bundle and thus allows us to extend P from M to a bundle $\tilde{P}$ on X. Moreover, it is easy to show that a connection A on P can be extended to a connection in $\tilde{P}$.

Lemma 19.7. *Let X be a 4-manifold (possibly with boundary). For $A \in \Omega^1(X, \mathfrak{g})$, let $\omega_{CS}(A) \in \Omega^3(X)$ be defined by (19.9). Then*

$$d\omega_{CS}(A) = \langle F_A \wedge F_A \rangle \in \Omega^4(X),$$

where $F_A = dA + \frac{1}{2}[A \wedge A] \in \Omega^2(X, \mathfrak{g})$ is the curvature of A. For $\mathrm{SU}(n)$, *this becomes*

$$d\omega_{CS}(A) = -\frac{1}{4\pi}\,\mathrm{tr}(F_A \wedge F_A).$$

Proof. Using (19.10) and Ad-invariance of the bilinear form $\langle\ ,\ \rangle$, we readily obtain

$$d\omega_{cs}(A) = \langle dA \wedge dA \rangle + \langle dA \wedge [A \wedge A] \rangle.$$

Now the result follows immediately from the following observation:

$$\langle [A \wedge A] \wedge [A \wedge A] \rangle = \langle A \wedge [A \wedge [A \wedge A]] \rangle = 0.$$

Here the first equality is the invariance of $\langle\ ,\ \rangle$, and the second equality holds since by the Jacobi identity $[A \wedge [A \wedge A]] = 0$. □

Corollary 19.8. *Let M be a 3-manifold without boundary, with a G-bundle P and a connection in P. Choose a trivialization $\varphi \colon P \to M \times G$ of P and let A be the one-form of the connection in this trivialization.*

Let X be a 4-manifold with boundary such that $\partial X = M$; then trivialization $P \simeq M \times G$ allows one to extend P to a bundle $\tilde{P} \simeq X \times G$ on X. Let $\tilde{A}$ be an extension of A to $\tilde{P}$. Then

$$S_{CS}[A] = -\int_M \omega_{CS}(A) = -\int_X \langle F_{\tilde{A}} \wedge F_{\tilde{A}} \rangle.$$

Note that the right-hand side depends on the choice of X, on the trivialization (which is used to extend P to X) and on the extension of a connection to X.

This gives another way to explain why the $S_{CS}[A]$ is well-defined up to an integer multiple of 2π, at least for $G = \mathrm{SU}(n)$. Let X_1, X_2 be two 4-manifolds such that $\partial X_1 = \partial X_2 = M$. Given a G bundle P on M with a connection, choose two trivializations φ_1, φ_2 of P; let A_1, A_2 be the one-forms of the connection in each of these trivializations.

Using trivialization φ_1, we can extend P to a (trivial) bundle $\tilde{P}_1 \simeq X_1 \times G$ on X_1, and choose an extension $\tilde{A}_1$ of A_1 to X_1. Similarly we can use φ_2 to extend the connection to X_2.

Consider now the manifold $X = X_1 \cup (-X_2)$, obtained by gluing X_1, X_2 along the common boundary M (the minus sign indicates that we consider X_2 with the opposite orientation). Then X is a smooth manifold without boundary, and the bundles $\tilde{P}_1$, $\tilde{P}_2$ can be glued together to give a G-bundle $\tilde{P}$ on X (which might not be trivial, as the trivializations φ_1 and φ_2 might be different), with a connection $\tilde{A}$.[1]

Then Corollary 19.8 implies

$$S_{CS}[A_1] - S_{CS}[A_2] = \frac{1}{4\pi}\int_X \mathrm{tr}(F_{\tilde{A}} \wedge F_{\tilde{A}}).$$

[1] One needs to be careful so that the gluing gives a smooth connection on X; it can be achieved by making a suitable gauge transformation in a neigborhood of the boundary. For details, we refer the reader to paper by D. Freed cited in Chapter 20.

We can now use the result of Example 17.3, which shows that for $G = \mathrm{SU}(n)$ the right-hand side can be written as

$$\frac{1}{4\pi}\int_X \mathrm{tr}(F_{\tilde{A}} \wedge F_{\tilde{A}}) = 2\pi c_2(\tilde{P}).$$

Since $c_2(\tilde{P}) \in H^4(X, \mathbb{Z}) \simeq \mathbb{Z}$ when X is connected, we conclude that

$$S_{CS}[A_1] - S_{CS}[A_2] \in 2\pi\mathbb{Z}$$

(compare with Theorem 19.5).

19.4. Equations of motion in the Chern–Simons theory

Recall that for any classical field theory, the classical equations of motion are given by the critical points $\delta S[\varphi] = 0$ of the action functional S. Let us apply it the Chern–Simons theory.

Lemma 19.9. *Using standard notation*

$$\delta S_{CS}[A] = \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} S[A + \varepsilon \cdot \delta A],$$

we have the following result

$$\delta S_{CS}[A] = 2\int_M \langle F_A \wedge \delta A\rangle.$$

Proof. Using (19.9), Stokes theorem and Ad-invariance of the bilinear form $\langle\ ,\ \rangle$, we obtain

$$\begin{aligned}\delta S_{CS}[A] = -\int_M \bigg(&\langle \delta A \wedge dA\rangle + \langle A \wedge d\delta A\rangle + \frac{1}{3}\langle \delta A \wedge [A \wedge A]\rangle \\ &+ \frac{1}{3}\langle A \wedge [\delta A \wedge A]\rangle + \frac{1}{3}\langle A \wedge [A \wedge \delta A]\rangle\bigg) \\ = -\int_M &\langle \delta A \wedge (2dA + [A \wedge A])\rangle \\ = 2\int_M &\langle F_A \wedge \delta A\rangle.\end{aligned}$$

□

This result can be interpreted as follows. Let $\mathcal{A}$ be the space of all connections in the trivial G-bundle on M; as discussed in Theorem 17.9, this is an infinite-dimensional affine space associated with the vector space $\Omega^1(M, \mathfrak{g})$, so at any point $A \in \mathcal{A}$, the tangent space is $T_A\mathcal{A} = \Omega^1(M, \mathfrak{g})$. Since we have a non-degenerate pairing

$$\Omega^2(M, \mathfrak{g}) \otimes \Omega^1(M, \mathfrak{g}) \to \mathbb{R}$$

$$\eta \otimes \omega \mapsto \int_M \langle \eta \wedge \omega\rangle,$$

we can think of $\Omega^2(M, \mathfrak{g})$ as the cotangent space: $T^*_A\mathcal{A} := \Omega^2(M, \mathfrak{g})$.

Then the curvature of a connection admits a natural interpretation as 1-form on $\mathcal{A}$, since to every connection $A \in \mathcal{A}$ it assigns its curvature $F_A \in \Omega^2(M, \mathfrak{g}) \simeq T^*_A\mathcal{A}$. From this point of view, Lemma 19.9 is interpreted as the identity

$$\mathrm{d}\, S_{CS}[A] = 2F_A \in \Omega^1(\mathcal{A})$$

of 1-forms on $\mathcal{A}$, where d is the exterior differential on $\mathcal{A}$.

Lemma 19.9 immediately implies the following result.

Theorem 19.10. *Critical points of Chern–Simons action are the flat connections on M:*

$$\delta S_{CS}[A] = 0 \iff F_A = 0.$$

Remark 19.11. Note that $F_A = 0$ is a first–order differential equation in components of A, and not of a second order as one would normally expect. This makes Chern–Simons field theory very different from more common field theories such as Klein–Gordon or Yang–Mills. In those theories, if our spacetime is of the form $N \times \mathbb{R}$, and we fix the initial conditions $\varphi|_{t=0}$, $\dot{\varphi}|_{t=0}$, then there is a unique solution of equations of motion with these conditions. For Chern–Simons it is not so: if $M = N \times \mathbb{R}$, then in order for a connection on N to be extendable to a solution of equations of motion, it is necessary that it is flat. Moreover, if this is the case, a flat extension is not unique.

Let us now consider the set of gauge equivalence classes of flat connections:

$$\{A \in \mathcal{A}(P) \mid F_A = 0\}/\mathcal{G}(P).$$

Assuming that M is connected, it is well-known that we have a bijection

$$\{A \in \mathcal{A}(P) \mid F_A = 0\}/\mathcal{G}(P) \simeq \mathrm{Hom}(\pi_1(M), G)/G$$

given by the holonomy of a connection (see discussion in Section 17).

Since $\pi_1(M)$ is finitely generated, space $X = \mathrm{Hom}(\pi_1(M), G)/G$ is finite-dimensional. In general, it is not a smooth manifold (it is a stratified space), so usually one considers one of the approaches below to replace it by a better behaved space.

(1) We can consider *simple connections*, which are characterized by the property that the centralizer of the image of $\pi_1(M)$ in G is finite. It is known that such connections form an open dense subset $R^0 \subset \mathrm{Hom}(\pi_1(M), G)$, and the quotient $X^0 = R^0/G$ is a smooth finite-dimensional manifold (but non-compact).

(2) Alternatively, one can replace (real) Lie group G by a complex algebraic group $G^{\mathbb{C}}$ (e.g., replace $\mathrm{SU}(n)$ by $\mathrm{SL}(n, \mathbb{C})$). Then the set $\mathrm{Hom}(\pi_1(M), G^{\mathbb{C}})$ is an affine algebraic variety, called the *representation variety* of M, and we can consider the geometric invariant theory quotient

$$X(M) = \mathrm{Hom}(\pi_1(M), G^{\mathbb{C}})//G^{\mathbb{C}}.$$

This is again an affine algebraic variety (generally, not smooth), called the *character variety* of M.

We will say more about the structure of $\mathrm{Hom}(\pi_1(M), G)/G$ in the case $M = \mathbb{R} \times N$ in Section 19.6.

19.5. Chern–Simons theory for manifolds with boundary

So far, we have studied Chern–Simons theory on a closed 3-manifold. Let us now consider whan happens if we allow manifolds with boundary.

Let M be an oriented compact 3-manifold with boundary: $\partial M = N$; as before, let P be a G-bundle on M and let A be a connection in P. Choose a trivialization of P and define the Chern–Simons action $S_{CS}(A)$ by the same formulas (19.9)–(19.10).

As before, so defined action depends on the choice of trivialization. In addition, we get an extra term coming from the boundary: it follows from (19.12) that

$$S_{CS}(A^g) - S_{CS}(A) = \int_{\partial M} \langle \theta_g \wedge g^{-1}Ag\rangle + \frac{1}{6}\int_M \langle \theta_g \wedge [\theta_g \wedge \theta_g]\rangle, \tag{19.13}$$

so we can not claim that $S_{CS}(A)$ is a well-defined real number modulo 2π. Instead, we will show that for every choice of a connection A on a 2-manifold $N = \partial M$, we can define a one-dimensional vector space $L_N(A)$ so that $e^{iS_{CS}(A)}$ is well-defined as an element in $L_N(A)$.

Recall that every compact oriented 2-manifold N without boundary is a boundary of some (not unique) 3-manifold M. Moreover, it is easy to show using Theorem 19.1, that every function $g\colon N \to G$ can be extended to a function $\tilde{g}\colon M \to G$.

Lemma 19.12. *For a compact oriented 2-manifold N and a function $g\colon N \to G$, define*

$$W_N(g) = \frac{1}{6}\int_M \langle \theta_{\tilde{g}} \wedge [\theta_{\tilde{g}} \wedge \theta_{\tilde{g}}]\rangle$$

where M is a compact 3-manifold with boundary such that $\partial M = N$ and $\tilde{g}\colon M \to G$ is an extension of $g\colon N \to G$. Then $W_N(g) \mod 2\pi$ only depends on N and doesn't depend on the choice of 3-manifold M and extension $\tilde{g}$.

Proof. Let M_1, M_2 be two 3-manifolds such that $\partial M_1 = \partial M_2 = N$. As in Section 19.3, form a new manifold $M = M_1 \cup (-M_2)$, obtained by gluing M_1 and M_2 (with reversed orientation) along N. Then

$$\frac{1}{6}\int_{M_1} \langle \theta_g \wedge [\theta_g \wedge \theta_g]\rangle - \frac{1}{6}\int_{M_2} \langle \theta_g \wedge [\theta_g \wedge \theta_g]\rangle = \frac{1}{6}\int_{M} \langle \theta_g \wedge [\theta_g \wedge \theta_g]\rangle \in 2\pi\mathbb{Z}.$$

□

Remark 19.13. The functional $W_N(g)$ is a two-dimensional analog of the Wess-Zumino term in the theory of quantum anomalies. It plays a fundamental role in two dimensional conformal field theory and is called the *Wess–Zumino–Novikov–Witten (WZNW) action.*

With the help of Lemma 19.12 we can rewrite (19.13) in the following form:

$$\begin{gathered} S_{CS}(A^g) - S_{CS}(A) = C(A,g) \mod 2\pi \\ C(A,g) = W_N(g) + \int_N \langle \theta_g \wedge g^{-1}Ag\rangle. \end{gathered} \tag{19.14}$$

Note that $C(A,g) \mod 2\pi$ only depends on restrictions of a one-form A and function g to the boundary $N = \partial M$.

Corollary 19.14. *For any $A \in \mathcal{A}_N$, $g_1, g_2 \in \mathcal{G}$, we have*

$$C(A, g_1g_2) = C(A, g_1) + C(A^{g_1}, g_2) \mod 2\pi. \tag{19.15}$$

Proof. Indeed, using $A^{g_1g_2} = (A^{g_1})^{g_2}$ and (19.14), we have

$$S_{CS}(A^{g_1g_2}) - S_{CS}(A) = S_{CS}((A^{g_1})^{g_2}) - S_{CS}(A^{g_1}) + S_{CS}(A^{g_1}) - S_{CS}(A). \quad □$$

Equation (19.15) admits a simple interpretation in terms of group cohomology. Recall that given a representation V of a group G, the cohomology $H^i(G, V)$ are defined by $H^i(G, V) = \mathrm{Ext}^i(\mathbb{C}, V)$, where Ext is considered in the category of representations of G (or, equivalently, $\mathbb{C}[G]$-modules). These cohomology can be computed explicitly using the bar resolution:

$$C^0(G, V) \to C^1(G, V) \to C^2(G, V) \to \dots$$
$$C^k(G, V) = \{f \colon G^k \to V\}.$$

In particular $C^1(G, V) = \{f \colon G \to V\}$, and the differential $d \colon C^1 \to C^2$ is given by

$$df(g_1, g_2) = g_1 f(g_2) - f(g_1 g_2) + f(g_1)$$

(the general formula for the differential can be found in standard textobooks on homological algebra). Thus, the space of 1-cocycles is given by

$$Z^1(G, V) = \{f \colon G \to V \mid f(g_1 g_2) = f(g_1) + g_1 f(g_2)\}.$$

Let us now apply this to the group $\mathcal{G} = C^\infty(N, G)$ of gauge transformations, and take $V = \Omega^0(\mathcal{A}_N)$ be the space of complex-valued functionals $\Psi \colon \mathcal{A}_N \to \mathbb{C}$ ("wave functions" in the physics terminology). It is a left $\mathcal{G}$-module with the action

$$(g \cdot \Psi)(A) = \Psi(A^g)$$

where as before, $A^g = g^{-1}Ag + g^{-1}dg$. Considering $C(A, g)$ as an element in $\Omega^0(\mathcal{A}_N)$, we immediately see that (19.15) is equivalent to $dC = 0 \mod 2\pi$; in other words, C is a 1-cocycle of group $\mathcal{G}$ with values in $\Omega^0(\mathcal{A}_N) \mod 2\pi$.

Remark 19.15. In physics terminology, the term $C(A, g)$ is called "quantum anomaly". The 1-cocycle condition allows to define a representation U of the gauge group $\mathcal{G}$ on the space $\Omega^0(\mathcal{A}_N)$ (the Hilbert space of wave functions in the physics terminology) by the formula

$$(U(g)\Psi)(A) = e^{iC(A,g)}\Psi(A^g).$$

Using (19.14), we can construct a one-dimensional complex vector space $L_N(A)$, which depends on N and connection A in a G-bundle on N, so that $\exp(iS_{CS}(A))$ is well-defined as an element of $L_N(A)$. To do that, we use the following approach: we first construct a vector space $L_N(A, \varphi)$ which depends on some auxiliary data (namely, trivialization φ of the bundle P_N) and then construct canonical isomorphisms $f_{\psi\varphi} \colon L_N(A, \varphi) \to L_N(A, \psi)$. This allows one to identify all the spaces $L_N(A, \varphi)$ for different choices of φ and thus construct a single space $L_N(A)$. More formally, we will use the following lemma.

Lemma 19.16. *Given a set X, consider the category $\mathscr{C}_X$, whose objects are collections of vector spaces $V_\alpha, \alpha \in X$, together with isomorphisms $f_{\beta\alpha} \colon V_\alpha \to V_\beta$ satisfying $f_{\alpha\alpha} = id$, $f_{\gamma\beta} f_{\beta\alpha} = f_{\gamma\alpha}$.*

Then $\mathscr{C}_X$ is equivalent to the category of vector spaces, with the equivalence given by the following functor:

$$\Gamma \colon \mathscr{C}_X \to \mathrm{Vect}$$
$$(V_\alpha, f_{\beta\alpha}) \mapsto V = \{(v_\alpha \in V_\alpha)_{\alpha \in X} \mid f_{\beta\alpha} v_\alpha = v_\beta\}.$$

In other words, a collection of vector spaces V_α together with collection of canonical isomorphisms $V_\alpha \simeq V_\beta$ can be replaced by a single vector space. A common example of such a situation is when X is a contractible topological space, and V is a local system over X; in this case, Γ is the functor of global sections.

Now, let A be a connection in a G-bundle P on N; as discussed before, such a bundle is automatically trivializable. Let X be the set of all trivializations of P; for each such choice of a trivialization φ, define the vector space $V_\varphi = \mathbb{C}$ and let A^φ be the one-form of connection in this trivialization. Recall (see Theorem 17.11) that if we are given two trivializations φ_α, φ_β related by

$$\varphi_\alpha = g\varphi_\beta, \quad g \in C^\infty(N, G)$$

then

$$A^\beta = g^{-1}A^\alpha g + g^{-1}dg = (A^\alpha)^g.$$

For such a pair of trivializations, define isomorphisms $f_{\beta\alpha}\colon V_\alpha \simeq V_\beta$ by

$$f_{\beta\alpha} = \exp(iC(A^\alpha, g)) \tag{19.16}$$

where $C(A^\alpha, g)$ is defined by (19.14). It follows from the cocycle condition (19.15) that so defined $f_{\beta\alpha}$ satisfy $f_{\alpha\alpha} = \mathrm{id}$, $f_{\gamma\beta}f_{\beta\alpha} = f_{\gamma\alpha}$.

This shows that we can canonically identify vector spaces V_α for different choice of trivialization φ_α and thus, by Lemma 19.16, we can replace this collection by a single vector space which doesn't depend on the choice of trivialization and is thus defined by the connection A on N. We will denote the resulting one-dimensional vector space by $L_N(A)$.

Lemma 19.17. *Let M be a compact oriented 3-manifold with boundary $N = \partial M$. Let A be a connection in a principal G-bundle P on M; denote by A_N the restriction of this connection to N. Then $\exp(iS_{CS}(A))$ is well-defined as an element in one-dimensional vector space $L_N(A_N)$.*

Proof. Choose a trivialization φ of P; abusing the language, let A^φ be the one-form of connection A in this trivialization. Define the number $z_\varphi = \exp(iS_{CS}(A^\varphi)) \in \mathbb{C}$ by formula (19.9); we will consider it as an element in the vector space V_φ. Then it follows from (19.14) that given two trivializations φ_α, φ_β, we have

$$z_\beta = f_{\beta\alpha}z_\alpha$$

which shows that $z = \exp(iS_{CS}(A))$ is well-defined as an element of $L_N(A)$. □

19.6. The phase space of Chern–Simons theory

Recall that in classical mechanics, for a system with configuration space N we had the spacetime $\mathbb{R} \times N$ and the phase space TN (or T^*N — in non-degenerate case, Legendre transform gives an isomorphism $TN \simeq T^*N$). For field theories such as the scalar field theory, the configuration space is taken to be the space of fields on N, and the phase space is the space of 1-jets of fields.

For Chern–Simons model, the situation is different. As discussed before, equations of motion for Chern–Simons theory are first order equations, and not every connection on N

can be extended to a solution of equations of motion on $\mathbb{R} \times N$ (it can only be done for flat connections). Thus, the approach above doesn't work for Chern–Simons theory.

Instead, we can use an alternative definition. Recall that in classical mechanics, for every $(\boldsymbol{q}, \dot{\boldsymbol{q}}) \in TN$, there is a unique solution of equations of motion on $\mathbb{R} \times N$ with this initial conditions. Thus, one can alternatively define the phase space as the space of solutions of equations of motion on $\mathbb{R} \times N$; we have used this in Section 5.3 to define the action of the Galielean group on the phase space.

Following the same idea, for a compact oriented 2-manifold N we define the phase space of Chern–Simons model $X(N)$ to be the space of solutions of equations of motion on $M = \mathbb{R} \times N$, or, equivalently, the space of gauge equivalence classes of flat connections on M. Since N and $\mathbb{R} \times N$ are homotopy equivalent, this is the same as the space of flat connections on N up to gauge equivalence; as discussed in Section 19.4, this gives

$$X(N) = \mathrm{Hom}(\pi_1(N), G)/G. \tag{19.17}$$

As discussed in Section 19.4, X is a finite-dimensional space, which in general is not smooth.

Theorem 19.18.

(1) *Let us call an element $\rho \in \mathrm{Hom}(\pi_1(N), G)$ simple if the centralizer in G of image of ρ is finite; denote by $\mathrm{Hom}^0(\pi_1(N), G)$ the set of simple morphisms. Then for a surface N of genus $g > 1$, the set of simple morphisms is open and dense in $\mathrm{Hom}(\pi_1(N), G)$, and the quotient space*

$$X^0(N) = \mathrm{Hom}^0(\pi_1(N), G)/G$$

is a smooth manifold of dimension

$$\dim X^0(N) = (2g - 2) \dim G.$$

(2) *The manifold $X^0(N)$ has a canonical symplectic structure.*

We skip the proof of this theorem, referring the reader to papers listed in Notes.

This phase space can also be constructed in a different way, via the Hamiltonian reduction. Namely, choose a principal G-bundle P over N (note that P must be necessarily trivializable) and consider the space $\mathcal{A}(P)$ of all connections in P. Recall that $\mathcal{A}(P)$ is an affine space over the vector space $\Omega^1(N, \mathfrak{g}_P)$.

Lemma 19.19. *The space $\mathcal{A}(P)$ has a natural structure of (infinite-dimensional) symplectic manifold.*

Proof. Since $\mathcal{A}(P)$ is an affine space, at every point $A \in \mathcal{A}(P)$ the tangent space $T_A\mathcal{A}(P)$ is the space $\Omega^1(N, \mathfrak{g}_P)$. Define the 2-form ω on $\mathcal{A}(P)$ by letting

$$\omega(\alpha, \beta) = \int_N \langle \alpha \wedge \beta \rangle, \qquad \alpha, \beta \in \Omega^1(N, \mathfrak{g}_P). \tag{19.18}$$

It is easy to show that this 2-form is non-degenerate and closed. □

Recall that the space of connections has a natural action of the gauge group $\mathcal{G}(P)$ described in Section 17.5.

Lemma 19.20. *The action of $\mathcal{G}(P)$ on $\mathcal{A}(P)$ is Hamiltonian; for $a \in Lie(\mathcal{G}(P)) = \Gamma(N, \mathfrak{g}_P)$, the corresponding Hamiltonian function on $\mathcal{A}(P)$ is*

$$H_a(A) = \int_N \langle F_A, a\rangle.$$

where $F_A \in \Omega^2(N, \mathfrak{g}_P)$ is the curvature of A.

In other words, if we define the (restricted) dual of $\Gamma(N, \mathfrak{g}_P)$ to be $\Omega^2(N, \mathfrak{g}_P)$, with the pairing given by $\int_N \langle F, a\rangle$, then the moment map $\mu\colon \mathcal{A} \to \Omega^2(N, \mathfrak{g}_P)$ is given by

$$\mu(A) = F_A. \tag{19.19}$$

Proof. To check that H_a is the Hamiltonian function corresponding to a, we need to check

$$\partial_\eta H_a = \omega(\xi_a, \eta)$$

(see (5.3)).

As is common in calculus of variations, we use notation δA instead of η for the tangent vector at A. Then, using $\delta F_A = d\delta A + [\delta A \wedge A]$, we get

$$\begin{aligned}\delta H_a &= \int_N \langle \delta F_A, a\rangle = \int_N \langle (d\delta A + [\delta A \wedge A]), a\rangle \\ &= \int_N \langle \delta A \wedge (da + [A, a])\rangle \\ &= \int_N \langle (-da + [a, A]) \wedge \delta A\rangle = \omega(-da + [a, A], \delta A).\end{aligned}$$

On the other hand, by formula (17.35) for the action of $a \in \Gamma(N, \mathfrak{g}_P)$ on A, we see that the vector field corresponding to a is

$$\xi_a = -da + [a, A].$$

Thus, $\delta H_a = \omega(\xi_a, \delta A)$, which shows that H_a is indeed the Hamiltonian function corresponding to $a \in \Gamma(N, \mathfrak{g}_P)$.

We leave it as an exercise to the reader to verify that $\{H_a, H_b\} = H_{[a,b]}$. □

As an immediate corollary, we get the following theorem.

Theorem 19.21. *The phase space $X(N)$ of flat connections in G–bundle P on N modulo gauge equivalence coincides with the Hamiltonian reduction of $\mathcal{A}(P)$ by the group of gauge transformations:*

$$X(N) = \mathcal{A}(P)//\mathcal{G}(P).$$

It can be shown that the symplectic structure on $X^0(N)$ mentioned in Theorem 19.18 coincides with the symplectic structure obtained by the Hamitlonian reduction. This gives a conceptual construction of this symplectic structure.

It is natural to ask if $X^0(N)$ is a cotangent bundle — or, in a weaker form, if it has a natural polarization (see Section 4.5 for the discussion of polarizations). It turns out that there is no natual polarization on $X^0(N)$ in the sense of Definition 4.12. However, by choosing a complex structure on N, we can obtain a complex polarization of $X^0(N)$;

this plays an important role in Wess–Zumino–Witten model. We refer the reader to papers listed in Notes to this Part for further discussion.

19.7. Exercises

Exercise 19.1.

(1) Show that a 2×2 matrix g is in SU(2) iff it has the form
$$g=\begin{pmatrix}\alpha & \beta\\ -\overline{\beta} & \overline{\alpha}\end{pmatrix},\qquad \alpha,\beta\in\mathbb{C},\quad |\alpha|^2+|\beta|^2=1.$$

(2) Deduce from it that the map
$$\begin{aligned}\imath\colon \mathrm{SU}(2)&\hookrightarrow \mathbb{C}^2=\mathbb{R}^4\\ \begin{pmatrix}\alpha & \beta\\ -\overline{\beta} & \overline{\alpha}\end{pmatrix}&\mapsto(\alpha,\beta)\end{aligned}\tag{19.20}$$
identifies SU(2) with the unit sphere $S^3\subset\mathbb{R}^4$.

(3) Let $(\ ,\ )_{\mathbb{R}^4}$ be the metric on SU(2) obtained as the pullback of the standard metric on $\mathbb{R}^4$ under the embedding (19.20). Show that the matrices below form an orthonormal basis of $\mathfrak{su}(2)$ with respect to $(\ ,\ )_{\mathbb{R}^4}$:
$$X_1=i\sigma_2=\begin{pmatrix}0&1\\-1&0\end{pmatrix},\qquad X_2=i\sigma_1=\begin{pmatrix}0&i\\i&0\end{pmatrix},\qquad X_3=i\sigma_3=\begin{pmatrix}i&0\\0&-i\end{pmatrix}.$$
Deduce from this that
$$(a,b)_{\mathbb{R}^4}=-\frac{1}{2}\operatorname{tr}(ab),\qquad a,b\in\mathfrak{su}(2).$$

Exercise 19.2.

(1) Let $\tilde{\lambda}\in\Omega^3(\mathrm{SU}(2))$ be the left-invariant form normalized by the condition
$$\tilde{\lambda}(X_1,X_2,X_3)=1,$$
where X_1,X_2,X_3 are defined in the previous exercise.
Show that $\int_{SU(2)}\tilde{\lambda}=2\pi^2$ (for appropriate choice of orientation on SU(2)).

(2) Show that
$$\tilde{\lambda}(a,b,c)=-\frac{1}{4}\operatorname{tr}(a[b,c]).$$

Exercise 19.3. Let $\imath_k\colon \mathrm{SU}(2)\to\mathrm{SU}(n)$ be the embedding given by
$$\imath_k\colon g\mapsto\begin{pmatrix}1&&&&\\&1&&&\\&&g&&\\&&&1&\\&&&&\ddots\end{pmatrix}\qquad g \text{ in positions } k,k+1$$
It is known that for any k, the map $\imath_k$ induces an isomorphism $H_3(\mathrm{SU}(2),\mathbb{Z})\to H_3(\mathrm{SU}(n),\mathbb{Z})$. Deduce from this that for any $n\geq 2$, the left-invariant 3-form on SU(n) given by
$$\lambda(a,b,c)=-\frac{1}{4\pi}\operatorname{tr}(a[b,c])$$
satisfies the normalization condition (19.3).

Exercise 19.4. Prove formula (19.7).

Exercise 19.5.

(1) For a one-form $A \in \Omega^1(M, \mathfrak{g})$ and a function $g \in C^\infty(M, G)$, denote $A' = g^{-1}Ag$. Prove that then $dA' = -[\theta_g, A'] + g^{-1} \cdot dA \cdot g$.

(2) Using the previous part and the Maurer–Cartan equation $d\theta + \frac{1}{2}[\theta \wedge \theta] = 0$, prove (19.12).

Exercise 19.6. For a connection A on a 2-manifold N, let $L_N(A)$ be the one-dimensional vector space defined in Section 19.5. Prove that if $N = N_1 \sqcup N_2$, then one has natural isomorphisms $L_N(A) \simeq L_{N_1}(A) \otimes L_{N_2}(A)$.

Notes and References

From the physics perspective, there are many excellent introductions to classical electrodynamics, from the classic text [**LL1971**] to the monographs [**Gri1999**], [**Jac1998**], [**Ryd1996**]. Correspondingly, we start Chapter 15 with the physics formulation of Maxwell's equations in Section 15.1, and give their invariant formulation by using the differential forms in Section 15.2. The key role there is played by the Hodge star operator in the 4-dimensional space $\mathbb{R}^{1,3}$ with the Minkowski metric.

In Section 15.4 we show, in invariant form, that Maxwell equations are the Euler-Lagrange equations for the action functional of the electromagnetic 4-potential, a 1-form A. In Section 15.5 we derive the stress-energy tensor of electromagnetic field by using the action of Poincaré group on the space of gauge equivalence classes of A. The obtained tensor is symmetric, so there is no need in ad hoc addition of a divergence-free term, as it is customary in physics textbooks. The integration on Riemannian manifolds and properties of the Hodge $\star$ operator are discussed in many textbooks on differential geometry, e.g. in [**War1983**] and [**DFN1984**].

The Hamiltonian formulation of Maxwell's equations in Section 16.4 is based on the Dirac formalism [**Dir1958**], as developed in [**Fad1969**]. We are following [**FS1991**] and unpublished Faddeev's lectures on Feynman path integral and gauge fields (Leningrad State University, 1974). In Section 16.5 we briefly mention QED (quantum electrodynamics); the interested reader can find its detailed account in the classic textbook [**Wei1995**].

Chapter 17 serves as a crash course on connections and curvature on principal and vector bundles and theory of characteristic classes. The reader can find more details in [**MS1974**], [**DFN1984**], [**KN1996a**], [**KN1996b**], [**Fra2012**], and in many other sources. For complex vector bundles, standard references are [**GH1994**] and [**Wel2008**].

Based on the material in Chapter 17, in Section 18.1 we introduce Yang-Mills fields as connections in a principal G-bundle over a pseudo-Riemannian manifold. In Section 18.2 we show that this theory naturally generalizes the abelian gauge theory — classical electrodynamics — to the non-abelian case, and in Section 18.3 we derive the stress-energy tensor for the Yang-Mills field on the Minkowski spacetime. It should be noted that this theory, for the basic case $G = \mathrm{SU}(2)$, was introduced by C. N. Yang and R. L. Mills in the classic paper [**YM1954**]. Nowadays there are many lectures and books introducing classical theory of Yang-Mills fields, written from mathematics and physics perspective. They include classic lectures [**Ati1979**], introductory texts [**NS1987**] and [**BM1994**], and books [**Rub2002**], [**Nak2003**] for mathematically oriented physicists.

Yang-Mills equations on the Minkowski spacetime are nonlinear partial differential equations, so one cannot use Fourier method for solving the Cauchy problem, and we refer to [**EM1982**] and [**CS1997**] for global existence of solutions. In Section 18.4 we briefly mention relation with the quantum chromodynamics (QCD) and the Standard Model of elementary particles; for more information, see [**Wei1996**] and references there.

Our exposition of the Hamiltonian formalism for the Yang-Mills theory in Section 18.5 follows [**Fad1969**], [**FS1991**] and Faddeev's 1974 unpublished lectures. Hamiltonian formulation is fundamental for the quantization of the Yang-Mills fields, and we refer to [**FS1991**]. The Faddeev-Popov operator M plays a fundamental role in developing perturbation theory (Feynman rules). It was rigorously proved in [**Sin1978**] that for large $A_k(\boldsymbol{x})$ the operator M may have a zero eigenvalue, so that the Coulomb gauge condition can intersect the orbits of gauge group more that once — the so-called *Gribov ambiguity*. However, in does not affect the perturbation theory, based on the Hamiltonian formulation of the Yang-Mills theory.

Section 18.6 is devoted to the self-duality equations, introduced in the classic paper [**BPST1975**]; for complete description of solutions in case $M = S^4$ and $G = \mathrm{SU}(2)$, we refer to Atiyah's lectures [**Ati1979**], and references therein. These equations were used by S. K. Donaldson [**Don1983**] for the study of topology of 4-manifolds. For detailed exposition of the Donaldson's theory, we refer to the monographs [**DK1990**] and [**FU1991**]. The Hitchin equations were introduced by N.J. Hitchin in [**Hit1987a**]. With each solution (A, Φ) of Hitchin's equations, where A is a connection in a complex vector bundle E, one associates a *Higgs bundle* — the pair (E, Φ). In [**Hit1987b**], the Higgs bundles were used to define the *Hitchin integrable system*, which plays important role in many areas of mathematics.

Finally, in Chapter 19 we introduce a very different kind of gauge theory, the Chern–Simons model. This model is based on Chern–Simons form, an example of *secondary characteristic class*, which was introduced by Chern and Simons in 1974 in [**CS1974**]. This form was used by physicists as an additional, "topological" term in gauge field theories in 1980s; pure Chern–Simons theory, which had no other terms, was studied in [**Zuc1987**]. (In special case of abelian groups, it has appeared earlier in the work of A. Schwarz [**Sch1977**]). Since this theory is purely topological (the action doesn't depend on metric), it was mostly considered a toy model by physicists — until the publication of Witten's paper [**Wit1989**], in which he showed that quantum Chern–Simons theory can be used to construct invariants of knots, in particular the Jones polynomial. This led to an avalanche of papers on Chern–Simons theory, invariants of knots and 3-manifolds, and topological quantum field theories in general; a review of some of related results can be found in [**Fre2009**].

Invariant definition of the functional W_N in Section is due to S.P. Novikov [**Nov1982**] and E. Witten [**Wit1984**]. The Chern-Simons and Wess-Zumino actions play a prominent role in theory of quantum anomalies, and we refer the reader to [**FS1984**] and references therein.

Our exposition mostly follows [**Fre1995**], to which we refer the reader for details. Most topology results we use in that section are standard; proof of Theorem 19.6 (every 3-manifold is a boundary) can be found in [**Kir1989**, Chapter VII]. For the description of the symplectic structure on the space of gauge equivalence classes of flat connections on a surface, we refer the reader to the original papers [**AB1983**], [**Gol1984**]; note that this was found independent of realization of this space as the phase space of Chern–Simons field theory.

Part 4

Theory of Gravity

Chapter 21

General Relativity

General relativity — Einstein's theory of gravity — is the geometric theory describing gravity as property of space and time. It generalizes special relativity by postulating that the metric of spacetime is not constant but is itself a field described by appropriate equations of motion which generalize Newton's law of universal gravitation. Here we introduce the general notion of a spacetime and briefly recall the necessary basic from Riemannian geometry.

21.1. Spacetime in general relativity

Definition 21.1. A smooth connected n-dimensional 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 $(1, n-1)$ at every $x \in M$. Such a metric ds^2 is called *Lorentzian metric.*

In particular, a 4-dimensional Lorentzian metric has signature $(+,-,-,-)$, so $g = \det g_{\mu\nu} < 0$.

The basic example of a Lorentzian manifold is Minkowski space $\mathbb{R}^{1,3}$; a more general example is $M = \mathbb{R} \times N$, where N is an $(n-1)$-dimensional Riemannian manifold, and the metric on M is defined by $g_M = dt^2 - g_N$. However, these are not the only examples.

Theorem 21.2. *Let M be a connected n-dimensional manifold. Then the following are equivalent*

(1) *M admits a Lorentzian metric.*
(2) *The tangent bundle TM can be split as $TM = L \oplus Q$, for some vector bundles L and Q of ranks 1 and $n-1$ respectively.*

Proof in one direction is quite easy: since it is well-known that any manifold M admits a Riemannian metric $\tilde{g}$, given a splitting $TM = L \oplus Q$ we can define the Lorentzian metric on M by $g = \tilde{g}|_L - \tilde{g}|_Q$. We refer the reader to standard texbooks for the proof in the opposite direction.

As an immediate corollary, we get the following result.

Corollary 21.3.

(1) *Every non-compact connected manifold admits a Lorentzian metric.*

(2) *A compact connected manifold admits a Lorentzian metric if and only if its Euler characteristic is zero.*

(3) *A connected manifold admits a Lorentzian metric if and only if it admits a nowhere vanishing vector field X.*

As in the case of Minkowski metric in Section 7.1, a tangent vector $v \in T_x M$ is called

- timelike if $g(v, v) > 0$
- null (lightlike) if $g(v, v) = 0$
- spacelike if $g(v, v) < 0$

A curve $\gamma\colon [u_1, u_2] \to M$ is called 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 $u \in [u_1, u_2]$.

It is easy to see that for every point x, the set of all timelike vectors is not connected and has two connected components. It is natural to think of one of them as going in positive time direction and the other, in the negative time direction. This motivates the following definition.

Definition 21.4. A Lorentzian manifold M is called *time-orientable* if admits a timelike vector field $X \in \mathrm{Vec}(M)$.

(Note that by definition, a timelike vector filed is non-vanishing.)

Such a vector field defines time orientation: a timelike or null vector $u \in T_x M$ is *future-directed* (or *past-directed*), if $u \cdot X_x > 0$ (or $u \cdot X_x < 0$).

Similarly, a timelike curve $\gamma\colon [u_1, u_2] \to M$ is future-directed (or past-directed), if $\gamma'(u)$ is future-directed (or past-directed) for all $u \in [u_1, u_2]$.

Note that not every Lorentzian manifold is time-orientable; however, it easily follows from Corollary 21.3 that every Lorentzian manifold admits a (possibly different) Lorentzian metric which is time-orientable.

We have just introduced all notions necessary for the definition of a spacetime in general relativity.

Definition 21.5. A *spacetime* of general relativity is time-oriented Lorentzian four-manifold M.

Denote by $\mathscr{M}$ be the space of all Lorentzian metrics on M, and by $\mathcal{G} = \mathrm{Diff}(M)$ — the group of all diffeomorphisms of M, so there is a natural $\mathcal{G}$-action on $\mathscr{M}$ by pullbacks. Explicitly, for $\Phi \in \mathcal{G}$ and $ds^2 = g_{\mu\nu}(x)dx^\mu dx^\nu$ we have

$$\Phi \cdot ds^2 = (\Phi^* g_{\mu\nu})(x)dx^\mu dx^\nu, \quad \text{where} \quad \Phi^* g_{\mu\nu}(x) = g_{\alpha\beta}(\Phi(x))\frac{\partial \Phi^\alpha(x)}{\partial x^\mu}\frac{\partial \Phi^\beta(x)}{\partial x^\nu}. \tag{21.1}$$

From a physicst point of view, this amounts to a change of coordinate systems on the spacetime M. In general relativity, the action of $\mathcal{G}$ on M replaces the Poincaré group action on the Minkowski spacetime $\mathbb{R}^{1,3}$, discussed in Section 7.1. Correspondingly, the special principle of relativity is replaced by

General Principle of Relativity. The field equations — the laws of motion — are invariant under the action of the diffeomorphism group $\mathcal{G}$ of the spacetime M.

Remark 21.6. It should be noted that all physical theories are diffeomorphism invariant: the physical laws do not depend on a coordinate system chosen to write them down. The general relativity describes the dynamics of spacetime, so its diffeomorphism invariance — a "general covariance" in physics language — manifests the geometric nature of the theory. As the result, the field equations are written in a covariant form in terms of tensors.

We have the following analogs of future and past lightcones of each point in Minkowski spacetime (see Section 7.1). Namely, 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, and 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.

Note that $x \in I_+^M(x)$ if and only if there exists a timelike curve starting and ending at x. From physical point of view, it means that our system admits time travel, which leads to numerous difficulties.

Proposition 21.7. *If the spacetime M is compact, there exists a closed timelike curve in M.*

Proof. The family $\{I_+^M(x)\}_{x\in M}$ is an open covering of M. By compactness, $M = I_+^M(x_1)\cup \cdots \cup I_+^M(x_m)$. If $x_1 \notin I_+^M(x_1)$, then $x_1 \in I_+^M(x_2) \cup \cdots \cup I_+^M(x_m)$, then $x_1 \in I_+^M(x_k)$ for some $2 \leq k \leq m$. Then $I_+^M(x_1) \subseteq I_+^M(x_k)$, so we can omit $I_+^M(x_1)$ from the covering. If $x_2 \notin I_+^M(x_2)$, then by the same argument we can remove $I_+^M(x_2)$ form the covering, and so on, until we find some $x_j \in I_+^M(x_j)$. $\square$

Thus compact spacetime allows for the time travel, so we will consider only non-compact spacetimes.

We say that 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 $x \in M$ 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 passes through V more then once. Clearly the strong causality condition implies the causality condition.

Definition 21.8. 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 describes the structure of globally hyperbolic spacetimes explicitly: they are foliated by smooth spacelike hypersurfaces.

Theorem 21.9. *Let M be a spacetime. Then the following are equivalent.*

(1) *M is globally hyperbolic.*

(2) *M is isometric to $\mathbb{R}\times N$ with the Lorentzian metric $\beta dt^2-\gamma_t$, where β is a smooth positive function on M, and γ_t is is a Riemannian metric on N depending smoothly on $t\in\mathbb{R}$.*

It is obvious from the construction that every timelike curve γ intersects every hypersurface $N_t=\{t\}\times N\subset M$ at most once, and in fact, every such curve can be reparametrized and extended so that intersects each N_t exactly once. Hypersurfaces with these properties are called *Cauchy hypersurfaces.*

Corollary 21.10. *On every globally hyperbolic spacetime M there exists a smooth function $h: M\to\mathbb{R}$ whose gradient $\boldsymbol{\nabla} h\in\operatorname{Vect}(M)$ is timelike and future-directed, and all level sets of h are spacelike Cauchy hypersurfaces. Moreover, given a Cauchy hypersurface N in M, there is a function h with the property $N=h^{-1}(0)$.*

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.

To better appreciate the importance of Theorem 21.9, let us discuss, from physics perspective, the basic notions of space and time in general relativity. Let M be a spacetime (we do not assume it is globally hyperbolic) and let (x^0,x^1,x^2,x^3) be an arbitrary local coordinate system with the only condition that $g_{00}>0$. Thus the vector field $\partial/\partial x^0$ is timelike, so we can think of x^0 as local time coordinate and of $\boldsymbol{r}=(x^1,x^2,x^3)$ as space coordinates.

Such a choice of local coordinate system allows one to talk about events happening at the same point in space: we say that two events $(x^0_1,\boldsymbol{r}_1)$ and $(x^0_2,\boldsymbol{r}_2)$ happen at the same point in space if $\boldsymbol{r}_1=\boldsymbol{r}_2$ (note that this notion depends on the choice of coordinate system). As in Section 8.1, we introduce a proper time for an observer at point $\boldsymbol{r}$ in space by defining the time interval between events $(x^0_1,\boldsymbol{r})$ and $(x^0_2,\boldsymbol{r})$ to be

$$\Delta\tau=\frac{1}{c}\int_{x^0_1}^{x^0_2}\sqrt{g_{00}(x^0,\boldsymbol{r})}\,dx^0.$$

This defines a proper time τ at point $\boldsymbol{r}$ in space, which is defined uniquely up to adding a constant (depending on $\boldsymbol{r}$).

Next, we need to define a notion of simultaneous events at different points in space, which of course depends on the choice of coordinate system. This is done by "synchronizing clocks", using physicists terminology. Namely, suppose that we have two nearby points A and B in space with coordinates $\boldsymbol{r}$ and $\boldsymbol{r}+d\boldsymbol{r}$. Suppose that at some time at B we send a signal from to point A, propagating at speed of light along γ, and as soon as it reaches point $\boldsymbol{r}$ at x^0, we immediately send it back. It is very easy to see that the signal should be send from B at time $x^0+dx^0_1$, and it comes back to B at time $x^0+dx^0_2$, where $dx^0_{1,2}$ are two roots of the quadratic equation[1]

$$ds^2=g_{00}(x^0,\boldsymbol{r})(dx^0)^2+2g_{0i}(x^0,\boldsymbol{r})dx^0dx^i+g_{ij}(x^0,\boldsymbol{r})dx^idx^j=0. \tag{21.2}$$

[1]It is the condition that lift $\tilde{\gamma}$ of a curve γ to spacetime is lightlike.

Then we say that the events $(x^0, \boldsymbol{r})$ and $(x^0+dx^0, \boldsymbol{r}+d\boldsymbol{r})$ in the spacetime are simultaneous (or that the clocks at points $\boldsymbol{r}$ and $\boldsymbol{r}+d\boldsymbol{r}$ in space are synchronized), if

$$x^0 + dx^0 = \frac{1}{2}(x^0 + dx^0_1 + x^0 + dx^0_2).$$

Using (21.2), we readily obtain

$$dx^0 = -\frac{g_{0i}}{g_{00}}dx^i.$$

This definition allows to synchronize clocks along any curve $\gamma(u) = \boldsymbol{r}(u)$ in space; events $(x^0, \boldsymbol{r}(0))$ and $(x^0(u), \boldsymbol{r}(u))$ are simultaneous (in a given coordinate system), if $x^0(u)$ is the solution of the following nonlinear ordinary differential equation

$$\frac{dx^0(u)}{du} = -\frac{g_{0i}(x^0(u), \boldsymbol{r}(u))}{g_{00}(x^0(u), \boldsymbol{r}(u))}\frac{dx^i(u)}{du}, \quad x^0(u)\big|_{u=0} = x^0.$$

However, this depends on the choice of γ connecting two points in space. Thus, in general it is impossible to synchronize clocks at all points in space, even locally.

However, if our coordinate system was chosen so that $g_{0i} = 0$ for all $i = 1, 2, 3$, then it follows from this formula that one can synchronize clocks, just by declaring two events simultaneous if they have the same x^0 coordinate. Such coordinate systems are called *"synchronous"*.

It is not very difficult to show that locally, any spacetime admits a synchronous coordinate system and thus, locally one can find a way to synchronize the clocks. The importance of Theorem 21.9 is that it implies that for a globally hyperbolic spacetime, one can also synchronize the clocks globally, which is a highly nontrivial result.

21.2. Particle in a gravitation field and Levi-Civita connection

The action of a relativistic particle of mass m in a gravitational field described by the metric tensor $g_{\mu\nu}(x)$ has the same form as in Section 8.2,

$$S(\gamma) = -mc\int ds = -mc\int\sqrt{g_{\mu\nu}u^\mu u^\nu}ds, \quad u^\mu = \frac{dx^\mu}{ds}. \tag{21.3}$$

In other words, the action functional is $-mc$ times the length functional in the pseudo-Riemannian geometry. Correspondingly, (cf. Example 1.6 in Chapter 1), the Euler-Lagrange equations are the geodesic equations with respect to the natural parameter for the metric $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$ (there is no difference between Riemannian and pseudo-Riemannian cases).

It is well-known that the geodesic equations can be written down explicitly in terms of Christoffel's symbols of the Levi-Civita connection.

Recall that a Levi-Civita connection for the metric $g_{\mu\nu}(x)$ on the spacetime M is a unique connection ∇ in the tangent bundle TM such that

$$\begin{aligned}\partial_X\langle Y, Z\rangle &= \langle\nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle,\\ \nabla_X Y - \nabla_Y X &= [X, Y]\end{aligned} \tag{21.4}$$

for all $X, Y, Z \in \mathrm{Vect}(M)$ — sections of TM, where $\langle\ ,\ \rangle$ is the pseudo-inner product of vector fields, determined by the metric on M, and covariant derivatives ∇_X, ∇_Y were introduced in (17.19). In other words, Levi-Civita connection is a metric connection with no torsion.

As was discussed in Section 17.2, a choice of local coordinates x^μ on a chart $U \subset M$ gives a trivialization of the tangent bundle TM over U, so $\nabla = d + A$, where $A = A_\mu(x)dx^\mu$ and $A_\mu(x) \in \operatorname{End} T_xM$. Using the basis $\partial_\mu = \dfrac{\partial}{\partial x^\mu}$ of vector fields over U, we can represent A_μ by matrices

$$(A_\mu)^\lambda_\nu = \Gamma^\lambda_{\nu\mu}, \tag{21.5}$$

so covariant derivative of a $(1,0)$-tensor — a vector field $V = v^\mu(x)\partial_\mu$ — is given by

$$(\nabla_\mu V)^\lambda = \frac{\partial v^\lambda}{\partial x^\mu} + \Gamma^\lambda_{\nu\mu} v^\nu. \tag{21.6}$$

Also, for $X = a^\mu(x)\partial_\mu$, $Y = b^\mu(x)\partial_\mu$ we have

$$\langle X, Y\rangle = g_{\mu\nu}(x)a^\mu(x)b^\nu(x).$$

As it is written in standard differential geometry textbooks, from here we readily obtain that unique solution of equations (21.4) is the connection $d + A$ with coefficients $\Gamma^\lambda_{\mu\nu}$ given by

$$\Gamma^\lambda_{\mu\nu} = \frac{1}{2} g^{\lambda\sigma}\left(\frac{\partial g_{\mu\sigma}}{\partial x^\nu} + \frac{\partial g_{\nu\sigma}}{\partial x^\mu} - \frac{\partial g_{\mu\nu}}{\partial x^\sigma}\right), \tag{21.7}$$

the Christoffel's symbols.

Covariant derivatives can be defined for a $(0,1)$-tensor — a section of the cotangent bundle T^*M, a 1-form $\theta = a_\mu dx^\mu$ — by

$$(\nabla_\mu \theta)_\lambda = \frac{\partial a_\lambda}{\partial x^\mu} - \Gamma^\nu_{\lambda\mu} a_\nu, \tag{21.8}$$

as well as for an arbitrary (p,q)-tensor, a section of the bundle $TM^{\otimes p} \otimes (T^*M)^{\otimes q}$. Equations (21.4) can be written succinctly as

$$\nabla_\lambda g_{\mu\nu} = 0 \quad \text{and} \quad \Gamma^\lambda_{\mu\nu} = \Gamma^\lambda_{\nu\mu}. \tag{21.9}$$

Remark 21.11. In classical tensor analysis, tensors of type $(p,0)$ are called *contravariant* tensors, and tensors of type $(0,q)$ — *covariant* tensors.

Thus the Euler-Lagrange equations for the action (21.3) of a free particle in a gravitational field is the geodesic equation

$$\frac{d^2x^\lambda}{ds^2} + \Gamma^\lambda_{\mu\nu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} = 0, \tag{21.10}$$

written in terms of the natural parameter.

It is very instructive to consider the case of a *weak gravitational field* in $\mathbb{R}^4$, when

$$g_{\mu\nu}(x) = \eta_{\mu\nu} + \frac{1}{c^2} g^{(2)}_{\mu\nu}(x) + O\left(\frac{1}{c^3}\right) \quad \text{as} \quad c \to \infty, \tag{21.11}$$

where $\eta_{\mu\nu}$ is Minkowski metric, and as in Section 7.4, we consider c as a parameter.

As a special case, consider the gravitational field given by

$$ds^2 = \left(1 + \frac{2\varphi}{c^2}\right)(dx^0)^2 - d\boldsymbol{r}^2 = (c^2 + 2\varphi)dt^2 - d\boldsymbol{r}^2 \tag{21.12}$$

for some function $\varphi = \varphi(x^0, \boldsymbol{r})$ on $\mathbb{R}^4$. In this case, using $\sqrt{1+x} = 1 + \frac{x}{2} + O(x^2)$, we see that natural parameter along a curve γ is given by $\int_\gamma ds$, where

$$ds = \sqrt{ds^2} = c\left(1 + \frac{\varphi}{c^2} - \frac{v^2}{2c^2}\right)dt + O\left(\frac{1}{c^2}\right) \quad \text{and} \quad \boldsymbol{v} = \frac{d\boldsymbol{r}}{dt}.$$

Therefore, the action (21.3) can be rewritten as

$$S = -mc\int_\gamma ds = \int L\,dt + O\left(\frac{1}{c}\right),$$

$$L = -mc^2 + \frac{mv^2}{2} - m\varphi$$

(compare with Section 8.2).

Thus, in the limit $c \to \infty$, motion of a particle in this gravitational field is the same as the motion of a classical particle in Newtonian mechanics in $\mathbb{R}^3$ with time-dependent potential function $V = m\varphi$. By results of Example 1.4, this motion is described by equation

$$\ddot{\boldsymbol{r}} = -\frac{\partial\varphi}{\partial\boldsymbol{r}}.$$

Introducing the potential force

$$\boldsymbol{F} = -m\frac{\partial\varphi}{\partial\boldsymbol{r}},$$

we obtain Newton's equation $m\ddot{\boldsymbol{r}} = \boldsymbol{F}$.

Remark 21.12. Alternatively, one could obtain the equations of motion for a particle in a weak gravitational force by using the equation for geodesic lines (21.10) and analyzing the behavior of Christoffel symbols as $c \to \infty$. We leave this as an exercise to the reader.

21.3. The Riemann curvature tensor

According to Section 17.6, the curvature of the Levi-Civita connection $\nabla = d + A$ is

$$F = dA + A \wedge A,$$

a 2-form on M with values in $\operatorname{End} TM$. We have

$$F = \tfrac{1}{2}F_{\mu\nu}dx^\mu \wedge dx^\nu, \quad \text{where} \quad F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu} + [A_\mu, A_\nu].$$

On 2-forms B with values in $\operatorname{End} TM$ the connection ∇ acts by

$$\nabla B = dB + [A \wedge B],$$

which gives the Bianchi identity

$$\nabla F = 0$$

for the curvature 2-form. Equivalently,

$$\nabla_\lambda F_{\mu\nu} + \nabla_\mu F_{\nu\lambda} + \nabla_\nu F_{\lambda\mu} = 0.$$

(compare with (17.46).)

Representing $F_{\mu\nu} \in \operatorname{End} TM$ as a matrix introduces a $(1,3)$-tensor on M, the Riemann curvature tensor

$$R^{\lambda}{}_{\rho\mu\nu} = (F_{\mu\nu})^{\lambda}_{\rho},$$

and using (21.5), we obtain

$$R^{\lambda}{}_{\rho\mu\nu} = \frac{\partial \Gamma^{\lambda}_{\rho\nu}}{\partial x^{\mu}} - \frac{\partial \Gamma^{\lambda}_{\rho\mu}}{\partial x^{\nu}} + \Gamma^{\lambda}_{\sigma\mu}\Gamma^{\sigma}_{\rho\nu} - \Gamma^{\lambda}_{\sigma\nu}\Gamma^{\sigma}_{\rho\mu}. \tag{21.13}$$

The Bianchi identity for the Riemann tensor has the form

$$\nabla_{\sigma} R^{\lambda}{}_{\rho\mu\nu} + \nabla_{\nu} R^{\lambda}{}_{\rho\sigma\mu} + \nabla_{\mu} R^{\lambda}{}_{\rho\nu\sigma} = 0. \tag{21.14}$$

The Ricci curvature tensor is a $(0,2)$-tensor on M, defined by

$$R_{\mu\nu} = R^{\lambda}{}_{\mu\lambda\nu},$$

and we have an explicit formula

$$R_{\mu\nu} = \frac{\partial \Gamma^{\lambda}_{\mu\nu}}{\partial x^{\lambda}} - \frac{\partial \Gamma^{\lambda}_{\mu\lambda}}{\partial x^{\nu}} + \Gamma^{\lambda}_{\mu\nu}\Gamma^{\sigma}_{\lambda\sigma} - \Gamma^{\sigma}_{\mu\lambda}\Gamma^{\lambda}_{\sigma\nu}. \tag{21.15}$$

It follows from (21.7) that

$$\begin{aligned} \Gamma^{\lambda}_{\mu\lambda} &= \frac{1}{2} g^{\lambda\sigma} \left(\frac{\partial g_{\mu\sigma}}{\partial x^{\lambda}} + \frac{\partial g_{\lambda\sigma}}{\partial x^{\mu}} - \frac{\partial g_{\mu\lambda}}{\partial x^{\sigma}} \right) \\ &= \frac{1}{2} g^{\lambda\sigma} \frac{\partial g_{\sigma\lambda}}{\partial x^{\mu}} \\ &= \frac{1}{2g} \frac{\partial g}{\partial x^{\mu}} = \frac{\partial \log \sqrt{-g}}{\partial x^{\mu}}, \end{aligned} \tag{21.16}$$

where we used Jacobi's formula for the derivative of the determinant of a matrix: if $a(x) = \det A(x)$, then

$$a^{-1} \frac{da}{dx} = \operatorname{tr} \left(A^{-1} \frac{dA}{dx} \right).$$

Thus we see that the Ricci tensor is symmetric, $R_{\mu\nu} = R_{\nu\mu}$, and determines a symmetric bilinear form $\operatorname{Ric} = R_{\mu\nu} dx^{\mu} dx^{\nu}$ on the tangent spaces.

Finally, the scalar curvature R is the trace of the Ricci curvature tensor,

$$R = R^{\mu}_{\mu} = g^{\mu\nu} R_{\mu\nu}.$$

Contracting λ and ν in the Bianchi identity (21.14), we get

$$2\nabla_{\mu} R_{\rho\sigma} - \nabla_{\sigma} R_{\rho\mu} = 0$$

and using (21.9) we obtain

$$2\nabla_{\mu} R^{\rho}_{\sigma} - \nabla_{\sigma} R^{\rho}_{\mu} = 0.$$

Finally contracting μ and ρ we obtain the identity

$$2\nabla_{\mu} R^{\mu}_{\sigma} - \nabla_{\sigma} R = 0,$$

or

$$\nabla_{\mu} \left(R^{\mu}_{\nu} - \frac{1}{2} \delta^{\mu}_{\nu} R \right) = 0. \tag{21.17}$$

The tensor

$$G^{\mu}_{\nu} = R^{\mu}_{\nu} - \frac{1}{2}\delta^{\mu}_{\nu} R \tag{21.18}$$

is often called the *Einstein tensor*. It follows from (21.17) that the tensor $G_{\mu\nu} = R - \frac{1}{2} g_{\mu\nu} R$ satisfies

$$\nabla^{\mu} G_{\mu\nu} = 0, \quad \text{where} \quad \nabla^{\mu} = g^{\mu\nu}\nabla_{\nu}.$$

21.4. Exercises

Exercise 21.1. Let $g_{\mu\nu}$ be the Lorentzian metric tensor with the condition $g_{00} > 0$. Prove that γ_{ij} is positive-definite 3×3 matrix if and only if

$$\det \begin{pmatrix} g_{00} & g_{01} \\ g_{10} & g_{11} \end{pmatrix} < 0, \ \det \begin{pmatrix} g_{00} & g_{01} & g_{02} \\ g_{10} & g_{11} & g_{12} \\ g_{20} & g_{21} & g_{22} \end{pmatrix} > 0.$$

Physicists say that these conditions hold for any choice of coordinates on M, which can be realized with the aid of "physical bodies".

Exercise 21.2. As in Theorem 21.9, let $M = \mathbb{R} \times N$ be a spacetime with the metric $ds^2 = \beta(dx^0)^2 - \gamma_{ij}(t,x)dx^i dx^j$. Show that

$$\Gamma^0_{00} = \frac{1}{2}\frac{\partial \log \beta}{\partial x^0}, \quad \Gamma^0_{0j} = \frac{1}{2}\frac{\partial \log \beta}{\partial x^j}, \quad \Gamma^0_{ij} = \frac{1}{2\beta}\frac{\partial \gamma_{ij}}{\partial x^0}, \quad \Gamma^i_{00} = \frac{1}{2}\gamma^{ij}\frac{\partial \beta}{\partial x^j}, \quad \Gamma^i_{0j} = \frac{1}{2}\gamma^{ik}\frac{\partial \gamma_{jk}}{\partial x^0}$$

and $\Gamma^i_{jk} = \gamma^i_{jk}$ — Christoffel's symbols of the metric $\gamma_{ij}(t,x)dx^i dx^j$.

Exercise 21.3. Show that for the weak gravitational field (21.12), we have the following formulas

$$\Gamma^i_{00} = -\frac{1}{2} g^{ii} \frac{\partial g_{00}}{\partial x^i} = \frac{1}{c^2}\frac{\partial \varphi}{\partial x^i}, \quad i = 1, 2, 3,$$
$$\Gamma^0_{00} = O(1/c^3)$$

and all other Christoffel's symbols are of order $O(1/c^2)$ or zero.

Exercise 21.4. Show that the curvature F of the connection ∇ on TM can be also defined by the formula

$$F(X,Y) = [\nabla_X, \nabla_Y] - \nabla_{[X,Y]}, \quad X, Y \in \text{Vect}(M).$$

Exercise 21.5. Show that $\text{Ric}_x(u,v)$ for $u, v \in T_xM$ is the trace of the linear map $T_xM \ni \xi \to F_x(\xi, u)v \in T_xM$.

Chapter 22

Einstein Equations

Gravitational field on a spacetime M is a Lorentzian metric $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$, and we denote by $\mathscr{M}$ the space of all gravitational fields on M. The Einstein field equations relate the gravitational field and the distribution of matter and energy on M.

22.1. Einstein field equations

Let $g_{\mu\nu}(x)$ be a gravitational filed on the spacetime M, interacting with matter and energy. The latter are described by the stress-energy tensor T^μ_ν of the matter fields, which we will carefully introduce and discuss in Section 22.3.

The *Einstein equations* are given by the following beautiful relation between $(1,1)$-tensors

$$G^\mu_\nu = \kappa T^\mu_\nu, \quad \text{where} \quad \kappa = \frac{8\pi G}{c^4}.$$

Here G is the Newtonian constant of gravitation, and κ is called the *Einstein gravitational constant*. Equivalently,

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G}{c^4}T_{\mu\nu},$$

where $R_{\mu\nu}$ is the Ricci curvature, R is the scalar curvature and $T_{\mu\nu}$ is the stress-energy tensor of matter fields.

When fully written out, the Einstein equations are a system of ten coupled, nonlinear, hyperbolic-elliptic partial differential equations. Solutions of the Einstein equations are metrics on the spacetime. The exact solutions of Einstein equations play an important role in cosmology, describing different models of the evolution of the universe.

The initial value formulation of Einstein equations is rather non-trivial and for strongly hyperbolic spacetime consists of prescribing the initial data on a Cauchy hypersurface, subject to constraints that follow from the Gauss-Codazzi relations in differential geometry. We refer to Section 23.2 for more details and to Chapter 24 for references to general relativity textbooks.

It follows from the Bianci identity (21.17) that Einstein equations imply

$$\nabla_\mu T^\mu_\nu = 0, \quad \nu = 0,1,2,3,$$

which confirms the conservation laws in Section 22.3.

Taking traces of the Einstein equations

$$R^\mu_\nu - \frac{1}{2}\delta^\mu_\nu R = \kappa T^\mu_\nu,$$

we obtain

$$R = -\kappa T, \quad \text{where} \quad T = T^\mu_\mu.$$

Thus Einstein equations can be also written as

$$R^\mu_\nu = \kappa\left(T^\mu_\nu - \frac{1}{2}\delta^\mu_\nu T\right). \tag{22.1}$$

In particular, the Einstein equations in an empty space (nor matter or energy) are Ricci flat equations

$$R_{\mu\nu} = 0.$$

It is very instructive to derive the Newton's law of universal gravitation from the Einstein equations, applied to a slowly moving particle in a weak gravitational field on $\mathbb{R}^4$, when

$$g_{00}(x) = 1 + \frac{2\varphi(x)}{c^2}, \quad g_{ij} = -\delta_{ij}, \quad g_{0i} = 0.$$

To find the metric $g_{\mu\nu}$ and hence the potential φ, we need to use Einstein equations in the presence of matter. Namely, suppose that we have a macroscopic body, discussed in Remark 22.6. As it is explained there, its energy-momentum tensor is

$$T^{\mu\nu} = M(x)c^2 u^\mu u^\nu,$$

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 M depends only on space variables and the energy-momentum tensor takes the form

$$T^\mu_\nu = Mc^2\delta^\mu_0\delta^0_\nu.$$

It follows from formula (21.15) and formulas in Exercise 21.2 that in a weak gravitational field $R^\mu_\nu = O(1/c^2)$, so the only nontrivial contribution to the Einstein equations (22.1) is

$$R^0_0 = \frac{4\pi G}{c^4}T = \frac{4\pi GM}{c^2}.$$

It follows from Exercise 21.3 that

$$R^0_0 = \frac{\partial \Gamma^i_{00}}{\partial x^i} + O\left(\frac{1}{c^3}\right) = \frac{1}{c^2}\boldsymbol{\nabla}^2\varphi + O\left(\frac{1}{c^3}\right),$$

so Einstein equations for the weak gravitational field reduce to the Poisson equation

$$\boldsymbol{\nabla}^2\varphi = 4\pi GM$$

for the potential φ. 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 a Newtonian potential

$$\varphi(\boldsymbol{r}) = -\frac{GM}{\boldsymbol{r}}.$$

Thus the force acting on a slow particle of mass m in a weak gravitational field generated by a particle of a mass M is Newtonian universal gravitation force!

22.2. Hilbert action

On the space $\mathscr{M}$ of smooth Lorentzian metrics on the spacetime M consider the celebrated Hilbert (or Einstein-Hilbert) functional

$$S(g_{\mu\nu}) = \int_M R\sqrt{-g}\,d^4x,$$

where R is the scalar curvature of the metric $ds^2 = g_{\mu\nu}dx^\mu dx^\nu \in \mathscr{M}$, $g = \det g_{\mu\nu}$ and $\sqrt{-g}\,d^4x$ is the corresponding volume form on M. However, since our spacetime is non compact, convergence of the integral is rather a non-trivial issue, as we will see in the next section.

Thus for deriving equations of motion we take the point of view that as in Section 11.5. Namely, for any open $D \subset M$ with compact closure we define the action S_D as the integral of the scalar curvature R over D,

$$S_D(g_{\mu\nu}) = \int_D R\sqrt{-g}\,d^4x.$$

Then the metric tensor $g_{\mu\nu}$ is a critical point of the action, if for any such D and any variation $u_{\mu\nu} = \delta g_{\mu\nu}$ with support in D,

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} S_D(g_{\mu\nu} + \varepsilon u_{\mu\nu}) = 0.$$

Proposition 22.1. *The Gateaux derivative of the Hilbert functional S_D in the direction u is given by*

$$\delta_u S = -\int_D \left(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R\right) u^{\mu\nu}\sqrt{-g}\,d^4x, \quad \textit{where} \quad u^{\mu\nu} = g^{\mu\alpha}g^{\nu\beta}u_{\alpha\beta}.$$

Thus the Euler-Lagrange equations of the Hilbert action are Einstein equations in empty spacetime.

Proof. For a (p,q)-tensor $I^{\lambda_1\cdots\lambda_p}_{\mu_1\cdots\mu_q}$ which depends on $g_{\mu\nu}$, put

$$\delta I^{\lambda_1\cdots\lambda_p}_{\mu_1\cdots\mu_q} = \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} I^{\lambda_1\cdots\lambda_p}_{\mu_1\cdots\mu_q}(g_{\mu\nu} + \varepsilon u_{\mu\nu}).$$

Writing

$$S_D(g_{\mu\nu}) = \int_D g^{\mu\nu}R_{\mu\nu}\sqrt{-g}\,d^4x,$$

we obtain

$$\begin{aligned}\delta_u S_D &= \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} S_D(g_{\mu\nu}+\varepsilon u_{\mu\nu})\\ &= \int_D (\delta g^{\mu\nu}R_{\mu\nu}+g^{\mu\nu}\delta R_{\mu\nu})\sqrt{-g}\,d^4x+\int_D R\delta(\sqrt{-g})d^4x,\end{aligned}$$

where $\delta g^{\mu\nu}=-u^{\mu\nu}$. By Jacobi's formula we have

$$\delta g=\frac{d}{d\varepsilon}\bigg|_{\varepsilon=0}\det(g_{\mu\nu}+\varepsilon u_{\mu\nu})=\frac{\partial g}{\partial g_{\mu\nu}}u_{\mu\nu}=g\,g^{\mu\nu}u_{\mu\nu}=-g\,g_{\mu\nu}\delta g^{\mu\nu},$$

so

$$\delta(\sqrt{-g})=-\frac{1}{2}\sqrt{-g}\,g_{\mu\nu}\delta g^{\mu\nu}, \tag{22.2}$$

and we obtain

$$\delta_u S_D=\int_D\left(R_{\mu\nu}-\frac{1}{2}g_{\mu\nu}R\right)\delta g^{\mu\nu}\sqrt{-g}\,d^4x+\int_D g^{\mu\nu}\delta R_{\mu\nu}\sqrt{-g}\,d^4x.$$

It remains to compute $\delta R_{\mu\nu}(x)$. It follows from (21.15) that

$$\delta R_{\mu\nu}=\frac{\delta\Gamma^{\sigma}_{\mu\nu}}{\partial x^{\sigma}}-\frac{\delta\Gamma^{\sigma}_{\mu\sigma}}{\partial x^{\nu}}+\delta\Gamma^{\lambda}_{\mu\nu}\Gamma^{\sigma}_{\lambda\sigma}-\delta\Gamma^{\sigma}_{\mu\lambda}\Gamma^{\lambda}_{\sigma\nu}+\Gamma^{\lambda}_{\mu\nu}\delta\Gamma^{\sigma}_{\lambda\sigma}-\Gamma^{\sigma}_{\mu\lambda}\delta\Gamma^{\lambda}_{\sigma\nu}.$$

Since Christoffel's symbols $\Gamma^{\lambda}_{\mu\nu}$ are matrix elements of a Levi-Civita connection on TM, their variations $\delta\Gamma^{\lambda}_{\mu\nu}$ are $(1,2)$-tensors on M. This can also seen from the formula

$$\delta\Gamma^{\lambda}_{\mu\nu}=\frac{1}{2}g^{\lambda\sigma}(\nabla_\mu u_{\nu\sigma}+\nabla_\nu u_{\mu\sigma}-\nabla_\sigma u_{\mu\nu}), \tag{22.3}$$

which easily follows from (21.7). Using formulas (21.6) and (21.8), we see that the above formula can be written in a covariant form

$$\delta R_{\mu\nu}=\nabla_\sigma\delta\Gamma^{\sigma}_{\mu\nu}-\nabla_\nu\delta\Gamma^{\sigma}_{\mu\sigma},$$

called the *Palatini identity*.

Since $\nabla_\sigma g^{\mu\nu}=0$, we obtain from the Palatini identity

$$g^{\mu\nu}\delta R_{\mu\nu}=\nabla_\sigma(g^{\mu\nu}\delta\Gamma^{\sigma}_{\mu\nu})-\nabla_\nu(g^{\mu\nu}\delta\Gamma^{\sigma}_{\mu\sigma}),$$

so that

$$g^{\mu\nu}\delta R_{\mu\nu}=\nabla_\sigma W^{\sigma},\quad\text{where}\quad W^{\sigma}=g^{\mu\nu}\delta\Gamma^{\sigma}_{\mu\nu}-g^{\mu\sigma}\delta\Gamma^{\rho}_{\mu\rho}.$$

Using formula (21.16) — the relation

$$\Gamma^{\nu}_{\mu\nu}=\frac{\partial}{\partial x^{\mu}}\log(\sqrt{-g}),$$

we obtain

$$\nabla_\mu W^{\mu}=\frac{\partial W^{\mu}}{\partial x^{\mu}}+\Gamma^{\mu}_{\nu\mu}W^{\nu}=\frac{1}{\sqrt{-g}}\frac{\partial}{\partial x^{\mu}}(\sqrt{-g}\,W^{\mu}).$$

Thus we have

$$g^{\mu\nu}\delta R_{\mu\nu}=\frac{1}{\sqrt{-g}}\frac{\partial}{\partial x^{\mu}}(\sqrt{-g}\,W^{\mu}) \tag{22.4}$$

and by the Stokes theorem

$$\int_D g^{\mu\nu}\delta R_{\mu\nu}\sqrt{-g}\,d^4x = \int_D \nabla_\mu W^\mu\sqrt{-g}\,d^4x = \int_D \frac{\partial}{\partial x^\mu}(\sqrt{-g}\,W^\mu) = 0$$

since it follows from (22.3) that $\delta\Gamma^\lambda_{\mu\nu} = 0$ on ∂D. □

Remark 22.2. The basic identity

$$\nabla_\mu X^\mu = \frac{1}{\sqrt{-g}}\frac{\partial}{\partial x^\mu}(\sqrt{-g}\,X^\mu), \tag{22.5}$$

where $X^\mu\partial_\mu$ is a vector field on M, is used in for deriving many formulas in general relativity.

Remark 22.3. It follows from (21.7) and (21.15) that Lagrangian of the Hilbert action $\sqrt{-g}R = \sqrt{-g}g^{\mu\nu}R_{\mu\nu}$ contains second derivatives of the metric tensor $g_{\mu\nu}$. However, by adding a total divergence term, we can have a Lagrangian containing only first derivatives of the gravitational field, as was customary in Part 2. Namely, the Lagrangian

$$\begin{aligned}\mathscr{L} &= \sqrt{-g}g^{\mu\nu}R_{\mu\nu} + \frac{\partial}{\partial x^\mu}(\sqrt{-g}g^{\mu\nu}\Gamma^\sigma_{\nu\sigma} - \sqrt{-g}g^{\nu\sigma}\Gamma^\mu_{\nu\sigma}) \\ &= \Gamma^\sigma_{\nu\sigma}\frac{\partial\sqrt{-g}g^{\mu\nu}}{\partial x^\mu} - \Gamma^\mu_{\nu\sigma}\frac{\partial\sqrt{-g}g^{\nu\sigma}}{\partial x^\mu} + \sqrt{-g}g^{\mu\nu}(\Gamma^\lambda_{\mu\nu}\Gamma^\sigma_{\lambda\sigma} - \Gamma^\sigma_{\mu\lambda}\Gamma^\lambda_{\sigma\nu})\end{aligned} \tag{22.6}$$

contains only first derivatives of the metric tensor $g_{\mu\nu}$. It follows from the proof of Proposition 22.1 that Euler-Lagrange equations for this Lagrangian are Einstein equations in empty spacetime.

22.3. Einstein equations with matter

Here we consider the interaction of a gravitational field $g_{\mu\nu}$ with the matter fields, denoted by φ and described by some field theory with the Lagrangian $\mathscr{L}_{\text{matter}}(g,\varphi)$ that also depends on the metric and its first derivatives. The corresponding action functional is

$$S_{\text{matter}}(g,\varphi) = \frac{1}{c}\int_M \mathscr{L}_{\text{matter}}(g,\varphi)\sqrt{-g}\,d^4x,$$

where we tacitly assuming that the integral is convergent. Since the Hilbert functional $S(g)$ is invariant under the action of the group $\mathcal{G}$ of diffeomorphisms of M, the functional S_{matter} should also be $\mathcal{G}$-invariant.

This is so for the following examples of the matter fields.

(1) A scalar field $\varphi : M \to \mathbb{R}$ with the Lagrangian

$$\mathscr{L} = \frac{1}{2}\left(g^{\mu\nu}\nabla_\mu\varphi\nabla_\nu\varphi - V(\varphi)\right).$$

Note that though for a scalar field $\nabla_\mu\varphi = \partial_\mu\varphi$, covariant derivatives are used for deriving Euler-Lagrange equations. We have

$$S(g,\varphi) = \int_M \mathscr{L}\sqrt{-g}\,d^4x = \frac{1}{2}\int_M (d\varphi\wedge\star d\varphi - V(\varphi)\star 1),$$

where $\star$ is the Hodge operator for the metric $g_{\mu\nu}$ on M, so $S(g,\varphi)$ is invariant under the action of $\mathcal{G}$ on $\varphi \in C^\infty(M,\mathbb{R})$.

(2) The electromagnetic field with the Lagrangian

$$\mathscr{L}_{EM} = -\frac{1}{16\pi}F_{\mu\nu}F^{\mu\nu} = -\frac{1}{16\pi}g^{\mu\alpha}g^{\nu\beta}F_{\mu\nu}F_{\alpha\beta},$$

where A is a connection in a principal U(1)-bundle over M and $F_{\mu\nu} = \nabla_\mu A_\nu - \nabla_\nu A_\mu = \partial_\mu A_\nu - \partial_\mu A_\nu$. We have

$$S_{EM}(g, A) = \int_M \mathscr{L}_{EM}\sqrt{-g}\, d^4x = -\frac{1}{8\pi}\int_M F \wedge \star F,$$

so $S_{EM}(g, A)$ is $\mathcal{G}$-invariant under the action of $\mathcal{G}$ on connections A by pullbacks on 1-forms.

(3) The Yang-Mills field with the Lagrangian

$$\mathscr{L}_{YM} = -\frac{1}{4}\langle F_{\mu\nu}, F^{\mu\nu}\rangle = -\frac{1}{4}g^{\mu\alpha}g^{\nu\beta}\langle F_{\mu\nu}, F_{\alpha\beta}\rangle,$$

where A is a connection in a principal G-bundle over M and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\mu A_\nu + [A_\mu, A_\nu]$. We have

$$S_{YM}(g, A) = \int_M \mathscr{L}_{YM}\sqrt{-g}\, d^4x = -\frac{1}{2}\int_M \langle F_A \wedge \star F_A\rangle,$$

so $S_{EM}(g, A)$ is $\mathcal{G}$-invariant under the action of $\mathcal{G}$ on connections A by pullbacks on 1-forms.

The following simple fact is fundamental.

Lemma 22.4. *Suppose that the functional S_{matter} is invariant under the $\mathcal{G}$-action (it is assumed that there is also a $\mathcal{G}$-action on the matter fields). Then the Noether current*

$$T_{\mu\nu} = \frac{2}{\sqrt{-g}}\frac{\delta S_{\text{matter}}}{\delta g^{\mu\nu}} \tag{22.7}$$

— the stress-energy tensor of the matter fields — satisfies conservation laws

$$\nabla_\mu T^\mu_\nu = 0$$

on the solutions of the classical equations of motion.

Note that for the Hilbert functional we have by Proposition 22.1 that $T_{\mu\nu} = 2G_{\mu\nu}$, where $G_{\mu\nu}$ is Einstein tensor (21.18), so the conservation law is the corollary (21.17) of the Bianchi identity and is satisfied for all metrics.

Proof. Let Φ_s be a one-parameter subgroup of $\mathcal{G}$,

$$X = \frac{d}{ds}\bigg|_{s=0}\Phi_s \in \mathrm{Vect}(M)$$

be the corresponding vector field, and $\mathcal{L}_X$ be the Lie derivative. It follows from formula (21.1) that

$$\begin{aligned}\mathcal{L}_X g_{\mu\nu} &= \partial_\lambda g_{\mu\nu}X^\lambda + g_{\alpha\nu}\partial_\mu X^\alpha + g_{\mu\beta}\partial_\nu X^\beta \\ &= \partial_\mu X_\nu + \partial_\nu X_\mu + (\partial_\lambda g_{\mu\nu} - \partial_\mu g_{\lambda\nu} - \partial_\nu g_{\mu\lambda})X^\lambda \\ &= \nabla_\mu X_\nu + \nabla_\nu X_\mu,\end{aligned}$$

so

$$\mathcal{L}_X g^{\mu\nu} = -\nabla^\mu X^\nu - \nabla^\nu X^\mu.$$

Using the invariance of S_{matter} under the one-parameter subgroup Φ_s, the Euler-Lagrange equations for the matter fields and the symmetry $T_{\mu\nu} = T_{\nu\mu}$, we obtain

$$\begin{aligned} 0 &= \int_M \frac{\delta S_{\text{matter}}}{\delta g^{\mu\nu}} \nabla_X g^{\mu\nu} d^4x \\ &= \frac{1}{2}\int_M T_{\mu\nu}(\nabla^\mu X^\nu + \nabla^\nu X^\mu)\sqrt{-g}\, d^4x \\ &= \int_M T_{\mu\nu}\nabla^\mu X^\nu \sqrt{-g}\, d^4x \\ &= \int_M \nabla_\mu (T^\mu_\nu X^\nu)\sqrt{-g}\, d^4x - \int_M (\nabla_\mu T^\mu_\nu) X^\nu \sqrt{-g}\, d^4x. \end{aligned}$$

It follows from (22.5) and the Stokes theorem that the first integral is zero, so

$$\int_M (\nabla_\mu T^\mu_\nu) X^\nu \sqrt{-g}\, d^4x = 0$$

for all $X \in \text{Vect}(M)$, and we obtain $\nabla_\mu T^\mu_\nu = 0$. □

We emphasize that thus defined stress-energy tensor $T_{\mu\nu}$ is symmetric. If the Lagrangian $\mathscr{L}_{\text{matter}}$ depends only on the metric $g_{\mu\nu}$ and its first derivatives, then

$$T_{\mu\nu} = \frac{2}{\sqrt{-g}}\left\{ \frac{\partial(\sqrt{-g}\,\mathscr{L}_{\text{matter}})}{\partial g^{\mu\nu}} - \frac{\partial}{\partial x^\lambda}\frac{\partial(\sqrt{-g}\,\mathscr{L}_{\text{matter}})}{\partial_\lambda g^{\mu\nu}} \right\}. \tag{22.8}$$

When Lagrangian $\mathscr{L}_{\text{matter}}$ depends only on the metric $g_{\mu\nu}$ and not on its derivatives, due to the relation

$$\frac{\partial g}{\partial g_{\mu\nu}} = g g^{\mu\nu},$$

formula (22.8) simplifies

$$T_{\mu\nu} = 2\frac{\partial \mathscr{L}_{\text{matter}}}{\partial g^{\mu\nu}} - g_{\mu\nu}\mathscr{L}_{\text{matter}}.$$

It can be shown that in the special case of Minkowski spacetime $\mathbb{R}^{1,3}$, this definition of stress-energy tensor coincides with the one defined earlier for (see (12.17)). For example, for a scalar field (22.7) gives

$$T_{\mu\nu} = \nabla_\mu\varphi\nabla_\nu\varphi - \frac{1}{2}g_{\mu\nu}(\nabla^\lambda\varphi\nabla_\lambda\varphi - V(\varphi)),$$

for the electromagnetic field we get

$$T_{\mu\nu} = \frac{1}{4\pi}\left(-g^{\alpha\beta}F_{\mu\alpha}F_{\nu\beta} + \frac{1}{4}g_{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}\right),$$

and for the Yang-Mills field

$$T_{\mu\nu} = -g^{\alpha\beta}\langle F_{\mu\alpha}, F_{\nu\beta}\rangle + \frac{1}{4}g^{\mu\nu}\langle F_{\alpha\beta}, F^{\alpha\beta}\rangle.$$

When M is a Minkowski spacetime with the Minkowski metric, we get formulas obtained in Example 12.2 in Chapter **??**, and in Sections 15.5, 18.3.

Remark 22.5. This approach can be used for a more geometric definition of the stress-energy tensor $T_{\mu\nu}$ of a relativistic field theory in the Minkowski spacetime M^4 with Minkowski metric $\eta_{\mu\nu}$. Namely, let $S(g)$ be the action of the theory in an external gravitational field on M, described by a metric $g_{\mu\nu}$. Then we have

$$T_{\mu\nu} = 2\frac{\delta S(g)}{\delta g^{\mu\nu}}\bigg|_{g_{\mu\nu}=\eta_{\mu\nu}}.$$

Finally, we can write down the total action of a gravitational field in the presence of a matter fields described by the Lagrangian $\mathscr{L}_{\text{matter}}(g,\varphi)$, which depends only on $g_{\mu\nu}$ and its first derivatives. It is, in the form used in physics textbooks,

$$S(g,\varphi) = -\frac{c^3}{16\pi G}\int_M R\sqrt{-g}\,d^4x + \frac{1}{c}\int_M \mathscr{L}_{\text{matter}}(g,\varphi)\sqrt{-g}\,d^4x, \tag{22.9}$$

and its Euler-Lagrange equations are Einstein equations with matter

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G}{c^4}T_{\mu\nu}.$$

Remark 22.6. Besides the fields, the matter in a spacetime could be a collection of particles like the dust, or a macroscopic body, described by the laws of statistical mechanics. The result is the following formula for the energy-momentum tensor of a macroscopic body:

$$T_{\mu\nu} = (p+\varepsilon)u_\mu u_\nu - pg_{\mu\nu},$$

where u_μ is velocity vector, p is the pressure and ε is the energy density of the body. Thus for macroscopic body with mass distribution $M(x)$ at zero pressure, discussed in Section 22.1, we have

$$T_{\mu\nu} = M(x)c^2 u_\mu u_\nu.$$

It should be noted that 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 the temperature. This equation must be given along with the Einstein equations.

22.4. Asymptotically flat spacetime

In physics, asymptotically flat spacetime M describes the situation when gravitational bodies and matter fields are located in a finite space. The simplest example is when $M = \mathbb{R}^4$ with the Lorentzian metic $g_{\mu\nu}(x)$, satisfying at fixed x^0 as $r = |\boldsymbol{x}| \to \infty$,

$$g_{\mu\nu}(x) = \eta_{\mu\nu} + O\left(\frac{1}{r}\right), \quad \partial_\sigma g_{\mu\nu}(x) = O\left(\frac{1}{r^2}\right) \quad \text{and} \quad \Gamma^\sigma_{\mu\nu}(x) = O\left(\frac{1}{r^2}\right), \tag{22.10}$$

where $\eta_{\mu\nu}$ is Minkowski metric. Corresponding condition on the stress-energy tensor of masses and matter fields is

$$T_{\mu\nu}(x) = O\left(\frac{1}{r^4}\right).$$

The subgroup $\mathcal{G}_0$ of the diffeomorphism group $\mathcal{G}$ that preserves asymptotically flat conditions (22.10) consists of the diffeomorphisms $\Phi(x)$ satisfying as $r \to \infty$,

$$\Phi^\mu(x) = \Lambda^\mu_\nu x^\nu + a^\mu + O\left(\frac{1}{r^2}\right),$$

$$\partial_\nu \Phi^\mu(x) = \Lambda^\mu_\nu + O\left(\frac{1}{r^2}\right), \quad \partial_\nu \partial_\sigma \Phi^\mu(x) = O\left(\frac{1}{r^{2+\delta}}\right)$$

for some $\delta > 0$, where $(\Lambda, a) \in \mathfrak{P}$, the Poincaré group. The group $\mathcal{G}_0$ has a normal subgroup $\mathcal{N}$ for which $\Lambda^\mu_\nu = \delta^\mu_\nu$ and $a^\mu = 0$, and

$$\mathcal{G}_0/\mathcal{N} = \mathfrak{P}.$$

Then it follows from estimates (22.10) that modified Lagrangian (22.6) satisfies

$$\mathscr{L} = O\left(\frac{1}{r^4}\right) \quad \text{as} \quad r \to \infty,$$

so the correct action for the asymptotically flat gravitational field is

$$S(g) = \int_{I\times\mathbb{R}^3} \mathscr{L} dx^0 d^3\boldsymbol{x}, \tag{22.11}$$

where I is a finite time interval. It is not difficult to show that $S(g)$ is invariant under the action of the Lie algebra $Lie(\mathcal{G}_0)$, and we leave this computation to the reader.

22.5. Exercises

Exercise 22.1. Prove formula (22.3) by direct computation, and by using Riemann normal coordinates.

Exercise 22.2. Prove formula (22.3).

Exercise 22.3. Derive equations of motion for the Lagrangians of matter fields in Section 22.3.

Chapter 23

Hamiltonian Formulation and Exact Solutions

Consider the gravitation field action for the asymptotically flat space

$$S(g) = \int_M \mathscr{L} d^4x,$$

where

$$\mathscr{L} = \Gamma^\sigma_{\nu\sigma}\frac{\partial\sqrt{-g}g^{\mu\nu}}{\partial x^\mu} - \Gamma^\mu_{\nu\sigma}\frac{\partial\sqrt{-g}g^{\nu\sigma}}{\partial x^\mu} + \sqrt{-g}g^{\mu\nu}(\Gamma^\lambda_{\mu\nu}\Gamma^\sigma_{\lambda\sigma} - \Gamma^\sigma_{\mu\lambda}\Gamma^\lambda_{\sigma\nu}) \tag{23.1}$$

is the modified Lagrangian (22.6). It contains 10 independent fields — components of the metric tensor $g_{\mu\nu}(x)$ — and is quadratic in the first derivatives of the metric. As was discussed in Section 22.4, the action is invariant under the action of the Lie algebra $Lie(\mathcal{G}_0)$, so Lagrangian $\mathscr{L}$ is singular, like in case of gauge theories. According to the Dirac formalism (see Chapter 6, Section 16.4 and Section 18.5) for the Hamiltonian formulation of the Einstein equations we need to rewrite (23.1) as a first order Lagrangian.

23.1. Palatini first order formalism

In this approach we use the same Lagrangian (23.1), but consider the metric tensor $g_{\mu\nu}$ on the spacetime M and a torsion-free connection $\nabla = d + A$ on the tangent bundle TM as independent fields. Since $\Gamma^\lambda_{\mu\nu} = (A_\mu)^\lambda_\nu$ are symmetric, $\Gamma^\lambda_{\mu\nu} = \Gamma^\lambda_{\nu\mu}$, it contains $50 = 10 + 40$ independent variables.

Consider the following functional, called the *Palatini action*,

$$S(g, \Gamma) = \int_M g^{\mu\nu} R_{\mu\nu}\sqrt{-g}\, d^4x,$$

where $R_{\mu\nu}$ is given by the formula (21.15),

$$R_{\mu\nu}=\frac{\partial\Gamma^{\lambda}_{\mu\nu}}{\partial x^{\lambda}}-\frac{\partial\Gamma^{\lambda}_{\mu\lambda}}{\partial x^{\nu}}+\Gamma^{\lambda}_{\mu\nu}\Gamma^{\sigma}_{\lambda\sigma}-\Gamma^{\sigma}_{\mu\lambda}\Gamma^{\lambda}_{\sigma\nu}.$$

Regardless of the convergent of the integral (according to Section 22.4, for asymptotically flat spacetime the integral of $\mathscr{L}$ over M is convergent), we define its variational derivatives as in Section 22.2. Then since Lagrangian (23.1) and $\sqrt{-g}\,g^{\mu\nu}R_{\mu\nu}$ differ by a total divergence,

$$\int_D g^{\mu\nu}R_{\mu\nu}\sqrt{-g}\,d^4x=\int_D\mathscr{L}d^4x.$$

We have the following result.

Proposition 23.1. *Euler-Lagrange equations for the Palatini action are Einstein equations.*

Proof. Since $\Gamma^{\lambda}_{\mu\nu}$ are symmetric, variation of the tensor $R_{\mu\nu}$ is given by

$$\delta R_{\mu\nu}=\frac{\delta\Gamma^{\sigma}_{\mu\nu}}{\partial x^{\sigma}}-\frac{\delta\Gamma^{\sigma}_{\mu\sigma}}{\partial x^{\nu}}+\delta\Gamma^{\lambda}_{\mu\nu}\Gamma^{\sigma}_{\lambda\sigma}-\delta\Gamma^{\sigma}_{\mu\lambda}\Gamma^{\lambda}_{\sigma\nu}+\Gamma^{\lambda}_{\mu\nu}\delta\Gamma^{\sigma}_{\lambda\sigma}-\Gamma^{\sigma}_{\mu\lambda}\delta\Gamma^{\lambda}_{\sigma\nu},\tag{23.2}$$

whereas variation of $\sqrt{-g}$ is given by the formula (22.2),

$$\delta(\sqrt{-g})=-\frac{1}{2}\sqrt{-g}\,g_{\mu\nu}\delta g^{\mu\nu}.$$

Denoting $R=g^{\mu\nu}R_{\mu\nu}$, we readily obtain

$$\begin{aligned}\delta S_{\mathrm{P}}&=\int_D\left(R_{\mu\nu}\delta g^{\mu\nu}+g^{\mu\nu}\delta R_{\mu\nu}+R\frac{\delta(\sqrt{-g})}{\sqrt{-g}}\right)\sqrt{-g}\,d^4x\\&=\int_D\left(R_{\mu\nu}-\frac{1}{2}g_{\mu\nu}R\right)\delta g^{\mu\nu}\sqrt{-g}\,d^4x+\int_M g^{\mu\nu}\delta R_{\mu\nu}\sqrt{-g}\,d^4x\\&=\int_D\left(\left(R_{\mu\nu}-\frac{1}{2}g_{\mu\nu}R\right)\delta g^{\mu\nu}+Q^{\mu\nu}_{\lambda}\delta\Gamma^{\lambda}_{\mu\nu}\right)\sqrt{-g}\,d^4x.\end{aligned}$$

Using (23.2) and Stokes' theorem, we can transform the last integral as

$$\int_M g^{\mu\nu}\delta R_{\mu\nu}\sqrt{-g}\,d^4x=\int_M Q^{\mu\nu}_{\lambda}\delta\Gamma^{\lambda}_{\mu\nu}\sqrt{-g}\,d^4x,$$

where

$$\begin{aligned}Q^{\mu\nu}_{\lambda}=&-\frac{1}{\sqrt{-g}}\frac{\partial(\sqrt{-g}g^{\mu\nu})}{\partial x^{\lambda}}+g^{\mu\nu}\Gamma^{\sigma}_{\lambda\sigma}-g^{\mu\sigma}\Gamma^{\nu}_{\lambda\sigma}-g^{\nu\sigma}\Gamma^{\mu}_{\lambda\sigma}\\&+\delta^{\nu}_{\lambda}\left(\frac{1}{\sqrt{-g}}\frac{\partial(\sqrt{-g}g^{\mu\sigma})}{\partial x^{\sigma}}+g^{\rho\sigma}\Gamma^{\mu}_{\rho\sigma}\right).\end{aligned}$$

Thus equation $\delta S=0$ yileds

$$R_{\mu\nu}-\frac{1}{2}g_{\mu\nu}R=0\quad\text{and}\quad Q^{\mu\nu}_{\lambda}=0.$$

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 the definition of covariant derivative,

$$\nabla_{\lambda}g^{\mu\nu}=\frac{\partial g^{\mu\nu}}{\partial x^{\lambda}}+\Gamma^{\mu}_{\lambda\sigma}g^{\sigma\nu}+\Gamma^{\nu}_{\lambda\sigma}g^{\mu\sigma},$$

we can rewrite equation $Q_\lambda^{\mu\nu} = 0$ as

$$- \nabla_\lambda g^{\mu\nu} + \frac{1}{2} g^{\mu\nu} g_{\sigma\rho} \nabla_\lambda g^{\sigma\rho} + \delta_\lambda^\nu \left(\nabla_\sigma g^{\mu\sigma} - \frac{1}{2} g^{\mu\alpha} g_{\sigma\rho} \nabla_\alpha g^{\sigma\rho} \right) = 0. \tag{23.3}$$

Equation (23.3) has free indices λ, μ and ν. Putting $\lambda = \nu$ and summing over ν gives

$$-\nabla_\nu g^{\mu\nu} + \frac{1}{2} g^{\mu\nu} g_{\sigma\rho} \nabla_\nu g^{\sigma\rho} + 4 \left(\nabla_\sigma g^{\mu\sigma} - \frac{1}{2} g^{\mu\alpha} g_{\sigma\rho} \nabla_\alpha g^{\sigma\rho} \right) = 0,$$

whence

$$\nabla_\nu g^{\mu\nu} = \frac{1}{2} g^{\mu\nu} g_{\sigma\rho} \nabla_\nu g^{\sigma\rho}.$$

Back substituting this formula to (23.3) gives

$$\nabla_\lambda g^{\mu\nu} = \frac{1}{2} g^{\mu\nu} g_{\sigma\rho} \nabla_\lambda g^{\sigma\rho}. \tag{23.4}$$

Contracting equation (23.4) with $g_{\mu\nu}$ and using $g_{\mu\nu} g^{\mu\nu} = 4$, we get

$$g_{\sigma\rho} \nabla_\lambda g^{\sigma\rho} = 0,$$

and putting it back to (23.4) we finally obtain

$$\nabla_\lambda g^{\mu\nu} = 0.$$

This shows that ∇ is the Levi-Civita connection. Thus in the Palatini formalism equations (21.7) for the Christoffel's symbols appear from the principle of the least action. □

23.2. Hamiltonian formalism in general relativity

Here for simplicity we consider only the case of globally hyperbolic spacetime M, discussed in Section 21.1, and put $c = 1$. We leave it to the reader to fill all necessary details for the general case.

By Theorem 21.2, we can always use syncrhonous coordinates and assume that the metric in the Palatini Lagrangian density $\mathscr{L}$ has the form

$$ds^2 = g_{00}(x)(dx^0)^2 + g_{ij}(x) dx^i dx^j,$$

where $g_{00} > 0$ and the matrix $-g_{ij}$ is positive-definite. It follows from (23.1) that it is convenient instead of the metric tensor $g_{\mu\nu}$ to use the matrix $h^{\mu\nu} = \sqrt{-g}\, g^{\mu\nu}$, contravariant tensor density of weight 1 of the metric trnsor $g_{\mu\nu}$, as independent fields. We have

$$h^{00} = \frac{\sqrt{-g}}{g_{00}}, \quad h^{ij} = \sqrt{-g}\, g^{ij} \quad \text{and} \quad h^{0i} = 0.$$

Let N be a Cauchy hypersurface in M, given by the equation $x^0 = const$. Consider the Lagrangian

$$L = \frac{1}{2} \int_N \mathscr{L} d^3\boldsymbol{x} = \frac{1}{2} \int_N \left(\Gamma_{\nu\sigma}^\sigma \partial_\mu h^{\mu\nu} - \Gamma_{\nu\sigma}^\mu \partial_\mu h^{\nu\sigma} + h^{\mu\nu} (\Gamma_{\mu\nu}^\lambda \Gamma_{\lambda\sigma}^\sigma - \Gamma_{\mu\lambda}^\sigma \Gamma_{\sigma\nu}^\lambda) \right) d^3\boldsymbol{x},$$

and collect the terms containing time derivative ∂_0:

$$\Gamma_{\nu\sigma}^\sigma \partial_0 h^{0\nu} - \Gamma_{\nu\sigma}^0 \partial_0 h^{\nu\sigma} = \Gamma_{0i}^i \partial_0 h^{00} - \Gamma_{ik}^0 \partial_0 h^{ik}. \tag{23.5}$$

Now observe that Lagrangian $\mathscr{L}$ linearly depends on coefficients Γ^{μ}_{00}, which can be considered as Lagrange multipliers. These terms are

$$\Gamma^0_{00}(h^{00}\Gamma^i_{0i} + h^{ik}\Gamma^0_{ik}) - \Gamma^i_{00}(\partial_i h^{00} - h^{00}(\Gamma^k_{ik} - \Gamma^0_{0i}))$$

and lead to constraints

$$h^{00}\Gamma^i_{0i} + h^{ik}\Gamma^0_{ik} = 0 \quad \text{and} \quad \partial_i h^{00} - h^{00}(\Gamma^k_{ik} - \Gamma^0_{0i}) = 0. \tag{23.6}$$

Substituting $\Gamma^i_{0i} = -\Gamma^0_{ik}h^{ik}/h^{00}$ to (23.5), we see that the time derivative terms can be written as

$$-\frac{\Gamma^0_{ik}}{h^{00}}\,\partial_0(h^{00}h^{ik}), \tag{23.7}$$

which suggests that natural dynamical variables are $q^{ik} = -h^{00}h^{ik}$ and $\pi_{ik} = \Gamma^0_{ik}/h^{00}$. Note that there are 6 independent fields q^{ik}, where $1 \leq i \leq k \leq 3$, and 6 conjugated momenta π_{ik} Thus kinetic term in the Lagrangian takes a canonical form

$$\frac{1}{2}\pi_{ik}\partial_0 q^{ik}.$$

Restriction of the matrix q^{ik} is a contravariant tensor density of weight 2 of the metric tensor g_{ik} on N and

$$q^{ij}g_{jk} = \gamma\delta^i_k, \quad \gamma = \det g_{ik}.$$

Since $g = h^{00}\gamma$ and h^{00} is a scalar on N, the matrix Γ^0_{ik} is a covariant symmetric tensor density of weight -1 on N.

Coefficients $\Gamma^{\lambda}_{\mu\nu}$ that are different from Γ^0_{ik} are non-dynamical and give additional constraints

$$\frac{\partial\mathscr{L}}{\partial\Gamma^{\lambda}_{\mu\nu}} = 0,$$

which can be written explicitly as

$$\begin{gathered} h^{00}\Gamma^i_{0k} + h^{ij}\Gamma^0_{jk} = 0, \\ \partial_k h^{ij} + h^{il}\Gamma^j_{lk} + h^{jl}\Gamma^i_{lk} - h^{ij}\Gamma^{\mu}_{k\mu} = 0. \end{gathered} \tag{23.8}$$

The solution of equations (23.6) and (23.8) is

$$\begin{gathered} \Gamma^i_{kl} = \gamma^i_{kl}, \\ \Gamma^i_{j0} = -\frac{1}{h^{00}}h^{il}\Gamma^0_{jl} = \frac{1}{h^{00}}q^{il}\pi_{jl}, \\ \Gamma^0_{i0} = \gamma^j_{ij} - \frac{\partial_i h^{00}}{h^{00}}, \end{gathered} \tag{23.9}$$

where γ^i_{jk} are Christoffel's symbols of the Levi-Civita connection of the Riemannian metric $\gamma_{ij}(\boldsymbol{x}) = -g_{ij}(\boldsymbol{x})$ on the Cauchy hypersurface N.

Substituting formulas (23.9) back to the Lagrangian L, we finally obtain

$$L = \int_N \left(\frac{1}{2}\pi_{ik}\partial_0 q^{ik} - H(\boldsymbol{x}) - \lambda(\boldsymbol{x})C_0(\boldsymbol{x})\right) d^3\boldsymbol{x}, \tag{23.10}$$

so according to the Dirac formalism, $H(\boldsymbol{x})$ is the Hamiltonian density, $\lambda(\boldsymbol{x})$ are Lagrange multipliers, and $C_0(\boldsymbol{x})$ are constraints. Explicitly,

$$C_0(\boldsymbol{x}) = \frac{1}{2}(q^{ij}q^{kl}(\pi_{ik}\pi_{jl} - \pi_{ij}\pi_{kl}) - \gamma R_3), \quad \gamma = -\det\gamma_{ij}, \tag{23.11}$$

where R_3 is the scalar curvature of the Riemannian metric γ_{ij} on N, $\lambda = 1 + 1/h^{00}$ and

$$H(\boldsymbol{x}) = C_0(\boldsymbol{x}) - \frac{1}{2}\partial_i\partial_k q^{ik}(\boldsymbol{x}). \tag{23.12}$$

Constraints $C_0(\boldsymbol{x})$ have a nice geometric meaning: according to generalized Gauss *Theorema Egregium*, the right hand side of (23.11) is the formula for the restriction of the scalar curvature R_4 of the metric $g_{\mu\nu}$ to the hypersurface N. Thus $C_0(\boldsymbol{x}) = 0$ gives Einstein equations in the empty spacetime, so

$$H(\boldsymbol{x}) = -\frac{1}{2}\partial_i\partial_k q^{ik}(\boldsymbol{x}).$$

Thus the Hamiltonian density is a total divergence, so one can erroneously conclude that by the Stokes theorem, the total energy of a gravitational field

$$E = \int_N H(\boldsymbol{x})d^3\boldsymbol{x} \tag{23.13}$$

is zero. However, this is not so for the asymptotically flat spacetime. Namely specializing (22.10) to the metric tensor,

$$g_{00} = 1, \quad g_{0i} = 0, \quad g_{ij} = -\delta_{ij}\left(1 + \frac{M}{4\pi r}\right) + O\left(\frac{1}{r^2}\right) \quad \text{as} \quad r \to \infty, \tag{23.14}$$

we obtain

$$q^{ij}(x) = \delta^{ij}\left(1 + \frac{M}{2\pi r}\right) + O\left(\frac{1}{r^2}\right),$$

so

$$E = -\lim_{R\to\infty}\frac{1}{2}\int_{|\boldsymbol{x}|\le R}\partial_i\partial_k q^{ik}(\boldsymbol{x})d^3\boldsymbol{x} = -\lim_{R\to\infty}\frac{M}{4\pi}\int_{|\boldsymbol{x}|=R}\frac{\partial}{\partial r}\frac{1}{r}dS = M.$$

Remark 23.2. As in physics textbooks, we can easily restore the dependence on c by introducing the coefficient $1/\kappa$ in the definition (23.13), where κ is the Einstein constant. Defining the leading term of the asymptotically flat metric (23.14) to be $-\delta_{ij}\left(1 + \frac{\kappa M c^2}{4\pi r}\right)$, we obtain the relativistic relation $E = Mc^2$.

23.3. The Schwarzschild solution

Consider the case of empty spacetime $M = \mathbb{R}^4$ with static (time independent) spherically symmetric metric

$$ds^2 = g_{00}(r)c^2dt^2 - g_{11}(r)dr^2 - r^2(d\theta^2 + \sin^2\theta\, d\varphi^2),$$

written in the 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 of radius R, on the assumption that the electric charge of the mass and angular momentum of the mass are all zero.

Computing Christoffel's symbols $\Gamma^{\lambda}_{\mu\nu}$, where $x^0 = ct, x^1 = r, x^2 = \theta, x^3 = \varphi$, and solving Einstein equations $R_{\mu\nu} = 0$, we obtain

$$g_{00}(r) = 1 - \frac{a}{r}, \quad g_{11} = \frac{1}{1 - \frac{a}{r}},$$

where a is a constant. Thus

$$ds^2 = \left(1 - \frac{a}{r}\right) c^2 dt^2 - \frac{dr^2}{1 - \frac{a}{r}} - r^2 d\Omega^2,$$

where $d\Omega^2$ is the induced metric on $S^2 \subset \mathbb{R}^3$. In the limit $r \to \infty$ we should have

$$g_{\mu\nu} = \eta_{\mu\nu} + \frac{1}{c^2} g^{(2)}_{\mu\nu} + O\left(\frac{1}{c^3}\right),$$

so

$$g^{(2)}_{00} = -\frac{ac^2}{r} = -\frac{2MG}{r},$$

where M is the mass of a body creating gravitational field. By definition, the quantity

$$a = \frac{2MG}{c^2}$$

is called the *Schwarzschild radius* and is denoted by r_s[1].

Thus the Schwarzschild metric is

$$ds^2 = \left(1 - \frac{r_s}{r}\right) c^2 dt^2 - \frac{dr^2}{1 - \frac{r_s}{r}} - r^2 d\Omega_2^2, \tag{23.15}$$

and this formula 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*.

According to Jebsen-Birkhoff's theorem, the Schwarzschild metric is the most general spherically symmetric solution of the Einstein equations in empty spacetime $\mathbb{R}^4$.

23.4. Cosmological constant and AdS and dS spacetimes

One can add to the Hilbert action a multiple of a volume form of the spacetime. The resulting action is

$$S = \int_M (R - 2\Lambda)\sqrt{-g}\, d^4x,$$

where Λ is called the *cosmological constant*. The minus sign in the action comes from thinking of the Lagrangian as "$T - V$": the cosmological constant playing the role of potential energy V. It immediately follows from the proof of Proposition 22.1, that the Euler-Lagrange equations for this action are

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = -\Lambda g_{\mu\nu}.$$

[1] For the Earth $r_s = 0.8.9$ mm, while for the Sun $r_s = 3$ km.

Taking the trace, we get $R = 4\Lambda$, so substituting this back we obtain

$$R_{\mu\nu} = \Lambda g_{\mu\nu}$$

— the Einstein equations in empty spacetime in the presence of a cosmological constant.

The basic examples of such spacetimes are provided by Lorentzian *maximally symmetric spaces*. This means that in addition to a transitive action of the isometry group on the space M, its differential at every $p \in M$ acts transitively on T_pM.

The n-dimensional *de Sitter space* dS_n is a symmetric space $\mathrm{O}(1,n)/\mathrm{O}(1,n-1)$. It can be realized as a one sheet hyperboloid in the Minkowski space $\mathbb{R}^{1,n}$,

$$(x^0)^2 - \sum_{i=1}^{n}(x^i)^2 = -R^2, \tag{23.16}$$

with the Lorentzian metric $g_{\mu\nu}$ induced by the Minkowski metric[2]. Topologically, the de Sitter space is isomorphic to $\mathbb{R} \times S^{n-1}$. Its isometry group is $\mathrm{O}(1,n)$, so dS_n is maximally symmetric space. Its is easy to compute its curvature tensor and to show that

$$R_{\mu\nu} = \frac{n-1}{2R^2} g_{\mu\nu}. \tag{23.17}$$

Remark 23.3. Replacing negative $-R^2$ by positive R^2 in (23.20), one gets a hyperboloid of two sheets. The negative of the induced metric is positive-definite, and each sheet is a copy of a hyperbolic n-space.

Considering dS_4 as the spacetime, we see that its metric $g_{\mu\nu}$ satisfies Einstein equations with positive cosmological constant $\Lambda = 3/R^2$. Using global coordinates on dS_4

$$x^0 = R\sinh\left(\frac{ct}{R}\right), \quad x^i = R\cosh\left(\frac{ct}{R}\right)n^i, \quad i = 1,2,3,4,$$

where $\boldsymbol{n} = (n^1, n^2, n^3, n^4) \in S^3$, de Sitter metric can be written as

$$ds^2 = c^2dt^2 - R^2\cosh^2\left(\frac{ct}{R}\right)d\Omega_3^2, \tag{23.18}$$

where $d\Omega_3^2$ is the round metric on the unit sphere S^3.

Thus we see that de Sitter metric is a time-dependent solution of the Einstein equations that describes the expanding spacetime. Indeed, the space — topological 3-sphere S^3 — first shrinks to a minimal sphere with the radius R, and then subsequently expands as $t \to \infty$.

It is also customary to use coordinates

$$x^0 = \sqrt{R^2 - r^2}\sinh\left(\frac{ct}{R}\right), \quad x^4 = \sqrt{R^2 - r^2}\cosh\left(\frac{ct}{R}\right),$$

where $r^2 = |\boldsymbol{x}|^2$ and $\boldsymbol{x} = (x^1, x^2, x^3)$ are Euclidean coordinates in $\mathbb{R}^3$. They are local coordinates in a coordinate patch $r < R$ and $x^4 > 0$, and in these coordinates the de Sitter metric is *static*,

$$ds^2 = \left(1 - \frac{r^2}{R^2}\right)c^2dt^2 - \left(1 - \frac{r^2}{R^2}\right)^{-1}dr^2 - r^2d\Omega_2^2. \tag{23.19}$$

---

[2] Here $R > 0$ is a constant and not a scalar curvature.

In general relativity these coordinates are called *static coordinates*, since the metric tensor $g_{\mu\nu}$ does not depend on t. It is instructive to compare the static solution (23.19) for $\Lambda > 0$ with the Schwarzschild solution (23.15) for $\Lambda = 0$.

The n-dimensional *anti de Sitter space* AdS_n is a symmetric space $\mathrm{O}(2, n-1)/\mathrm{O}(1, n-1)$. It can be realized as the following hyperquadric in $\mathbb{R}^{n+1}$,

$$(x^0)^2 + (x^1)^2 - \sum_{i=2}^{n}(x^i)^2 = -R^2, \tag{23.20}$$

with the Lorentzian metric $g_{\mu\nu}$ induced by the following pseudo-Riemannian metric on $\mathbb{R}^{n+1}$

$$ds^2 = (dx^0)^2 + (dx^1)^2 - \sum_{i=2}^{n}(dx^i)^2.$$

Topologically, the anti de Sitter space is isomorphic to $S^1 \times \mathbb{R}^{n-1}$, so its time direction is compact. The isometry group is $\mathrm{O}(2, n-1)$, so AdS_n is a maximally symmetric space. It is easy to compute its curvature tensor and to show that

$$R_{\mu\nu} = -\frac{n-1}{2R^2}g_{\mu\nu}. \tag{23.21}$$

Considering AdS_4 as the spacetime, we see that its metric $g_{\mu\nu}$ satisfies Einstein equations with negative cosmological constant $\Lambda = -3/R^2$. Using global coordinates on AdS_4

$$x^0 = R\cosh\rho\,\sin\left(\frac{ct}{R}\right), \quad x^1 = R\cosh\rho\,\cos\left(\frac{ct}{R}\right), \quad x^i = R\sinh\rho\, n^i, \quad i = 2, 3, 4,$$

where $\boldsymbol{n} = (n^2, n^3, n^4) \in S^2$, anti de Sitter metric can be written as

$$ds^2 = \cosh^2\!\rho\, c^2dt^2 - R^2dr^2 - R^2\sinh^2\!\rho\, d\Omega_2^2. \tag{23.22}$$

Introducing $r = R\sinh\rho$, we can rewrite (23.22) as

$$ds^2 = \left(1 + \frac{r^2}{R^2}\right)c^2dt^2 - \left(1 + \frac{r^2}{R^2}\right)^{-1}dr^2 - r^2d\Omega_2^2. \tag{23.23}$$

Though for several reasons the anti de Sitter space cannot be considered as a model for the physical spacetime, it serves as a useful mathematical tool.

To summarize, we have seen that Schwarzschild, de Sitter and anti de Sitter metrics are simplest solutions of the Einstein equations in the empty spacetime for cosmological constant Λ being, respectfully, zero, positive and negative.

23.5. 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 spacetime $\mathcal{M} = M \times S_r^1$, where the fifth dimension in the circle of a very small radius

$$r = \sqrt{\frac{\hbar G}{c^3}} \sim 10^{-35}\, m$$

— the Planck's length ℓ_{P}, where $\hbar$ is the Planck constant.

Denote the coordinates on $\mathcal{M}$ 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

$$d\tilde{s}^2 = \tilde{g}_{ab}d\tilde{x}^a\tilde{x}^b = g_{\mu\nu}dx^\mu dx^\nu - (A_\mu dx^\mu - d\theta)^2,$$

and 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.

It is easy to establish the following basic properties.

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!

From (21.7) we obtain the following formulas for Christoffel's symbols for the metric $\tilde{g}_{ab}$:

$$\begin{aligned} &\tilde{\Gamma}^\mu_{\alpha\beta} = \Gamma^\mu_{\alpha\beta} + \frac{1}{2}g^{\mu\sigma}(A_\alpha F_{\sigma\beta} + A_\beta F_{\sigma\alpha}), \\ &\tilde{\Gamma}^\mu_{\alpha 4} = \frac{1}{2}g^{\mu\sigma}F_{\alpha\sigma}, \\ &\tilde{\Gamma}^4_{\alpha\beta} = A_\mu\Gamma^\mu_{\alpha\beta} - \frac{1}{2}\left(A^\mu(A_\alpha F_{\beta\mu} + A_\beta F_{\alpha\mu}) - \frac{\partial A_\alpha}{\partial x^\beta} - \frac{\partial A_\beta}{\partial x^\alpha}\right), \\ &\tilde{\Gamma}^4_{\alpha 4} = \frac{1}{2}A^\mu F_{\alpha\mu}, \\ &\tilde{\Gamma}^a_{44} = 0, \end{aligned} \tag{23.24}$$

where

$$F_{\alpha\beta} = \frac{\partial A_\beta}{\partial x^\alpha} - \frac{\partial A_\alpha}{\partial x^\beta}.$$

Now consider a free particle of mass m on the five-dimensional spacetime $\mathcal{M}$ with the action

$$S = -mc\int d\tilde{s} = -mc\int\sqrt{\tilde{g}_{ab}\frac{d\tilde{x}^a}{d\tilde{s}}\frac{d\tilde{x}^b}{d\tilde{s}}}\,d\tilde{s}. \tag{23.25}$$

The corresponding equations of motion are geodesic equations with respect to the natural parameter $\tilde{s}$,

$$\frac{du^a}{d\tilde{s}} + \tilde{\Gamma}^a_{bc} u^b u^c = 0, \quad \text{where} \quad u^a = \frac{d\tilde{x}^a}{d\tilde{s}}. \tag{23.26}$$

Using formulas (23.24), we obtain from (23.26)

$$\frac{d}{d\tilde{s}}(u^4 - A_\mu u^\mu) = 0, \tag{23.27}$$

so $u^4 - A_\mu u^\mu = \xi$ is a constant. Observing that 5-vector u^a has unit length,

$$1 = g_{\mu\nu} u^\mu u^\nu + (u^4 - A_\mu u^\nu)^2,$$

we obtain $g_{\mu\nu} u^\mu u^\nu = 1 - \xi^2$, so

$$\frac{ds}{d\tilde{s}} = \sqrt{1 - \xi^2}.$$

Whence

$$\frac{dx^\mu}{ds} = u^\mu \frac{d\tilde{s}}{ds} = \frac{u^\mu}{\sqrt{1-\xi^2}}$$

and form (23.26) we obtain

$$\frac{d^2 x^\mu}{ds^2} + \Gamma^\mu_{\alpha\beta} \frac{dx^\alpha}{ds} \frac{dx^\beta}{ds} = -\frac{\xi}{\sqrt{1-\xi^2}} g^{\mu\sigma} F_{\alpha\sigma} \frac{dx^\alpha}{ds}. \tag{23.28}$$

Finally, putting

$$\xi = \frac{e}{\sqrt{m^2 c^4 + e^2}}$$

we see that the right hand side of (23.28) becomes

$$\frac{e}{mc^2} g^{\mu\sigma} F_{\alpha\sigma} \frac{dx^\alpha}{ds}.$$

Thus from the action (23.25) for a free particle moving on a 5-dimensional spacetime $\mathcal{M}$ we obtained the equations of motion for a charged particle, moving on 4-dimensional spacetime M in external gravitational and magnetic fields! These equations are Euler-Lagrange equations for the action

$$-mc \int ds - \frac{e}{c} \int A_\mu dx^\mu,$$

so the action (23.25) in five dimensions unifies electromagnetic and gravitational actions in four dimensions. This remarkable result is the so-called *first Kaluza miracle.*

Next, by a direct and rather lengthy computation we get the following relation between scalar curvatures of the metrics on $\tilde{M}$ and on M

$$\tilde{R} = R + \frac{1}{4} F_{\mu\nu} F^{\mu\nu}, \tag{23.29}$$

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 $\tilde{G} = 2\pi r G$, replacing A_μ by $\varkappa A_\mu$, where $\varkappa = 2\sqrt{G}/c^2$, and trivially integrating over S^1_r we 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

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

on M in the presence of the gravitational 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 spacetime M.

Though mathematically very elegant, Kaluza-Klein theory is not physical: it gives unrealistic predictions for the masses of particles. Thus 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 $\Box_4 = \partial^\mu\partial_\mu$ for the Minkowski metric. Corresponding coefficients of the Fourier series expansion

$$\Phi(x,\theta) = \sum_{n=-\infty}^{\infty}\varphi_n(x)e^{\frac{in\theta}{r}}$$

satisfy Klein-Gordon equations

$$(\Box_4 + m_n^2)\varphi_n = 0$$

with the masses

$$m_n^2 = \frac{n^2}{r^2}.$$

However, these masses are very large! Thus assuming that $n = 1$ corresponds to the electron, the obtained mass would $m_e \sim 3{\cdot}10^{30}$ MeV, while the actual electron mass is only 0.5 MeV.

23.6. Exercises

Exercise 23.1. Prove formulas (23.9) and (23.10)–(23.12).

Exercise 23.2. Apply Dirac formalism for a Hamiltonian formulation of Einstein equations for a general asymptotically flat spacetime, and find a geometric interpretation of corresponding constraints.

Exercise 23.3. Derive the formula for the Schwarzschild solution in Section 23.3.

Exercise 23.4. Compute the Riemann curvature tensor for the de Sitter and anti de Sitter spaces dS_n and AdS_n, and prove equations (23.17) and (23.21).

Exercise 23.5. Derive formulas (23.24).

Exercise 23.6. Using (23.24), rewrite equations (23.26) as

$$\frac{du^\mu}{d\tilde{s}} + \Gamma^\mu_{\alpha\beta} u^\alpha u^\beta = -g^{\mu\sigma} A_\alpha F_{\sigma\beta} u^\alpha u^\beta - g^{\mu\sigma} F_{\alpha\sigma} u^\alpha u^4,$$

$$\frac{du^4}{d\tilde{s}} + A_\mu \Gamma^\mu_{\alpha\beta} u^\alpha u^\beta = -A^\sigma F_{\alpha\sigma} u^\alpha u^4 + A^\sigma A_\alpha F_{\beta\sigma} u^\alpha u^\beta + \frac{\partial A_\alpha}{\partial x^\beta} u^\alpha u^\beta$$

and derive equations (23.27)–(23.28).

Exercise 23.7. Prove the second Kaluza miracle — relation (23.29).

Notes and References

There are many excellent textbooks and monographs on general relativity. For basic physics principles, we refer to classic texts [**LL1971**], [**HE1973**], and to unique textbook [**Zee2013**], written from the point of view of the action principle. More mathematically oriented sources include the monograph [**Wal1984**] and the textbook [**CB2015**]. The necessary basic material on pseudo-Riemannian geometry can be found in [**O'N1983**]; see also [**DFN1984**, **DFN1985**] for a succinct introduction to the field.

The proof of Theorem 21.2 about existence of Lorentzian metrics can be found in [**O'N1983**, p. 149].

Our definition of globally hyperbolic spacetime in Section 21.1 follows [**HE1973**]. For the proof of Theorem 21.9 and other results, we refer the reader to [**BS2005**, **BS2006**]; in [**BS2007**] it was proved that in Definition 21.8 of a globally hyperbolic spacetime the strong causality condition can be replaced by the causality condition. The initial value formulation of Einstein equations is discussed in [**Wal1984**]; for the definition of the energy-momentum tensor of the macroscopic body, we refer to [**LL1971**].

Regarding the Hilbert-Einstein action functional S in Section 22.2, we remark that if one fixes only the values of a metric tensor $g_{\mu\nu}$ on ∂D, then its variation δS will contain the boundary term. It is quite remarkable that one can add to S the so-called *Gibbons-Hawking-York (GHY) boundary term* [**Yor1972**, **GH1977**], so that δS will be still given by Proposition 22.1. The GHY boundary term is the integral over ∂D with respect to the volume form of the trace of a second fundamental form of the induced metric on ∂D.

The Hamiltonian approach to general relativity was developed by Dirac and was formulated in geometric terms by R. Arnowitt, S. Deser and C. W. Misner in [**ADM1959**], nowadays called the *Arnowitt-Deser-Misner (ADM) formalism*. Our exposition in Section 23.2 is a simplified version of L.D. Faddeev elegant formulation of the ADM formalism, and we refer to [**Fad1982**] for details and further references. Positivity of mass for asymptotically flat spacetime (the ADM mass) is a non-trivial result, proved by R. Schoen and S. T. Yau [**SY1979**] and subsequently simplified by Witten [**Wit1981**] (see also an exposition in [**Fad1982**]).

In Sections 23.3 and 23.4 we present simplest solutions of the vacuum Einstein equations for zero, positive and negative values of the cosmological constant — the Schwarzschild metric, the dS metric and the AdS metric. We refer to the textbook [**Zee2013**] for a detailed discussion of the properties of these solutions and of the Jebsen-Birkhoff theorem.

In Section 23.5 we briefly discussed Kaluza-Klein theory, a beautiful though unsuccessful attempt to the unified field theory, which is an important precursor to string theory. We refer the reader to the book [**Lee1984**] which contains English translation of original papers by T. Kaluza and O. Klein, and to [**Zee2013**] for a detailed exposition of Kaluza-Klein theory and its generalizations.

Bibliography

[AM1978] Ralph Abraham and Jerrold E. Marsden, *Foundations of mechanics*, Benjamin/Cummings Publishing Co., Inc., Advanced Book Program, Reading, Mass., 1978. Second edition, revised and enlarged.

[Are1971] R. Arens, *Classical relativistic particles*, Commun. Math. Phys. **21** (1971), 139–149.

[Arn1989] V. I. Arnol'd, *Mathematical methods of classical mechanics*, 2nd ed., Graduate Texts in Mathematics, vol. 60, Springer-Verlag, New York, 1989.

[AG1990] V.I. Arnold and A.B. Givental, *Symplectic geometry*, Dynamical systems IV, 1990, pp. 1-136.

[AKN1997] V.I. Arnold, V.V. Kozlov, and A.I. Neishtadt, *Mathematical aspects of classical and celestial mechanics*, Encyclopaedia of Mathematical Sciences, vol. 3, Springer-Verlag, Berlin, 1997.

[ADM1959] R. Arnowitt, S. Deser, and C. W. Misner, *Dynamical Structure and Definition of Energy in General Relativity*, Phys. Rev. **116** (1959), no. 5, 1322–1330.

[Ati1979] M. F. Atiyah, *Geometry on Yang-Mills fields*, Scuola Normale Superiore, Pisa, 1979.

[Ati1990] Michael Atiyah, *The geometry and physics of knots*, Lezioni Lincee. [Lincei Lectures], Cambridge University Press, Cambridge, 1990.

[AB1983] M. F. Atiyah and R. Bott, *The Yang-Mills equations over Riemann surfaces*, Philos. Trans. Roy. Soc. London Ser. A **308** (1983), no. 1505, 523–615.

[AHDM1978] M. F. Atiyah, N. J. Hitchin, V. G. Drinfel'd, and Yu. I. Manin, *Construction of instantons*, Phys. Lett. A **65** (1978), no. 3, 185–187.

[dAI1995] José A. de Azcárraga and José M. Izquierdo, *Lie groups, Lie algebras, cohomology and some applications in physics*, Cambridge Monographs on Mathematical Physics, Cambridge University Press, Cambridge, 1995.

[BM1994] John Baez and Javier P. Muniain, *Gauge fields, knots and gravity*, Series on Knots and Everything, vol. 4, World Scientific Publishing Co., Inc., River Edge, NJ, 1994.

[Bar1954] V. Bargmann, *On unitary ray representations of continuous groups*, Ann. of Math. (2) **59** (1954), 1–46.

[BPST1975] A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Yu. S. Tyupkin, *Pseudoparticle solutions of the Yang-Mills equations*, Phys. Lett. B **59** (1975), no. 1, 85–87.

[BS2005] Antonio N. Bernal and Miguel Sánchez, *Smoothness of time functions and the metric splitting of globally hyperbolic spacetimes*, Comm. Math. Phys. **257** (2005), no. 1, 43–50.

[BS2006] ———, *Further results on the smoothability of Cauchy hypersurfaces and Cauchy time functions*, Lett. Math. Phys. **77** (2006), no. 2, 183–197.

[BS2007] ———, *Globally hyperbolic spacetimes can be defined as 'causal' instead of 'strongly causal'*, Classical Quantum Gravity **24** (2007), no. 3, 745–749.

[BLOT1990] N. N. Bogolubov, A. A. Logunov, A. I. Oksak, and I. T. Todorov, *General principles of quantum field theory*, Mathematical Physics and Applied Mathematics, vol. 10, Kluwer Academic Publishers Group, Dordrecht, 1990.

[BS1983] N. N. Bogoliubov and D. V. Shirkov, *Quantum fields*, Benjamin/Cummings Publishing Co., Inc., Advanced Book Program, Reading, MA, 1983.

[Bry1995] R.L. Bryant, *An introduction to Lie groups and symplectic geometry*, Geometry and quantum field theory (Park City, UT, 1991), 1995, pp. 5–181.

[CS1974] Shiing Shen Chern and James Simons, *Characteristic forms and geometric invariants*, Ann. of Math. (2) **99** (1974), 48–69.

[CS1997] Piotr T. Chruściel and Jalal Shatah, *Global existence of solutions of the Yang-Mills equations on globally hyperbolic four-dimensional Lorentzian manifolds*, Asian J. Math. **1** (1997), no. 3, 530–548.

[CB2015] Yvonne Choquet-Bruhat, *Introduction to general relativity, black holes, and cosmology*, Oxford University Press, Oxford, 2015.

[Del1999] Pierre Deligne, *Notes on spinors*, Quantum fields and strings: a course for mathematicians, Vol. 1, 2 (Princeton, NJ, 1996/1997), Amer. Math. Soc., Providence, RI, 1999, pp. 99–135.

[DF1999] Pierre Deligne and Daniel S. Freed, *Classical field theory*, Quantum fields and strings: a course for mathematicians, Vol. 1, 2 (Princeton, NJ, 1996/1997), Amer. Math. Soc., Providence, RI, 1999, pp. 137–225.

[Dir1958] P.A.M. Dirac, *Generalized Hamiltonian Dynamics*, Proc. R. Soc. Lond. A **246** (1958), no. 16, 3755–3772.

[Don1983] S. K. Donaldson, *An application of gauge theory to four-dimensional topology*, J. Differential Geom. **18** (1983), no. 2, 279–315.

[DK1990] S. K. Donaldson and P. B. Kronheimer, *The geometry of four-manifolds*, Oxford Mathematical Monographs, The Clarendon Press, Oxford University Press, New York, 1990.

[DFN1984] B.A. Dubrovin, A.T. Fomenko, and S.P. Novikov, *Modern geometry — methods and applications. Part I*, Graduate Texts in Mathematics, vol. 93, Springer-Verlag, New York, 1984.

[DFN1985] ———, *Modern geometry — methods and applications. Part II*, Graduate Texts in Mathematics, vol. 104, Springer-Verlag, New York, 1985.

[EM1982] Douglas M. Eardley and Vincent 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), no. 2, 171–191.

[Fad1969] L.D. Faddeev, *The Feynman integral for singular Lagrangians*, Theoret. and Math. Phys **1** (1969), no. 1, 1–13.

[Fad1982] L. D. Faddeev, *The energy problem in Einstein's theory of gravitation*, Soviet Phys. Uspekhi **25** (1982), no. 3, 130–142.

[FS1984] L. D. Faddeev and S.L. Shatashvili, *Algebraic and Hamiltonian methods in the theory of non-Abelian anomalies*, Theoret. and Math. Phys **60** (1984), no. 2, 770–778.

[FS1991] L. D. Faddeev and A. A. Slavnov, *Gauge fields: An introduction to quantum theory*, Frontiers in Physics, vol. 83, Addison-Wesley, Redwood City, CA, 1991.

[Fad1998] L.D. Faddeev, *A mathematician's view of the development of physics*, Les relations entre les mathématiques et la physique théorique, 1998, pp. 73–79.

[FT2007] L.D. Faddeev and L.A. Takhtajan, *Hamiltonian methods in the theory of solitons*, Reprint of 1987 English edition, Classics in Mathematics, Springer-Verlag, New York, 2007.

[FF2000] B. L. Feigin and D. B. Fuchs, *Cohomologies of Lie groups and Lie algebras*, Lie groups and Lie algebras, II, Encyclopaedia Math. Sci., vol. 21, Springer, Berlin, 2000, pp. 125–223.

[Fla1982] M. Flato, *Deformation view of physical theories*, Czechoslovak J. Phys **B32** (1982), 472–475.

[Fra2012] Theodore Frankel, *The geometry of physics: An introduction*, Cambridge University Press, Cambridge, 2012.

[FU1991] Daniel S. Freed and Karen K. Uhlenbeck, *Instantons and four-manifolds*, Mathematical Sciences Research Institute Publications, vol. 1, Springer-Verlag, New York, 1991. Second edition.

[Fre1995] Daniel S. Freed, *Classical Chern-Simons theory. I*, Adv. Math. **113** (1995), no. 2, 237–303.

[Fre2009] ———, *Remarks on Chern-Simons theory*, Bull. Amer. Math. Soc. (N.S.) **46** (2009), no. 2, 221–254.

[Gal1967] A. Galindo, *Lie algebra extensions of the Poincaré algebra*, J. Math. Phys **8** (1967), 768–774.

[GF1963] I. M. Gelfand and S. V. Fomin, *Calculus of variations*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1963.

[GH1977] G. W. Gibbons and S. W. Hawking, *Action integrals and partition functions in quantum gravity*, Phys. Rev. D **15** (1977), no. 10, 2752–2756.

[God1969] C. Godbillon, *Géometrie differéntielle et mécanique analytique*, Hermann, Paris, 1969.

[Gol1984] William M. Goldman, *The symplectic nature of fundamental groups of surfaces*, Adv. in Math. **54** (1984), no. 2, 200–225.

[Gol1980] H. Goldstein, *Classical mechanics*, Addison Wesley, 1980.

[GH1994] Phillip Griffiths and Joseph Harris, *Principles of algebraic geometry*, Wiley Classics Library, John Wiley & Sons, Inc., New York, 1994. Reprint of the 1978 original.

[Gri1999] D.H. Griffiths, *Introduction to electrodynamics*, Prentice-Hall, 1999.

[Jac1998] J.D. Jackson, *Classical electrodynamics*, Wiley, 1998.

[Hit1987a] N. J. Hitchin, *The self-duality equations on a Riemann surface*, Proc. London Math. Soc. (3) **55** (1987), no. 1, 59–126.

[Hit1987b] Nigel Hitchin, *Stable bundles and integrable systems*, Duke Math. J. **54** (1987), no. 1, 91–114.

[HE1973] S. W. Hawking and G. F. R. Ellis, *The large scale structure of space-time*, Cambridge Monographs on Mathematical Physics, vol. 1, Cambridge University Press, London-New York, 1973.

[Kir1989] Robion C. Kirby, *The topology of* 4*-manifolds*, Lecture Notes in Mathematics, vol. 1374, Springer-Verlag, Berlin, 1989.

[Kir1976] A.A. Kirillov, *Elements of the theory of representations*, Grundlehren der Mathematischen Wissenschaften, Band 220, Springer-Verlag, Berlin, 1976.

[Kir2004] ———, *Lectures on the orbit method*, Graduate Studies in Mathematics, vol. 64, Amer. Math. Soc., Providence, RI, 2004.

[Kir2008] Alexander Jr. Kirillov, *An introduction to Lie groups and Lie algebras*, Cambridge Studies in Advanced Mathematics, vol. 113, Cambridge University Press, Cambridge, UK, 2008.

[KN1996a] Shoshichi Kobayashi and Katsumi Nomizu, *Foundations of differential geometry. Vol. I*, Wiley Classics Library, John Wiley & Sons, Inc., New York, 1996. Reprint of the 1963 original.

[KN1996b] ———, *Foundations of differential geometry. Vol. II*, Wiley Classics Library, John Wiley & Sons, Inc., New York, 1996. Reprint of the 1969 original.

[LL1976] L.D. Landau and E.M. Lifshitz, *Mechanics. Course of theoretical physics. Vol. 1*, Pergamon Press, Oxford, 1976.

[LL1971] ———, *The classical theory of fields. Course of theoretical physics. Vol. 2*, Pergamon Press, Oxford, 1971.

[LM1989] H. Blaine Lawson Jr. and Marie-Louise Michelsohn, *Spin geometry*, Princeton Mathematical Series, vol. 38, Princeton University Press, Princeton, NJ, 1989.

[Lee1984] H. C. Lee (ed.), *An introduction to Kaluza-Klein theories*, World Scientific Publishing Co., Singapore, 1984. Papers from the workshop held in Deep River, Ont., August 11–16, 1983.

[Leu1965] H. Leutwyler, *A no-interaction theorem in classical relativistic Hamiltonian particle mechanics*, Nuovo Cimento **37** (1965), no. 2, 556–567.

[MS1974] John W. Milnor and James D. Stasheff, *Characteristic classes*, Annals of Mathematics Studies, vol. No. 76, Princeton University Press, Princeton, NJ; University of Tokyo Press, Tokyo, 1974.

[Mne2019] Pavel Mnev, *Quantum field theory: Batalin-Vilkovisky formalism and its applications*, University Lecture Series, vol. 72, American Mathematical Society, Providence, RI, 2019.

[Nak2003] Mikio Nakahara, *Geometry, topology and physics*, Graduate Student Series in Physics, Institute of Physics, Bristol, 2003. Second edition.

[NS1987] Charles Nash and Siddhartha Sen, *Topology and geometry for physicists*, Academic Press, Inc., London, 1987. Reprint of the 1983 edition.

[O'N1983] Barrett O'Neill, *Semi-Riemannian geometry. With applications to relativity*, Pure and Applied Mathematics, vol. 103, Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York, 1983.

[Nov1982] S.P. Novikov, *The Hamiltonian formalism and a many-valued analogue of Morse theory*, Russian Math. Surveys **37** (1982), no. 5, 1–56.

[Rub2002] Valery Rubakov, *Classical theory of gauge fields*, Princeton University Press, Princeton, NJ, 2002.

[Ryd1996] Lewis H. Ryder, *Quantum field theory*, Second ed., Cambridge University Press, Cambridge, 1996.

[Sch1977] A. S. Schwarz, *The partition function of degenerate quadratic functional and Ray-Singer invariants*, Lett. Math. Phys. **2** (1977/78), no. 3, 247–252.

[SY1979] Richard Schoen and Shing Tung Yau, *On the proof of the positive mass conjecture in general relativity*, Comm. Math. Phys **65** (1979), no. 1, 45–76.

[Sin1978] I. M. Singer, *Some remarks on the Gribov ambiguity*, Comm. Math. Phys. **60** (1978), no. 1, 7-12.

[Ste1983] S. Sternberg, *Lectures on differential geometry*, Chelsea Publishing Co., New York, 1983.

[Tak2008] Leon A. Takhtajan, *Quantum mechanics for mathematicians*, Graduate Studies in Mathematics, vol. 95, Amer. Math. Soc., Providence, RI, 2008.

[Wal1984] Robert M. Wald, *General relativity*, University of Chicago Press, Chicago, IL, 1984.

[War1983] Frank W. Warner, *Foundations of differentiable manifolds and Lie groups*, Graduate Texts in Mathematics, vol. 94, Springer-Verlag, New York, 1983.

[Wei1995] Steven Weinberg, *The quantum theory of fields. Vol. I*, Cambridge University Press, Cambridge, 1995.

[Wei1996] ———, *The quantum theory of fields. Vol. II*, Cambridge University Press, Cambridge, 1996.

[Wel2008] Jr. Wells Raymond O., *Differential analysis on complex manifolds*, Graduate Texts in Mathematics, vol. 65, Springer, New York, 2008.

[Wit1981] Edward Witten, *A new proof of the positive energy theorem*, Comm. Math. Phys. **80** (1981), no. 3, 381–402.

[Wit1984] ———, *Non-abelian bosonization in two dimensions*, Comm. Math. Phys. **92** (1984), no. 4, 455–472.

[Wit1989] ———, *Quantum field theory and the Jones polynomial*, Comm. Math. Phys. **121** (1989), no. 3, 351–399.

[YM1954] C.N. Yang and R.L. Mills, *Conservation of isotopic spin and isotopic gauge invariance*, Phys. Rev **96** (1954), no. 1, 191–195.

[Yor1972] James W. York, *Role of conformal three-geometry in the dynamics of gravitation*, Phys. Rev. Lett. **28** (1972), no. 16, 1082–1085.

[Zee2013] Anthony Zee, *Einstein gravity in a nutshell*, Princeton University Press, Princeton, NJ, 2013.

[Zuc1987] Gregg J. Zuckerman, *Action principles and global geometry*, Mathematical aspects of string theory (San Diego, Calif., 1986), Adv. Ser. Math. Phys., vol. 1, World Sci. Publishing, Singapore, 1987, pp. 259–284.