

Classical Mechanics 220

Eric D'Hoker

*Department of Physics and Astronomy, UCLA
Los Angeles, CA 90095, USA*

9 November 2015

Contents

1	Review of Newtonian Mechanics	5
1.1	Some History	5
1.2	Newton's laws	6
1.3	Comments on Newton's laws	7
1.4	Work	8
1.5	Dissipative forces	9
1.6	Conservative forces	10
1.7	Velocity dependent conservative forces	12
1.8	Charged particle in the presence of electro-magnetic fields	14
1.9	Physical relevance of conservative forces	15
1.10	Appendix 1: Solving the zero-curl equation	15
2	Lagrangian Formulation of Mechanics	17
2.1	The Euler-Lagrange equations in general coordinates	17
2.2	The action principle	20
2.3	Variational calculus	21
2.4	Euler-Lagrange equations from the action principle	23
2.5	Equivalent Lagrangians	24
2.6	Symmetry transformations and conservation laws	24
2.7	General symmetry transformations	26
2.8	Noether's Theorem	28
2.9	Examples of symmetries and conserved charges	29
2.10	Systems with constraints	31
2.11	Holonomic versus non-holonomic constraints	33
2.12	Lagrangian formulation for holonomic constraints	36
2.13	Lagrangian formulation for some non-holonomic constraints	38
2.14	Examples	38
3	Quadratic Systems: Small Oscillations	41
3.1	Equilibrium points	41
3.2	Mechanical stability of equilibrium points	42
3.3	Small oscillations near a general solution	44
3.4	Magnetic stabilization	45
3.5	Lagrange Points	46
3.6	Stability near the non-colinear Lagrange points	49

4	Hamiltonian Formulation of Mechanics	51
4.1	Canonical position and momentum variables	51
4.2	Derivation of the Hamilton's equations	52
4.3	Some Examples of Hamiltonian formulation	53
4.4	Variational Formulation of Hamilton's equations	54
4.5	Poisson Brackets and symplectic structure	55
4.6	Time evolution in terms of Poisson brackets	56
4.7	Canonical transformations	57
4.8	Symmetries and Noether's Theorem	59
4.9	Poisson's Theorem	59
4.10	Noether charge reproduces the symmetry transformation	60
5	Lie groups and Lie algebras	62
5.1	Definition of a group	62
5.2	Matrix multiplication groups	63
5.3	Orthonormal frames and parametrization of $SO(N)$	65
5.4	Three-dimensional rotations and Euler angles	67
5.5	Definition of a Lie group	68
5.6	Definition of a Lie algebra	68
5.7	Relating Lie groups and Lie algebras	69
5.8	Symmetries of the degenerate harmonic oscillator	70
5.9	The relation between the Lie groups $SU(2)$ and $SO(3)$	73
6	Motion of Rigid Bodies	74
6.1	Inertial and body-fixed frames	74
6.2	Kinetic energy of a rigid body	75
6.3	Angular momentum of a rigid body	76
6.4	Changing frames	77
6.5	Euler-Lagrange equations for a freely rotating rigid body	78
6.6	Relation to a problem of geodesics on $SO(N)$	79
6.7	Solution for the maximally symmetric rigid body	80
6.8	The three-dimensional rigid body in terms of Euler angles	80
6.9	Euler equations	82
6.10	Poinsot's solution to Euler's equations	82
7	Special Relativity	84
7.1	Basic Postulates	84
7.2	Lorentz vector and tensor notation	86
7.3	General Lorentz vectors and tensors	88

7.4	Relativistic invariance of the wave equation	90
7.5	Relativistic invariance of Maxwell equations	91
7.6	Relativistic kinematics	94
7.7	Relativistic dynamics	96
7.8	Lagrangian for a massive relativistic particle	97
7.9	Particle collider versus fixed target experiments	98
7.10	A physical application of time dilation	99
8	Fluid Mechanics	100
8.1	Conservation of mass	101
8.2	Conservation of momentum and Newton's third law	101
8.3	Euler's equations	103
8.4	Ideal fluids	103
8.5	Viscous fluids	110
9	Completely integrable systems	114
9.1	Criteria for integrability, and action-angle variables	114
9.2	Standard examples of completely integrable systems	114
9.3	More sophisticated integrable systems	115
9.4	Elliptic functions	116
9.5	Solution by Lax pair	117
9.6	The Korteweg de Vries (or KdV) equation	119
9.7	The KdV soliton	119
9.8	Integrability of KdV by Lax pair	121
10	Global properties of Hamiltonian mechanics	123
10.1	The Poincaré Recurrence Theorem	123
10.2	Definition of a manifold	124
10.3	Examples	125
10.4	Maps between manifolds	126
10.5	Vector fields and tangent space	127
10.6	Poisson brackets and Hamiltonian flows	129
10.7	Stokes's theorem and grad-curl-div formulas	130
10.8	Differential forms: informal definition	131
10.9	Structure relations of the exterior differential algebra	132
10.10	Integration and Stokes's Theorem on forms	134
10.11	Frobenius Theorem for Pfaffian systems	134

1 Review of Newtonian Mechanics

A basic assumption of classical mechanics is that the system under consideration can be understood in terms of a fixed number N_p of point-like objects. Each such object is labeled by an integer $n = 1, \dots, N_p$, has a mass $m_n > 0$, and may be characterized by a position $\mathbf{x}_n(t)$. The positions constitute the dynamical degrees of freedom of the system. On these objects and/or between them, certain *forces* may act, such as those due to gravity and electro-magnetism. The goal of classical mechanics is to determine the time-evolution of the position $\mathbf{x}_n(t)$ due to the forces acting on body n , given a suitable set of *initial conditions*.

A few comments are in order. The point-like nature of the objects described above is often the result of an approximation. For example, a planet may be described as a point-like object when studying its revolution around the sun. But its full volume and shape must be taken into account if we plan to send a satellite to its surface, and the planet can then no longer be approximated by a point-like object. In particular, the planet will rotate as an extended body does. This extended body may be understood in terms of smaller bodies which, in turn, may be treated as point-like. A point-like object is often referred to as a *material point* or a *particle*, even though its size may be that of a planet or a star.

In contrast with quantum mechanics, classical mechanics allows simultaneous specification of both the position and the velocity (or momentum) for each of its particles. In contrast with quantum field theory, classical mechanics assumes that the number of particles is fixed, with fixed masses. In contrast with statistical mechanics, classical mechanics assumes that the positions and velocities of all particles can (in principle) be known to arbitrary accuracy.

1.1 Some History

Historically, one of the greatest difficulties that needed to be overcome was to observe and describe the motion of bodies *in the absence of any forces*. Friction on the ground and in the air could not easily be reduced with the tools available prior to the Renaissance. It is the motion of the planets which would produce the first reliable laws of mechanics. Based on the accurate astronomical observations which Tycho Brahe (1546-1601) made with the naked eye on the positions of various planets (especially Mars), Johannes Kepler (1571-1630) proposed his quantitative and precise mathematical laws of planetary motion. Galileo Galilei (1564-1642) investigated the motion of bodies on Earth, how they fall, how they roll on inclined planes, and how they swing in a pendulum. He demonstrated with the help of such experiments that bodies with different masses fall to earth at the same rate (ignoring air friction), and deduced the correct (quadratic) mathematical relation between height and elapsed time during such falls. He may not have been the first one to derive such laws, but

Galileo formulated the results in clear quantitative mathematical laws.

Galileo proposed that a body in uniform motion will stay so unless acted upon by a force, and he was probably the first to do so. Of course, some care is needed in stating this law precisely as the appearance of uniform motion may change when our reference frame in which we make the observation is changed. In a so-called *inertial frame*, which we shall denote by \mathcal{R} , the motion of a body on which no forces act is along a straight line at constant velocity and constant direction. A frame \mathcal{R}' which moves with constant velocity with respect to \mathcal{R} is then also inertial. But a frame \mathcal{R}'' which accelerates with respect to \mathcal{R} is not inertial, as a particle in uniform motion now sweeps out a parabolic figure. Galileo stated, for the first time, that the laws of mechanics should be the same in different inertial frames, a property that is referred to as *the principle of Galilean Relativity*, and which we shall discuss later.

Isaac Newton (1642-1727) developed the mathematics of differential and integral calculus which is ultimately needed for the complete formulation of the laws of mechanics. These laws form the foundation of mechanics, and were laid out in his *Philosophiae Naturalis Principia Mathematica*, or simply *the Principia*, published in 1687. Remarkably, the mathematics used in the Principia is grounded in classical Greek geometry, supplemented with methods borrowed from infinitesimal calculus. Apparently, Newton believed that a formulation in terms of Greek geometry would enjoy more solid logical foundations than a formulation in terms of Descartes analytic geometry and his own calculus.

1.2 Newton's laws

Newton's laws may be stated as follows,

1. Space is rigid and flat 3-dimensional, with distances measured with the Euclidean metric. Time is an absolute and universal parameter. Events are described by giving the position vector and the time (\mathbf{x}, t) . Events at any two points in space occur *simultaneously* if their time parameters are the same. It is assumed that in this space-time setting, an inertial frame exists.
2. The laws of motion of a material point of mass $m > 0$ and position $\mathbf{x}(t)$ are expressed in terms of the momentum of the material point, defined in terms of the mass, position and velocity by,

$$\mathbf{p} = m\mathbf{v} \qquad \mathbf{v} = \dot{\mathbf{x}} = \frac{d\mathbf{x}}{dt} \qquad (1.1)$$

Newton's second law may then be cast in the following form,

$$\frac{d\mathbf{p}}{dt} = \mathbf{F}(\mathbf{x}) \qquad (1.2)$$

where \mathbf{F} is the force to which the material point with position \mathbf{x} is subject. The second law holds in any inertial frame.

3. The law of action and reaction states that if a body B exerts a force $\mathbf{F}_{B \rightarrow A}$ on body A , then body A exerts a force $\mathbf{F}_{A \rightarrow B} = -\mathbf{F}_{B \rightarrow A}$ on body B . In terms of momentum, the law states that the momentum transferred from A to B by the action of the force is opposite to the momentum transferred from B to A .
4. The law of gravity between two bodies with masses m_A and m_B , and positions \mathbf{x}_A and \mathbf{x}_B respectively, is given by

$$\mathbf{F}_{B \rightarrow A} = -\mathbf{F}_{A \rightarrow B} = -Gm_A m_B \frac{\mathbf{r}}{|\mathbf{r}|^3} \quad \mathbf{r} = \mathbf{x}_A - \mathbf{x}_B \quad (1.3)$$

where G is Newton's constant of gravity. Here, $\mathbf{F}_{B \rightarrow A}$ is the force exerted by body B on body A , while $\mathbf{F}_{A \rightarrow B}$ is the force exerted by A on B . Since masses m_A and m_B are positive, the gravitational force between two massive bodies is always attractive.

1.3 Comments on Newton's laws

Some immediate comments on Newton's laws may be helpful.

- The definition of momentum given in (1.1) holds in *any frame*, including non-inertial frames. The expression for momentum given in (1.1) will require modification, however, in relativistic mechanics, as we shall develop later in this course.
- The mass parameter m may depend on time. For example, the mass of a rocket will depend on time through the burning and expelling of its fuel. Therefore, the popular form of Newton's second law, $\mathbf{F} = m\mathbf{a}$ with the acceleration given by $\mathbf{a} = \dot{\mathbf{v}}$, holds only in the special case where the mass m is constant.
- The total mass M , the total momentum \mathbf{P} , and the center of mass \mathbf{X} and of a system of N_p particles labelled by $n = 1, \dots, N_p$ are given respectively by the sum of the masses, the sum of the momenta, and a weighed sum of the positions of those particles,

$$M = \sum_{n=1}^{N_p} m_n \quad \mathbf{P} = \sum_{n=1}^{N_p} \mathbf{p}_n \quad \mathbf{X} = \sum_{n=1}^{N_p} \frac{m_n}{M} \mathbf{x}_n \quad (1.4)$$

If the masses are independent of time, then we have $M\dot{\mathbf{X}} = \mathbf{P}$. A first result of the third law (item 3 above) is that the total momentum \mathbf{P} in an isolated system (on which no

external forces act) is conserved during the motion of the system. A second result of the same law is that total *angular momentum*, defined by,

$$\mathbf{L} = \sum_{n=1}^{N_p} \mathbf{x}_n \times \mathbf{p}_n \quad (1.5)$$

is also conserved for an isolated system. Here \times denotes the cross-product of two 3-dimensional vectors.

- The measured value of Newton's constant of gravity is,

$$G = 6.67384(80) \times 10^{-11} \frac{m^3}{kg \times s^2} \quad (1.6)$$

- The force described by Newton's law of gravity acts instantaneously, and at a distance. Both properties will ultimately be negated, the first by special relativity, the second by field theory, such as general relativity.

1.4 Work

When a particle moves in the presence of a force, a certain amount of work is being done on the particle. The expression for the work δW done by a force \mathbf{F} under an infinitesimal displacement $d\mathbf{x}$ on a single particle is given by,

$$\delta W = \mathbf{F} \cdot d\mathbf{x} \quad (1.7)$$

The work done by the force on the particle along a path \mathcal{C}_{12} between two points $\mathbf{x}(t_1)$ and $\mathbf{x}(t_2)$ on the trajectory of the particle is given by the following integral,

$$W_{12} = \int_{\mathcal{C}_{12}} \mathbf{F} \cdot d\mathbf{x} \quad (1.8)$$

If the mass of the particle is time-independent, and we have $\mathbf{F} = m\ddot{\mathbf{x}}$, then the integral for the work may be carried out using Newton's equations, and we find,

$$W_{12} = \int_{\mathcal{C}_{12}} m\dot{\mathbf{v}} \cdot \mathbf{v} dt = T_2 - T_1 \quad (1.9)$$

where T is the kinetic energy,

$$T = \frac{1}{2} m \mathbf{v}^2 \quad (1.10)$$

and T_1, T_2 are the kinetic energies corresponding to the two boundary points on the trajectory. Thus, the work done on the particle is reflected in the change of kinetic energy of the particle.

Work is additive. Thus, in a system with N_p particles with positions $\mathbf{x}_n(t)$, subject to forces \mathbf{F}_n , the infinitesimal work is given by,

$$\delta W = \sum_{n=1}^{N_p} \mathbf{F}_n \cdot d\mathbf{x}_n \quad (1.11)$$

For a system in which all masses are independent of time, the integral between two points in time, t_1 and t_2 , may again be calculated using Newton's second law, and given by,

$$W_{12} = \sum_{n=1}^{N_p} \int_{\mathcal{C}_{12}} \mathbf{F}_n \cdot d\mathbf{x}_n = T_2 - T_1 \quad (1.12)$$

where the total kinetic energy for the N_p particles is given by,

$$T = \sum_{n=1}^{N_p} \frac{1}{2} m_n \mathbf{v}_n^2 \quad (1.13)$$

1.5 Dissipative forces

For a general force \mathbf{F} and given initial and final data the result of performing the line integral of (1.7), which defines the work, will depend upon the specific trajectory traversed. This occurs, for example, when the system is subject to *friction* and/or when *dissipation* occurs. A specific example for a single particle is given by the friction force law,

$$\mathbf{F} = -\kappa \mathbf{v} \quad \kappa > 0 \quad (1.14)$$

where κ itself may be a function of \mathbf{v} . Evaluating the work done along a *closed path*, where the points $\mathbf{x}(t_1)$ and $\mathbf{x}(t_2)$ coincide, gives

$$W_{12} = - \oint \kappa \mathbf{v} \cdot d\mathbf{x} = - \oint \kappa \mathbf{v}^2 dt \quad (1.15)$$

The integral on the right is always negative, since its integrand is manifestly positive. For a very small closed path, the work tends to zero, while for larger paths it will be a finite number since no cancellations can occur in the negative integral. As a result, the work done will depend on the path. The force of friction always takes energy away from the particle it acts on, a fact we know well from everyday experiences.

1.6 Conservative forces

A force is *conservative* if the work done depends only on the initial and final data, but not on the specific trajectory followed. A sufficient condition to have a conservative force is easily obtained when \mathbf{F} depends only on \mathbf{x} , and not on t and $\dot{\mathbf{x}}$.

Considering first the case of a single particle, and requiring that the work done on the particle vanish along all closed paths, we find using Stokes's theorem,

$$0 = \oint_{\mathcal{C}} \mathbf{F}(\mathbf{x}) \cdot d\mathbf{x} = \int_{\mathcal{D}} d^2\mathbf{s} \cdot (\nabla \times \mathbf{F}(\mathbf{x})) \quad (1.16)$$

where \mathcal{D} is any two-dimensional domain whose boundary is \mathcal{C} , and $d^2\mathbf{s}$ is the corresponding infinitesimal surface element. Vanishing of this quantity for all \mathcal{D} requires

$$\nabla \times \mathbf{F}(\mathbf{x}) = 0 \quad (1.17)$$

Up to global issues, which need not concern us here, this means that the force \mathbf{F} derives from a scalar potential V , which is defined, up to an arbitrary additive constant, by

$$\mathbf{F}(\mathbf{x}) = -\nabla V(\mathbf{x}) \quad (1.18)$$

The case of multiple particles may be treated along the same line. We assume again that the forces depend only on positions $\mathbf{x}_n(t)$, but not on velocities. From considering just one particle at the time, and varying its trajectory, it is immediate that the force \mathbf{F}_n on each particle n must be the gradient of a scalar potential, $\mathbf{F}_n = -\nabla_{\mathbf{x}_n} V^{(n)}$. Simultaneously varying trajectories of different particles yields the stronger result, however, that all these potentials $V^{(n)}$ are equal. To see this, it will be useful to introduce a slightly different notation, and define coordinates y_i and components of force f_i as follows,

$$\begin{aligned} x_n^1 &= y_{3n-2} & F_n^1 &= f_{3n-2} \\ x_n^2 &= y_{3n-1} & F_n^2 &= f_{3n-1} \\ x_n^3 &= y_{3n} & F_n^3 &= f_{3n} \end{aligned} \quad (1.19)$$

where $\mathbf{x}_n = (x_n^1, x_n^2, x_n^3)$, $\mathbf{F}_n = (F_n^1, F_n^2, F_n^3)$, with $n = 1, 2, \dots, N_p$ throughout. It will be convenient throughout to introduce the number N of dynamical degrees of freedom,

$$N = 3N_p \quad (1.20)$$

Vanishing of the work integral along any closed path \mathcal{C} may be recast in the following form,

$$0 = \sum_{n=1}^{N_p} \int_{\mathcal{C}} \mathbf{F}_n \cdot d\mathbf{x}_n = \sum_{i=1}^N \oint_{\mathcal{C}} dy_i f_i \quad (1.21)$$

We now use the higher-dimensional generalization of Stokes's Theorem to recast this line-integral in terms of an integral over a two-dimensional domain \mathcal{D} whose boundary is \mathcal{C} , but this time in the N -dimensional space of all y_i ,

$$\sum_{i=1}^N \oint_{\mathcal{C}_{12}} dy_i f_i = \sum_{i,j=1}^N \int_{\mathcal{D}} d^2 y_{ij} \left(\frac{\partial f_i}{\partial y_j} - \frac{\partial f_j}{\partial y_i} \right) \quad (1.22)$$

where $d^2 y_{ij} = dy_i dy_j$ is the area element for the plane spanned by the directions i, j . Since \mathcal{C} is arbitrary, \mathcal{D} is arbitrary, and it follows that the integrand all by itself must vanish,

$$\frac{\partial f_i}{\partial y_j} - \frac{\partial f_j}{\partial y_i} = 0 \quad (1.23)$$

for all $i, j = 1, 2, \dots, N$. Again ignoring global issues, this equation is solved generally in terms of a single potential V . A proof of this result is given in Appendix 1.10. Recasting the result in terms of the original coordinates and forces gives,¹

$$\mathbf{F}_n = -\nabla_{\mathbf{x}_n} V \quad (1.24)$$

where V is a function of \mathbf{x}_n . The work relation (1.26) still holds, but T is now the total kinetic energy of (1.13) and V is the total potential energy derived above.

1.6.1 Energy conservation

In terms of this potential, the work may be evaluated explicitly, and we find,

$$W_{12} = \sum_{n=1}^{N_p} \int_{\mathcal{C}_{12}} d\mathbf{x}_n \cdot \mathbf{F}_n = - \sum_{n=1}^{N_p} \int_{\mathcal{C}_{12}} d\mathbf{x}_n \cdot \nabla_{\mathbf{x}_n} V(\mathbf{x}) = V_1 - V_2 \quad (1.25)$$

Relation (1.9) between work and kinetic energy may be reinterpreted in terms of the total energy of the system, $T + V$, and reduces to the conservation thereof,

$$T_1 + V_1 = T_2 + V_2 \quad (1.26)$$

whence the name of *conservative force*.

¹Throughout, we shall use the notation $\nabla_{\mathbf{x}} U$ for the vector with Cartesian coordinates $(\frac{\partial U}{\partial x^1}, \frac{\partial U}{\partial x^2}, \frac{\partial U}{\partial x^3})$, and $\nabla_{\mathbf{v}} U$ for $(\frac{\partial U}{\partial v^1}, \frac{\partial U}{\partial v^2}, \frac{\partial U}{\partial v^3})$, with $\mathbf{x} = (x^1, x^2, x^3)$ and $\dot{\mathbf{x}} = \mathbf{v} = (v^1, v^2, v^3)$.

1.7 Velocity dependent conservative forces

The notion of conservative force generalizes to *velocity dependent forces*. Consider, for example, the case of the force acting on a particle with charge e due to a magnetic field \mathbf{B} ,

$$\mathbf{F} = e \mathbf{v} \times \mathbf{B} \quad (1.27)$$

The work done by this force is given by,

$$W_{12} = e \int_{c_{12}} d\mathbf{x} \cdot (\mathbf{v} \times \mathbf{B}) = e \int_{c_{12}} dt \mathbf{v} \cdot (\mathbf{v} \times \mathbf{B}) = 0 \quad (1.28)$$

and vanishes for any particle trajectory. Thus the magnetic part of the Lorentz force, though velocity dependent, is in fact conservative.

Therefore, it is appropriate to generalize the notion of conservative force to include the Lorentz force case, as well as others like it. This goal may be achieved by introducing a potential U , which generalizes the potential V above in that it may now depend on both \mathbf{x} and $\dot{\mathbf{x}}$. For the sake of simplicity of exposition, we begin with the case of a single particle.

Let us consider the total differential of $U(\mathbf{x}, \dot{\mathbf{x}})$, given by,

$$\begin{aligned} dU &= d\mathbf{x} \cdot \nabla_{\mathbf{x}} U + d\dot{\mathbf{x}} \cdot \nabla_{\dot{\mathbf{x}}} U \\ &= d\mathbf{x} \cdot \nabla_{\mathbf{x}} U - d\mathbf{x} \cdot \frac{d}{dt} (\nabla_{\dot{\mathbf{x}}} U) + d(\dot{\mathbf{x}} \cdot \nabla_{\dot{\mathbf{x}}} U) \end{aligned} \quad (1.29)$$

and integrate both sides from point 1 to point 2 on the particle trajectory. Using the relation given on the second line of (1.29), we obtain,

$$\int_{c_{12}} dU = U_2 - U_1 = \int_{c_{12}} d\mathbf{x} \cdot \left(\nabla_{\mathbf{x}} U - \frac{d}{dt} (\nabla_{\dot{\mathbf{x}}} U) \right) + \dot{\mathbf{x}} \cdot \nabla_{\dot{\mathbf{x}}} U \Big|_1^2 \quad (1.30)$$

Here, U_1 and U_2 are the values of the potential at the initial and final points on the trajectory. A slight rearrangement allows us to put the result in the following form,

$$(\dot{\mathbf{x}} \cdot \nabla_{\dot{\mathbf{x}}} U - U) \Big|_1^2 = \int_{c_{12}} d\mathbf{x} \cdot \left(-\nabla_{\mathbf{x}} U + \frac{d}{dt} (\nabla_{\dot{\mathbf{x}}} U) \right) \quad (1.31)$$

This means that *any force of the form*,

$$\mathbf{F} = -\nabla_{\mathbf{x}} U + \frac{d}{dt} (\nabla_{\dot{\mathbf{x}}} U) \quad (1.32)$$

is conservative in the sense defined above, namely that the work done onto the particle by the force is independent of the path followed between points 1 and 2,

$$W_{12} = \int_{\mathcal{C}_{12}} d\mathbf{x} \cdot \mathbf{F} = (\dot{\mathbf{x}} \cdot \nabla_{\dot{\mathbf{x}}} U - U) \Big|_1^2 \quad (1.33)$$

and depends only on the position and velocity at the end points.

The generalization to the case of multiple particles proceeds along the same lines as for velocity independent conservative forces and is straightforward. The result is that a set of velocity-dependent forces \mathbf{F}_n on an assembly of N_p particles is conservative provided there exists a single function U which depends on both positions and velocities, such that

$$\mathbf{F}_n = -\nabla_{\mathbf{x}_n} U + \frac{d}{dt} (\nabla_{\dot{\mathbf{x}}_n} U) \quad (1.34)$$

The work done along a path \mathcal{C}_{12} on the system is then given as follows,

$$W_{12} = \sum_{n=1}^{N_p} \int_{\mathcal{C}_{12}} d\mathbf{x}_n \cdot \mathbf{F}_n = \left(\sum_{n=1}^{N_p} \dot{\mathbf{x}}_n \cdot \nabla_{\dot{\mathbf{x}}_n} U - U \right) \Big|_1^2 \quad (1.35)$$

and depends only on the positions and velocities at the end points.

1.7.1 Energy conservation

Energy conservation may be read off from equations (1.12) and (1.35), and is given by,

$$H = T + U - \sum_{n=1}^{N_p} \dot{\mathbf{x}}_n \cdot \nabla_{\dot{\mathbf{x}}_n} U \quad (1.36)$$

To clarify the structure of H , we use the following equation holds on the kinetic energy,

$$0 = -2T + \sum_{n=1}^{N_p} \dot{\mathbf{x}}_n \cdot \nabla_{\dot{\mathbf{x}}_n} T \quad (1.37)$$

Adding this identity to the right side of (1.36), we see that all its terms may be combined into a function of a single quantity $L = T - U$, which is referred to as the *Lagrangian*, in terms of which the conserved quantity is given by,

$$H = \sum_{n=1}^{N_p} \dot{\mathbf{x}}_n \cdot \nabla_{\dot{\mathbf{x}}_n} L - L \quad (1.38)$$

which we recognize as the standard relation giving the *Hamiltonian*.

Note that U is allowed to be an arbitrary velocity-dependent potential, so the above derivation and results go far beyond the more familiar form $L = T - V$ where V is velocity-independent. While for a Lagrangian $L = T - V$ with velocity-independent V the conserved energy is simply given $H = T + V$, the corresponding conserved energy for a Lagrangian $L = T - U$ with velocity-dependent potential U is not given simply by $T = U$; rather the correct relation is given by (1.38).

1.8 Charged particle in the presence of electro-magnetic fields

The Lorentz force acting on a particle with charge e in the presence of electro-magnetic fields \mathbf{E} and \mathbf{B} is given by,

$$\mathbf{F} = e(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad (1.39)$$

We shall treat the general case where \mathbf{E} and \mathbf{B} may depend on both space and time. This force is velocity dependent. To show that it is conservative in the generalized sense discussed above, we recast the fields in terms of the electric potential Φ and the vector potential \mathbf{A} ,

$$\begin{aligned} \mathbf{E} &= -\nabla_{\mathbf{x}}\Phi - \frac{\partial \mathbf{A}}{\partial t} \\ \mathbf{B} &= \nabla_{\mathbf{x}} \times \mathbf{A} \end{aligned} \quad (1.40)$$

This means that we must solve half of the set of all Maxwell equations. Using the identity,

$$\mathbf{v} \times (\nabla_{\mathbf{x}} \times \mathbf{A}) = \nabla_{\mathbf{x}}(\mathbf{v} \cdot \mathbf{A}) - (\mathbf{v} \cdot \nabla_{\mathbf{x}})\mathbf{A} \quad (1.41)$$

the force may be recast as follows,

$$\mathbf{F} = -\nabla_{\mathbf{x}}(e\Phi - e\mathbf{v} \cdot \mathbf{A}) + \frac{d}{dt}(-e\mathbf{A}) \quad (1.42)$$

Introducing the following velocity dependent potential,

$$U = e\Phi - e\mathbf{v} \cdot \mathbf{A} \quad (1.43)$$

and using the fact that $\nabla_{\mathbf{x}}U = -e\mathbf{A}$, we see that the Lorentz force is indeed conservative in the generalized sense. The corresponding Lagrangian is given by,

$$L = \frac{1}{2}m\mathbf{v}^2 - e\Phi + e\mathbf{v} \cdot \mathbf{A} \quad (1.44)$$

The total energy of the charged particle, given by (1.38) will be conserved provided the electro-magnetic fields have no *explicit* time dependence.

1.9 Physical relevance of conservative forces

It would seem that in any genuine physical system, there must always be some degree of dissipation, or friction, rendering the system non-conservative. When dissipation occurs, the system actually transfers energy (and momentum) to physical degrees of freedom that have not been included in the description of the system. For example, the friction produced by a body in motion in a room full of air will transfer energy from the moving body to the air. If we were to include also the dynamics of the air molecules in the description of the system, then the totality of the forces on the body and on the air will be conservative. To summarize, it is a fundamental tenet of modern physics that, if all relevant degrees of freedom are included in the description of a system, then all forces will ultimately be conservative. Conservative systems may be described by Lagrangians, as was shown above (at least for the special case when no constraints occur).

Indeed, the four fundamental forces of Nature, gravity, electro-magnetism, weak and strong forces are all formulated in terms of Lagrangians, and thus effectively correspond to conservative forces.

Of course, friction and dissipation remain very useful effective phenomena. In particular, the whole set of phenomena associated with *self-organization of matter*, including life itself, are best understood in terms of systems subject to a high degree of dissipation. It is this dissipation of heat (mostly) that allow our bodies to self-organize, and dissipate entropy along with heat.

1.10 Appendix 1: Solving the zero-curl equation

To solve the N -dimensional vanishing-curl-equation (1.23),

$$\partial_i f_j - \partial_j f_i = 0 \qquad \partial_i = \frac{\partial}{\partial y_i} \qquad 1 \leq i < j \leq N \qquad (1.45)$$

we proceed by first solving the $j = N$ equations, for all $i < N$, namely,

$$\partial_N f_i = \partial_i f_N \qquad (1.46)$$

This in turn may be achieved by first solving the single partial differential equation in the variable y_N given by,

$$\partial_N V_N(y_1, \dots, y_N) = f_N(y_1, \dots, y_N) \qquad (1.47)$$

Locally in y_N , this equation is solved by integrating f_N in the variable y_N , while keeping the variables y_1, \dots, y_{N-1} fixed. The solution V_N is of course not unique, as one may always shift

it by a function of the variables y_1, \dots, y_{N-1} which is independent of y_N . We shall choose an arbitrary such solution V_N and eliminate f_N from the $j = N$ equations to obtain,

$$\partial_N f_i = \partial_i \partial_N V_N \quad (1.48)$$

We integrate this equation in the variable y_N , and the general solution is given by,

$$f_i(y_1, \dots, y_N) = \partial_i V_N(y_1, \dots, y_N) + f_i^{(1)}(y_1, \dots, y_{N-1}) \quad (1.49)$$

Substituting the solution of (1.48), as well as $f_1 = \partial_1 V_N$ into (1.45) for $j = N - 1$, we find that g_i satisfies,

$$\partial_{N-1} f_i^{(1)} = \partial_i f_{N-1}^{(1)} \quad (1.50)$$

Equation (1.50) is the analog of equation (1.46) but with one fewer variable. It may be solved analogously, thereby solving the entire system in terms of a single potential V .

2 Lagrangian Formulation of Mechanics

Newton's equations are expressed in an inertial frame, parametrized by Cartesian coordinates. Lagrangian mechanics provides a reformulation of Newtonian mechanics in terms of arbitrary coordinates, which is particularly convenient for generalized conservative forces, and which naturally allows for the inclusion of certain constraints to which the system may be subject. Equally important will be the fact that Lagrangian mechanics may be derived from a variational principle, and that the Lagrangian formulation allows for a systematic investigation into the symmetries and conservation laws of the system. Finally, the Lagrangian formulation of classical mechanics provides the logical starting point for the functional integral formulation of quantum mechanics.

Joseph Louis Lagrange (1736-1813) was born in Turin, Italy. He worked first at the Prussian Academy of Sciences, and was subsequently appointed the first professor of analysis at the Ecole Polytechnique in Paris which had been founded by Napoleon Bonaparte in 1794, five years after the French revolution. In 1807, Napoleon elevated Lagrange to the nobility title of Count. Besides the work in mechanics which bears his name, Lagrange developed the variational calculus, obtained fundamental results in number theory and group theory, thereby laying the ground for the work of Galois in algebra.

In this section, we shall derive the Euler-Lagrange equations from Newton's equations for systems with generalized conservative forces, in terms of arbitrary coordinates, and including certain constraints. We shall show that the Euler-Lagrange equations result from the variational principle applied to the *action functional*, investigate symmetries and conservation laws, and derive Noether's theorem.

2.1 The Euler-Lagrange equations in general coordinates

Newton's equations for a system of N_p particles, subject to generalized conservative forces, are formulated in an inertial frame, parametrized by Cartesian coordinates associated with the position vectors $\mathbf{x}_n(t)$, and take the form,

$$\frac{d\mathbf{p}_n}{dt} = -\nabla_{\mathbf{x}_n} U + \frac{d}{dt} (\nabla_{\dot{\mathbf{x}}_n} U) \quad (2.1)$$

The momentum \mathbf{p}_n may be expressed in terms of the total kinetic energy T , as follows,

$$\mathbf{p}_n = \nabla_{\dot{\mathbf{x}}_n} T \quad T = \sum_{n=1}^{N_p} \frac{1}{2} m_n \dot{\mathbf{x}}_n^2 \quad (2.2)$$

Using furthermore the fact that $\nabla_{\mathbf{x}_n} T = 0$, it becomes clear that equations (2.1) may be recast in terms of a single function L , referred to as *the Lagrangian*, which is defined by,

$$L \equiv T - U \quad (2.3)$$

in terms of which equations (2.1) become the famous Euler-Lagrange equations,

$$\frac{d}{dt} (\nabla_{\dot{\mathbf{x}}_n} L) - \nabla_{\mathbf{x}_n} L = 0 \quad (2.4)$$

These equations were derived in an inertial frame, and in Cartesian coordinates.

The remarkable and extremely useful property of the Euler-Lagrange equations is that they actually take the same form in an *arbitrary coordinate system*. To see this, it will be convenient to make use of the slightly different notation for Cartesian coordinates, introduced already in (1.19), namely for all $n = 1, 2, \dots, N_p$ we set,

$$x_n^1 = y_{3n-2} \quad x_n^2 = y_{3n-1} \quad x_n^3 = y_{3n} \quad (2.5)$$

In terms of these coordinates, the Euler-Lagrange equations take the form,

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{y}_i} \right) - \frac{\partial L}{\partial y_i} = 0 \quad (2.6)$$

for $i = 1, 2, \dots, N$. We shall again use the notation $N = 3N_p$ for the total number of degrees of freedom of the system. Next, we change variables, from the Cartesian coordinates y_i to a set of arbitrary coordinates q_i . This may be done by expressing the coordinates y_1, \dots, y_N as functions of the coordinates q_1, \dots, q_N ,

$$y_i = y_i(q_1, \dots, q_N) \quad i = 1, \dots, N \quad (2.7)$$

We want this change of coordinates to be faithful. Mathematically, we want this to be a diffeomorphism, namely differentiable and with differentiable inverse. In particular, the Jacobian matrix with entries $\partial y_i / \partial q_j$ must be invertible at all points. Denoting the Lagrangian in Cartesian coordinates by $L^{(y)}$, we shall denote the Lagrangian in the system of arbitrary coordinates q_i now simply by L . These two Lagrangians are equal for corresponding configurations in y and q related by (2.7),

$$L(q, \dot{q}) \equiv L^{(y)}(y, \dot{y}) \quad (2.8)$$

where we shall use the following shorthand throughout,

$$L(q, \dot{q}) = L(q_1, \dots, q_N; \dot{q}_1, \dots, \dot{q}_N) \quad (2.9)$$

and similarly for $L^{(y)}(y, \dot{y})$. To compare the Euler-Lagrange equations in both coordinate systems, we begin by computing the following variations,

$$\begin{aligned}\delta L^{(y)}(y, \dot{y}) &= \sum_{i=1}^N \left(\frac{\partial L^{(y)}}{\partial y_i} \delta y_i + \frac{\partial L^{(y)}}{\partial \dot{y}_i} \delta \dot{y}_i \right) \\ \delta L(q, \dot{q}) &= \sum_{j=1}^N \left(\frac{\partial L}{\partial q_j} \delta q_j + \frac{\partial L}{\partial \dot{q}_j} \delta \dot{q}_j \right)\end{aligned}\tag{2.10}$$

Now a variation δq_i produces a variation in each y_j in view of the relations (2.7) which, together with their time-derivatives, are calculated as follows,

$$\begin{aligned}\delta y_i &= \sum_{j=1}^N \frac{\partial y_i}{\partial q_j} \delta q_j \\ \delta \dot{y}_i &= \sum_{j=1}^N \left(\frac{\partial y_i}{\partial q_j} \delta \dot{q}_j + \frac{d}{dt} \left(\frac{\partial y_i}{\partial q_j} \right) \delta q_j \right)\end{aligned}\tag{2.11}$$

Under these variations, the left sides of equations (2.10) coincide in view of (2.8), and so the right sides must also be equal to one another. Identifying the coefficients respectively of $\delta \dot{q}_j$ and δq_j throughout gives the following relations,

$$\begin{aligned}\frac{\partial L}{\partial \dot{q}_j} &= \sum_{i=1}^N \frac{\partial y_i}{\partial q_j} \frac{\partial L^{(y)}}{\partial \dot{y}_i} \\ \frac{\partial L}{\partial q_j} &= \sum_{i=1}^N \left(\frac{\partial y_i}{\partial q_j} \frac{\partial L^{(y)}}{\partial y_i} + \frac{d}{dt} \left(\frac{\partial y_i}{\partial q_j} \right) \frac{\partial L^{(y)}}{\partial \dot{y}_i} \right)\end{aligned}\tag{2.12}$$

The Euler-Lagrange equations in terms of coordinates y_i and q_i are then related as follows,

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_j} - \frac{\partial L}{\partial q_j} = \sum_{i=1}^N \frac{\partial y_i}{\partial q_j} \left(\frac{d}{dt} \frac{\partial L^{(y)}}{\partial \dot{y}_i} - \frac{\partial L^{(y)}}{\partial y_i} \right)\tag{2.13}$$

If the Euler-Lagrange equations of (2.6) are satisfied in the inertial frame parametrized by Cartesian coordinates y_i , then the Euler-Lagrange equations for the Lagrangian defined by (2.8) in arbitrary coordinates q_i will be satisfied as follows,

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_j} \right) - \frac{\partial L}{\partial q_j} = 0\tag{2.14}$$

and vice-versa. The Euler-Lagrange equations take the same form in *any coordinate system*.

2.2 The action principle

Consider a mechanical system described by generalized coordinates $q_i(t)$, with $i = 1, \dots, N$, and a Lagrangian L which depends on these generalized positions, and associated generalized velocities $\dot{q}_i(t)$, and possibly also explicitly depends on time t ,

$$L(q, \dot{q}, t) = L(q_1, \dots, q_N; \dot{q}_1, \dots, \dot{q}_N; t) \quad (2.15)$$

One defines the *action* of the system by the following integral,

$$S[q] \equiv \int_{t_1}^{t_2} dt L(q_1, \dots, q_N; \dot{q}_1, \dots, \dot{q}_N; t) \quad (2.16)$$

The action is not a function in the usual sense. Its value depends on the functions $q_i(t)$ with t running through the entire interval $t \in [t_1, t_2]$. Therefore, $S[q]$ is referred to as a *functional* instead of a function, and a new notation with square brackets is being used to remind the reader of that distinction. An alternative way of describing a functional $S[q]$ is by stating that $S[q]$ depends on the *path* which $q_i(t)$ sweeps out in the N -dimensional space (or manifold) as t runs from t_1 to t_2 .

The Action Principle goes back to Pierre Louis Maupertuis (1698-1759) and Leonhard Euler (1707-1783), and was formulated in its present form by Sir William Rowan Hamilton (1805-1865). Lagrange was Euler's student.

Euler was a child prodigy and became the leading mathematician of the 18-th century. His work in mathematics and physics covers topics from the creation of the variational calculus and graph theory to the resolution of practical problems in engineering and cartography. Euler initiated the use of our standard notations for functions " $f(x)$ ", for the trigonometric functions, and introduced Euler $\Gamma(z)$ and $B(x, y)$ functions. He was incredibly prolific, at times producing a paper a week. Condorcet's eulogy of Euler included the famous quote "He ceased to calculate and to live."

The action principle applies to all mechanics systems governed by conservative forces, for which Newton's equations are equivalent to the Euler-Lagrange equations. Its statement is that the solutions to the Euler-Lagrange equations are precisely the *extrema* or *stationary points* of the action functional $S[q]$.

Before we present a mathematical derivation of the action principle, it is appropriate to make some comments. First, by extremum we mean that in the space of all possible paths $q_i(t)$, which is a huge space of infinite dimension, a trajectory satisfying the Euler-Lagrange

equations can *locally* be a *maximum*, *minimum*, or *saddle-point* of the action. There is no need for the extremum to be a global maximum or minimum. Second, the action principle is closely related with the minimum-path-length approach to geometrical optic, and in fact Hamilton appears to have been particularly pleased by this similarity.

2.3 Variational calculus

Standard differential calculus on ordinary functions was extended to the *variational calculus on functionals* by Euler, Lagrange, Hamilton and others. To get familiar with the corresponding techniques, we start with a warm-up for which we just have a single degree of freedom, $q(t)$ so that $N = 1$. The action is then a functional of a single function $q(t)$,

$$S[q] = \int_{t_1}^{t_2} dt L(q, \dot{q}, t) \quad (2.17)$$

We wish to compare the value taken by the action $S[q]$ on a path $q(t)$ with the value taken on a different path $q'(t)$, keeping the end points fixed. In fact, we are really only interested in a path $q'(t)$ that differs infinitesimally from the path $q(t)$, and we shall denote this infinitesimal deviation by $\delta q(t)$. It may be convenient to think of the infinitesimal variation δq as parametrized by the size of the variation ε together with a *fixed finite (non-infinitesimal) function* $s(t)$, so that we have,

$$\delta q(t) = \varepsilon s(t) \quad (2.18)$$

Variations are considered here to linear order in ε . The deformed paths are given as follows,

$$\begin{aligned} q'(t) &= q(t) + \varepsilon s(t) + \mathcal{O}(\varepsilon^2) \\ \dot{q}'(t) &= \dot{q}(t) + \varepsilon \dot{s}(t) + \mathcal{O}(\varepsilon^2) \end{aligned} \quad (2.19)$$

Keeping the end-points fixed amounts to the requirement,

$$s(t_1) = s(t_2) = 0 \quad (2.20)$$

We are now ready to compare the values takes by the Lagrangian on these two neighboring paths at each point in time t ,

$$\begin{aligned} \delta L(q, \dot{q}, t) &= L(q', \dot{q}', t) - L(q, \dot{q}, t) \\ &= L(q + \delta q, \dot{q} + \delta \dot{q}, t) - L(q, \dot{q}, t) \\ &= \frac{\partial L}{\partial q} \delta q + \frac{\partial L}{\partial \dot{q}} \delta \dot{q} = \varepsilon \left(\frac{\partial L}{\partial q} s + \frac{\partial L}{\partial \dot{q}} \dot{s} \right) + \mathcal{O}(\varepsilon^2) \end{aligned} \quad (2.21)$$

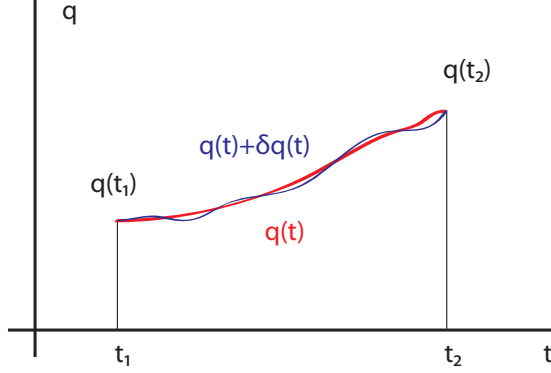


Figure 1: The path of a solution to the Euler-Lagrange equations (red), and an infinitesimal variation thereof (blue), both taking the same value at the boundary points.

Thus, the variation of the action may be expressed as follows,

$$\delta S[q] = S[q'] - S[q] = \varepsilon \int_{t_1}^{t_2} dt \left(\frac{\partial L}{\partial q} s + \frac{\partial L}{\partial \dot{q}} \dot{s} \right) + \mathcal{O}(\varepsilon^2) \quad (2.22)$$

Integrating the second term in the integral by parts, we see that the boundary terms involving $s(t)$ cancel in view of (2.20), so that we are left with,

$$\delta S[q] = \varepsilon \int_{t_1}^{t_2} dt \left(\frac{\partial L}{\partial q} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}} \right) s(t) + \mathcal{O}(\varepsilon^2) \quad (2.23)$$

The action is stationary on a path $q(t)$ provided an arbitrary first order variation in $q(t)$ leaves the value of the action unchanged. Thus, a path $q(t)$ will be extremal or stationary provided that for all functions $s(t)$ in the interval $t \in [t_1, t_2]$, and obeying (2.20) on the boundary of the interval, we have,

$$\int_{t_1}^{t_2} dt \left(\frac{\partial L}{\partial q} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}} \right) s(t) = 0 \quad (2.24)$$

This implies that the integrand must vanish at every point $t \in [t_1, t_2]$,

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = 0 \quad (2.25)$$

which is the Euler-Lagrange equation for a single degree of freedom.

One may give an alternative, but equivalent, formulation of the variational problem. Let $q'(t)$ be a deformation of $q(t)$, as specified by (2.19) by the parameter ε and the arbitrary function $s(t)$. We define the *functional derivative* of the action $S[q]$, denoted as $\delta S[q]/\delta q(t)$, by the following relation,

$$\left. \frac{\partial S[q']}{\partial \varepsilon} \right|_{\varepsilon=0} \equiv \int_{t_1}^{t_2} dt s(t) \frac{\delta S[q]}{\delta q(t)} \quad (2.26)$$

Equation (2.22) allows us to compute this quantity immediately in terms of the Lagrangian, and we find,

$$\frac{\delta S[q]}{\delta q(t)} = \frac{\partial L}{\partial q} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}} \quad (2.27)$$

Paths for which the action is extremal are now analogous to points at which an ordinary function is extremal: the first derivative vanishes, but the derivative is functional.

Variational problems are ubiquitous in mathematics and physics. For example, given a metric on a space, the curves of extremal length are the geodesics, whose shapes are determined by a variational problem. The distance functions on the Euclidean plane, on the round sphere with unit radius, and on the hyperbolic half plane are given as functionals of the path. In standard coordinates, they are given respectively by,

$$\begin{aligned} D_P(1, 2) &= \int_1^2 dt \sqrt{\dot{x}^2 + \dot{y}^2} \\ D_S(1, 2) &= \int_1^2 dt \sqrt{\dot{\theta}^2 + (\sin \theta)^2 \dot{\phi}^2} \\ D_H(1, 2) &= \int_1^2 dt \frac{\sqrt{\dot{x}^2 + \dot{y}^2}}{y} \quad y > 0 \end{aligned} \quad (2.28)$$

It is readily shown that the geodesics are respectively straight lines, grand circles, and half circles centered on the $y = 0$ axis.

2.4 Euler-Lagrange equations from the action principle

The generalization to the case of multiple degrees of freedom is straightforward. We consider arbitrary variations $\delta q_i(t)$ with fixed end points, so that,

$$\delta q_i(t_1) = \delta q_i(t_2) = 0 \quad (2.29)$$

for all $i = 1, \dots, N$. The variation of the action of (2.16) is then obtained as in (2.22),

$$\delta S[q] = \int_{t_1}^{t_2} dt \sum_{i=1}^N \left(\frac{\partial L}{\partial q_i} \delta q_i + \frac{\partial L}{\partial \dot{q}_i} \delta \dot{q}_i \right) \quad (2.30)$$

Using the cancellation of boundary contributions in the process of integration by parts the second term in the integral, we find the formula generalizing (2.23), namely,

$$\delta S[q] = \int_{t_1}^{t_2} dt \sum_{i=1}^N \left(\frac{\partial L}{\partial q_i} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} \right) \delta q_i \quad (2.31)$$

Applying now the action principle, we see that setting $\delta S[q] = 0$ for all variations $\delta q(t)$ satisfying (2.29) requires that the Euler-Lagrange equations,

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0 \quad (2.32)$$

be satisfied.

2.5 Equivalent Lagrangians

Two Lagrangians, $L(q; \dot{q}; t)$ and $L'(q; \dot{q}; t)$ are equivalent to one another if they differ by a total time derivative of a function which is *local* in t ,

$$L'(q(t); \dot{q}(t); t) = L(q(t); \dot{q}(t); t) + \frac{d}{dt} \Lambda(q(t); t) \quad (2.33)$$

We stress that this equivalence must hold for all configurations $q(t)$, not just the trajectories satisfying the Euler-Lagrange equations. The Euler-Lagrange equations for equivalent Lagrangians are the same. The proof is straightforward, since it suffices to show that the Euler-Lagrange equations for the difference Lagrangian $L' - L$ vanishes identically. It is, of course, also clear that the corresponding actions $S[q]$ and $S'[q]$ are the same up to boundary terms so that, by the variational principle, the Euler-Lagrange equations must coincide.

2.6 Symmetry transformations and conservation laws

The concept of symmetry plays a fundamental and *dominant* role in modern physics. Over the past 60 years, it has become clear that the structure of elementary particles can be organized in terms of symmetry principles. Closely related, symmetry principles play a key role in quantum field theory, in Landau's theory of second order phase transitions, and in Wilson's unification of both non-perturbative quantum field theory and phase behavior. Symmetry plays an essential role in various more modern discoveries such as topological quantum computation and topological insulators, supergravity, and superstring theory.

Here, we shall discuss the concept and the practical use of symmetry in the Lagrangian formulation (and later also in Hamiltonian formulation) of mechanics. Recall that Newton's

equations, say for a system of N particles interacting via a two-body force, are invariant under Galilean transformations,

$$\begin{aligned}\mathbf{x} &\rightarrow \mathbf{x}' = R(\mathbf{x}) + \mathbf{v}t + \mathbf{x}_0 \\ t &\rightarrow t' = t + t_0\end{aligned}\tag{2.34}$$

where \mathbf{x}_0 and t_0 produce translations in space and in time, while \mathbf{v} produces a boost, and R a rotation. Transformations under which the equations are invariant are referred to as *symmetries*. It is well-known that continuous symmetries are intimately related with conserved quantities. Time translation invariance implies conservation of total energy, while space translation invariance implies conservation of total momentum, and rotation symmetry implies conservation of angular momentum. We shall see later what happens to boosts.

We shall now proceed to define symmetries for general systems, expressed in terms of a set of arbitrary generalized coordinates $q(t)$. For the simplest cases, the derivation may be carried out without appeal to any general theory. Consider first the case of time translation invariance. We begin by calculating the time derivative of L ,

$$\frac{dL}{dt} = \frac{\partial L}{\partial t} + \sum_{i=1}^N \left(\frac{\partial L}{\partial q_i} \dot{q}_i + \frac{\partial L}{\partial \dot{q}_i} \ddot{q}_i \right)\tag{2.35}$$

Using the Euler-Lagrange equations, we eliminate $\partial L/\partial q_i$, and obtain,

$$\frac{dH}{dt} = \frac{\partial L}{\partial t}\tag{2.36}$$

where the total energy (i.e. the Hamiltonian) H is defined by,

$$H = \sum_{i=1}^N \frac{\partial L}{\partial \dot{q}_i} \dot{q}_i - L\tag{2.37}$$

The value E of the total energy function H will be conserved during the time-evolution of the system, provided $q(t)$ obeys the Euler-Lagrange equations, and the Lagrangian L has no *explicit* time-dependence. When this is the case, L is invariant under time translations, and the associated function H is referred to as a conserved quantity, or as a *first integral* of motion. Since H is time-independent, its value E may be fixed by initial conditions.

Note any system of a single dynamical variable $q(t)$, governed by a Lagrangian $L(q, \dot{q})$ which has no explicit time dependence, may be *integrated by quadrature*. Since energy is conserved for this system, and may be fixed by initial conditions, we have

$$E = \frac{\partial L}{\partial \dot{q}} \dot{q} - L\tag{2.38}$$

which is a function of q and \dot{q} only. Solve this equation for \dot{q} as a function of q , and denote the resulting expression by

$$\dot{q} = v(q, E) \quad (2.39)$$

The complete solution is given by

$$t - t_0 = \int_{q_0}^q \frac{dq'}{v(q', E)} \quad (2.40)$$

where q_0 is the value of the position at time t_0 , and q is the position at time t . For example, if $L = m\dot{q}^2/2 - V(q)$, then the integral becomes,

$$t - t_0 = \int_{q_0}^q dq' \sqrt{\frac{2(E - V(q'))}{m}} \quad (2.41)$$

2.7 General symmetry transformations

More generally, a symmetry is defined to be a transformation on the generalized coordinates,

$$q_i(t) \rightarrow q'_i(q_i; t) \quad (2.42)$$

which leaves the Lagrangian invariant, up to equivalence,

$$L(q'; \dot{q}'; t) = L(q; \dot{q}; t) + \frac{d\Lambda}{dt} \quad (2.43)$$

for some Λ which is a local function of $q_i(t)$, t , and possibly also of \dot{q}_i . Under composition of maps, *symmetries form a group*.

An alternative, but equivalent, way of looking at symmetries is as follows. One of the fundamental results of Lagrangian mechanics is that the Euler-Lagrange equations take the same form in all coordinate systems, provided the Lagrangian is mapped as follows,

$$L'(q'; \dot{q}'; t) = L(q; \dot{q}; t) \quad (2.44)$$

To every change of coordinates $q \rightarrow q'$, there is a corresponding new Lagrangian L' . A symmetry is such that the Lagrangian L' coincides with L , up to equivalence.

A transformation may be discrete, such as parity $q'_i = -q_i$, or continuous, such as translations and rotations. Continuous symmetries lead to conservation laws, while discrete symmetries do no. Thus, we shall focus on continuous transformations and symmetries. By definition, a continuous symmetry is parametrized by a continuous dependence on a set of

real parameters ε_α . For example, in space translations, we would have 3 real parameters corresponding to the coordinates of the translation, while in rotations, we could have three Euler angles. We shall concentrate on a transformation generated by a single real parameter ε . Thus, we consider coordinates which depend on ε ,

$$q'_i(t) = q_i(t, \varepsilon) \quad (2.45)$$

For a rotation in the q_1, q_2 -plane for example, we have,

$$\begin{aligned} q'_1(t, \varepsilon) &= q_1(t) \cos \varepsilon - q_2(t) \sin \varepsilon \\ q'_2(t, \varepsilon) &= q_1(t) \sin \varepsilon + q_2(t) \cos \varepsilon \end{aligned} \quad (2.46)$$

The study of continuous symmetries is greatly simplified by the fact that we can study them first infinitesimally, and then integrate the result to finite transformations later on, if needed. Thus, the finite rotations of (2.46) may be considered infinitesimally,

$$\begin{aligned} q'_1(t, \varepsilon) &= q_1(t) - \varepsilon q_2(t) + \mathcal{O}(\varepsilon^2) \\ q'_2(t, \varepsilon) &= q_2(t) + \varepsilon q_1(t) + \mathcal{O}(\varepsilon^2) \end{aligned} \quad (2.47)$$

A general continuous transformation may be considered infinitesimally, by writing,

$$q'_i(t, \varepsilon) = q_i(t) + \varepsilon \delta q_i + \mathcal{O}(\varepsilon^2) \quad (2.48)$$

where δq_i may be a function of q_i, t , as well as of \dot{q}_i . Alternatively, we have,

$$\delta q_i = \left. \frac{\partial q'_i(t, \varepsilon)}{\partial \varepsilon} \right|_{\varepsilon=0} \quad (2.49)$$

so that δq_i is a *tangent vector* to the configuration space of q_i at time t , pointing in the direction of the transformation.

Next, we impose the condition (2.43) for a transformation to be a symmetry. For our infinitesimal transformation, this relation simplifies, and we find the following condition,

$$\left. \frac{\partial L(q'; \dot{q}'; t)}{\partial \varepsilon} \right|_{\varepsilon=0} = \frac{d\Lambda}{dt} \quad (2.50)$$

for some local function Λ which may depend on $q_i(t), t$, and $\dot{q}_i(t)$.

Some comments are in order here. First, for a transformation to be a symmetry, one must be able to find the function Λ *without using the Euler-Lagrange equations* ! Second, given a Lagrangian, it is generally easy to find some symmetry transformations (such as time translation symmetry, for which the sole requirement is the absence of explicit time dependence of L), but it is considerably more difficult to find *all possible* symmetry transformations, even at the infinitesimal level. We shall provide some general algorithms later on.

2.8 Noether's Theorem

Noether's theorem(s) relating symmetries and conservation laws provides one of the most powerful tools for systematic investigations of classical mechanics with symmetries.

Emmy Noether (1882-1935) was born in Germany to a mathematician father, and went on to become one of the most influential women mathematicians of all times. After completing her dissertation, she had great difficulty securing an academic position, and worked for many years, *without pay*, at the Erlangen Mathematical Institute. David Hilbert and Felix Klein nominated her for a position at Göttingen, but the faculty in the humanities objected to appointing women at the university. After the rise of Nazism in Germany in 1933, Noether emigrated to the US, and held a position at Bryn Mawr College. Her work covers many areas of abstract algebra, including Galois theory and class field theory, as well as in mathematical physics with her work on symmetries.

Noether's Theorem states that, to every infinitesimal symmetry transformation of (2.47), and thus satisfying (2.50), there corresponds a first integral,

$$Q = \sum_{i=1}^N \frac{\partial L}{\partial \dot{q}_i} \delta q_i - \Lambda \quad (2.51)$$

which is *conserved*, i.e. it remains constant in time along any trajectory that satisfies the Euler-Lagrange equations,

$$\frac{dQ}{dt} = 0 \quad (2.52)$$

The quantity Q is referred to as the *Noether charge*.

Given the set-up we have already developed, the Theorem is easy to prove. We begin by writing out the condition (2.50) using (2.47), and we find,

$$\sum_{i=1}^N \left(\frac{\partial L}{\partial q_i} \delta q_i + \frac{\partial L}{\partial \dot{q}_i} \delta \dot{q}_i \right) = \frac{d\Lambda}{dt} \quad (2.53)$$

Using the relation $\delta \dot{q}_i = d(\delta q_i)/dt$, as well as the Euler-Lagrange equations to eliminate $\partial L/\partial q_i$, it is immediate that,

$$\sum_{i=1}^N \left(\delta q_i \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} + \frac{\partial L}{\partial \dot{q}_i} \frac{d}{dt} \delta q_i \right) = \frac{d\Lambda}{dt} \quad (2.54)$$

from which (2.52) follows using the definition of Q in (2.51).

2.9 Examples of symmetries and conserved charges

We provide here some standard examples of symmetries and conservation laws.

2.9.1 Time translation invariance

Using Noether's Theorem, we can re-derive total energy conservation. The transformation of time translation acts as follows on the position variables,

$$q'_i(t, \varepsilon) = q_i(t + \varepsilon) = q_i(t) + \varepsilon \dot{q}_i(t) + \mathcal{O}(\varepsilon^2) \quad (2.55)$$

so that $\delta q_i(t) = \dot{q}_i(t)$. Next, compute the transformation of the Lagrangian,

$$\left. \frac{\partial L(q'; \dot{q}'; t)}{\partial \varepsilon} \right|_{\varepsilon=0} = \sum_{i=1}^N \left(\frac{\partial L}{\partial q_i} \dot{q}_i + \frac{\partial L}{\partial \dot{q}_i} \ddot{q}_i \right) = \frac{dL}{dt} - \frac{\partial L}{\partial t} \quad (2.56)$$

From this, we see that we can have time translation symmetry if and only if $\partial L / \partial t = 0$. In this case, we have $\Lambda = L$, and the Noether charge Q then coincides with H . The examples of space translation and rotation symmetries are analogous but, in almost all cases have $\Lambda = 0$.

2.9.2 Invariance under translation of a canonical variable

Translations in linear combinations of the dynamical variables q_i , are given by,

$$q'_i(t) = q_i(t) + \varepsilon a_i \quad (2.57)$$

where a_i are constants (some of which, but not all of which, may vanish). Under this transformation, the Lagrangian changes as follows,

$$\frac{\partial L(q', \dot{q}'; t)}{\partial \varepsilon} = \sum_{i=1}^N \frac{\partial L}{\partial q_i} a_i \quad (2.58)$$

which is a total time derivative only if $\Lambda = 0$, so that the Euler-Lagrange equations imply,

$$\frac{d}{dt} P_a = 0 \quad P_a = \sum_{i=1}^N a_i \frac{\partial L}{\partial \dot{q}_i} \quad (2.59)$$

The momentum in the direction of translation P_a is then conserved, and the conjugate position variable is said to be a *cyclic variable*.

2.9.3 Boost invariance

The simplest system with boost invariance is a free particle moving in one dimension, with Lagrangian, $L = m\dot{q}^2/2$. A boost acts as follows,

$$q'(t, \varepsilon) = q(t) + \varepsilon t \quad (2.60)$$

where ε is the boost velocity. Proceeding as we did earlier, we find $\Lambda = mq(t)$, and thus,

$$Q = m\dot{q}t - mq(t) \quad (2.61)$$

Charge conservation here is possible because of the *explicit time dependence* of Q . We verify indeed that $\dot{Q} = 0$ provided the Euler-Lagrange equations are obeyed, namely $m\ddot{q} = 0$. The meaning of the conserved charge is clarified by solving the equation of motion, to obtain $q(t) = v_0 t + q_0$. Thus, we have $Q = -mq_0$, referring to the time-independence of a distinguished reference position of the particle. This is not really a very interesting conserved quantity, but Noether's Theorem nonetheless demonstrates its existence, and its conservation.

2.9.4 Gauge invariance

A particle with mass m and electric charge e subject to external electric and magnetic fields \mathbf{E} and \mathbf{B} is governed by the Lagrangian,

$$L = \frac{1}{2}m\dot{\mathbf{x}}^2 - e\Phi + e\dot{\mathbf{x}} \cdot \mathbf{A} \quad (2.62)$$

where $\mathbf{B} = \nabla \times \mathbf{A}$ and $\mathbf{E} = -\partial\mathbf{A}/\partial t - \nabla\Phi$. The scalar and gauge potential Φ, \mathbf{A} corresponding to given electro-magnetic fields \mathbf{E} and \mathbf{B} are not unique, and allows for arbitrary local *gauge transformations*,

$$\begin{aligned} \Phi(t, \mathbf{x}) &\rightarrow \Phi'(t, \mathbf{x}) = \Phi(t, \mathbf{x}) - \partial_t \Theta(t, \mathbf{x}) \\ \mathbf{A}(t, \mathbf{x}) &\rightarrow \mathbf{A}'(t, \mathbf{x}) = \mathbf{A}(t, \mathbf{x}) + \nabla \Theta(t, \mathbf{x}) \end{aligned} \quad (2.63)$$

for an arbitrary scalar function Θ . To derive the behavior of the Lagrangian under gauge transformations, we calculate,

$$\begin{aligned} L' - L &= -e\Phi' + e\dot{\mathbf{x}} \cdot \mathbf{A}' + e\Phi - e\dot{\mathbf{x}} \cdot \mathbf{A} \\ &= e\partial_t \Theta(t, \mathbf{x}) + e\dot{\mathbf{x}} \cdot \nabla \Theta \end{aligned} \quad (2.64)$$

and hence we have,

$$L' - L = \frac{d}{dt}(e\Lambda) \quad (2.65)$$

so that gauge transformations are symmetries of L . Since they act on the fields and not on the dynamical variable $\mathbf{x}(t)$, a derivation of the Noether charge will have to be postponed until we write down the Lagrangian also for the electro-magnetic fields.

2.10 Systems with constraints

It often happens that the motion of a particle or a body is subject to one or several *constraints*. A first example of a system with a constraint is illustrated in Figure 2 by a particle moving inside a “cup” which mathematically is a surface Σ . The motion of a particle with trajectory $\mathbf{x}(t)$ is subject to an external force \mathbf{F} (think of the force of gravity) which pulls the particle downwards. But the impenetrable wall of the cup produces a contact force \mathbf{f} which keeps the particle precisely on the surface of the cup. (This regime of the system is valid for sufficiently small velocities; if the particle is a bullet arriving at large velocity, it will cut through the cup.) While the force \mathbf{F} is known here, in general the contact force \mathbf{f} is not known. But what is known is that the particle stays on the surface Σ at all times. Such a system is referred to as a *constrained system*.

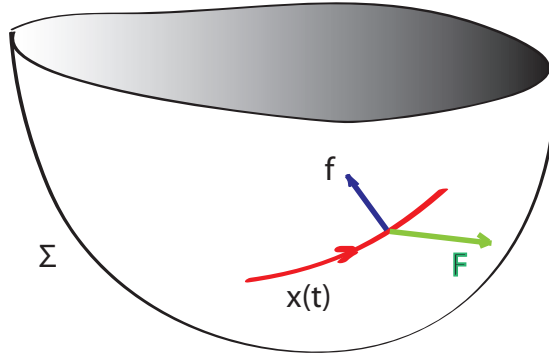


Figure 2: Particle constrained to move on a surface Σ . The trajectory $\mathbf{x}(t)$ is indicated in red, the external force \mathbf{F} in green, and the contact force \mathbf{f} in blue.

A second example of a system with a constraint is illustrated in Figure 3 by a solid body rolling on a surface Σ . The solid body has more degrees of freedom than the particle, since in addition to specifying its position (say the center of mass of the solid body), we also

need to specify three angles giving its relative orientation. The position together with the orientation must be such that the solid body touches the surface Σ at all time, but there are further conditions on the velocity as well. If the motion is without slipping and sliding, the velocity of the solid body must equal the velocity of the surface Σ at the point of contact.

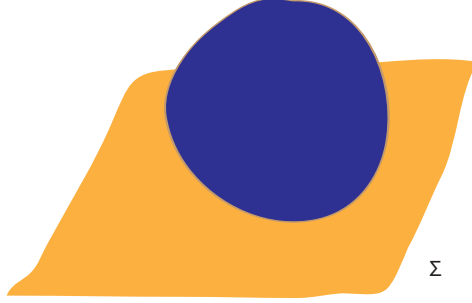


Figure 3: Solid “body” in blue constrained to move on a surface Σ .

The fact that the contact forces \mathbf{f} are not known complicates the study of constrained systems. We can make progress, however, by studying the different components of the contact force. To this end, we decompose the contact force \mathbf{f} , at the point on Σ where the contact occurs, into its component \mathbf{f}_\perp perpendicular to the surface Σ , and its component \mathbf{f}_\parallel which is tangent to the surface,

$$\mathbf{f} = \mathbf{f}_\perp + \mathbf{f}_\parallel \quad (2.66)$$

Now if the body is constrained to move along the surface Σ , then the motion of its contact point is *parallel* to Σ . Thus, the force \mathbf{f}_\perp is perpendicular to the motion of the body, and as a result does zero work,

$$\delta W_\perp = d\mathbf{x} \cdot \mathbf{f}_\perp = 0 \quad (2.67)$$

Following our earlier definitions, we see that the component \mathbf{f}_\perp of the contact force is *conservative*. The component \mathbf{f}_\parallel is parallel to Σ and, in general, will produce work on the moving body, and thus \mathbf{f}_\parallel will *not* be conservative. We shall mostly be interested in conservative forces, and will now extend the Lagrangian formulation to include the case where the contact forces are all conservative. A schematic illustration of the difference between conservative and non-conservative contact forces is given in figure 4.

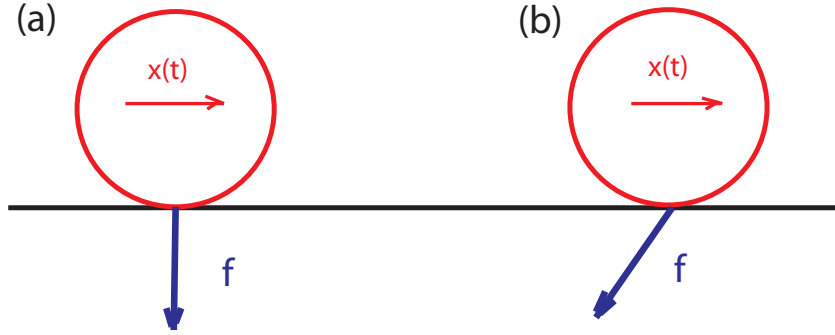


Figure 4: Schematic representation of a solid body (here a circle) constrained to move along a surface Σ (here a horizontal line). Figures (a) and (b) respectively represent the cases without and with friction.

2.11 Holonomic versus non-holonomic constraints

For systems with conservative external forces and conservative contact forces, the equations of motion may be obtained in the Lagrangian formulation, at least for certain special classes of constraints. One general form of a constraint may be expressed as the vanishing of a function ϕ_α which may depend on position, velocity, and possibly even explicitly on time. In keeping with the formulation of Lagrangian mechanics in generalized coordinates $q(t)$, we express constraints formulated as equalities in terms of generalized coordinates as well,

$$\phi_\alpha(q; \dot{q}; t) = 0 \quad \alpha = 1, \dots, A \quad (2.68)$$

where we continue to use the abbreviation $\phi_\alpha(q; \dot{q}; t) = \phi_\alpha(q_1, \dots, q_N; \dot{q}_1, \dots, \dot{q}_N; t)$ familiar from earlier discussions of Lagrangian mechanics. The role of the index α is to label the A different functionally independent constraints to which the system is subject. There may also be constraints expressed in the form of inequalities. Different types of constraints may have to be treated by different methods. We begin by discussing some of these differences, through the use of various specific examples.

Example 1

A particle is constrained to move on the surface of a sphere, in the presence of external forces, such as a gravitational field. The constraint to motion on a sphere is the result, for example, of the particle being suspended by an inelastic rod or string to form a pendulum. The dynamical degrees of freedom may be chosen to be the Cartesian position of the particle $\mathbf{x}(t)$, subject to the external gravitational force $\mathbf{F} = m\mathbf{g}$, where \mathbf{g} is the gravitational

acceleration vector. The system is subject to a single constraint $\phi = 0$ with,

$$\phi(\mathbf{x}; \dot{\mathbf{x}}; t) = (\mathbf{x}(t) - \mathbf{x}_0)^2 - R^2 \quad (2.69)$$

where \mathbf{x}_0 is the center of the sphere, and R is its radius.

Example 2

This example is somewhat more complicated, and is illustrated in Figure 5. We consider a wheel attached to an axle (in red) which in turn has its other end point attached to a fixed point. The wheel is free to roll, without slipping and sliding, on a plane.

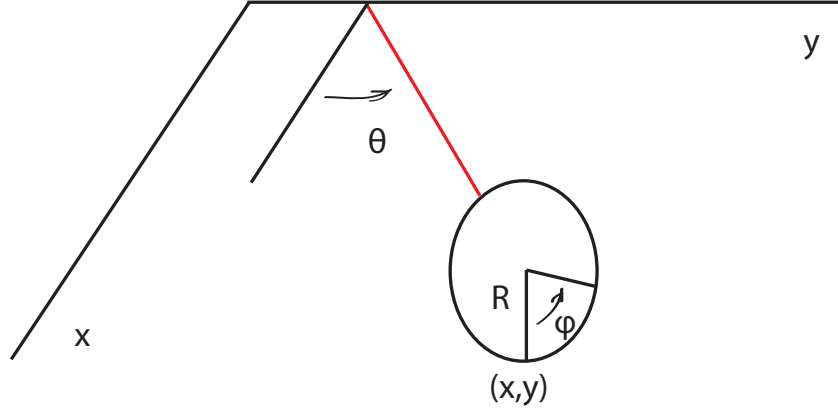


Figure 5: Wheel attached to an axle, and free to roll on a plane.

The degrees of freedom introduced in Figure 5 consist of the point of contact on the plane described by Cartesian coordinates (x, y) , the angle θ giving the position of the axle, and the angle φ giving the rotational position of the wheel. The absence of slipping and sliding requires matching the components of the velocity of the contact points, and are given by,

$$\begin{aligned} \phi_1 &= \dot{x} - R\dot{\varphi} \sin \theta \\ \phi_1 &= \dot{y} + R\dot{\varphi} \cos \theta \end{aligned} \quad (2.70)$$

Under the assumptions, any motion of the system is subject to the above two constraints.

Example 3

Finally, a last example is that of a particle allowed to move under the influence of external forces, but constrained to remain above a surface or above a plane. Such constraints are

expressed through inequalities,

$$\psi(q, \dot{q}, t) > 0 \quad (2.71)$$

Often, such constraints can be approximated by subjecting the system to a potential energy which vanishes in the allowed regions, but tends to ∞ in the forbidden regions.

2.11.1 Definition

A constraint is *holonomic* provided it is given by an equality on a set of functions which may depend on positions q and explicitly on time t , but *not* on the velocities \dot{q} . Thus, any set of holonomic constraints is given by a set of equations,

$$\phi_\alpha(q_1, \dots, q_N; t) = 0 \quad \alpha = 1, \dots, A < N \quad (2.72)$$

All other constraints are *non-holonomic*, including the constraints by inequalities.

2.11.2 Reducibility of certain velocity dependent constraints

An important caveat, which is very useful to know about, is that a constraint given by an equality $\phi(q, \dot{q}, t) = 0$ where ϕ is *linear in \dot{q}* , may actually be holonomic, even though it exhibits \dot{q} -dependence. The velocity-dependent constraint $\phi(q, \dot{q}, t) = 0$ is then *reducible* to a holonomic one. This is the case when we have,

$$\phi(q, \dot{q}, t) = \frac{d}{dt}\psi(q; t) \quad (2.73)$$

where $\psi(q; t)$ depends on positions q but not on velocities \dot{q} . The original constraint $\phi(q, \dot{q}, t) = 0$ may be replaced with the holonomic constraint $\psi(q; t) = \psi_0$ where ψ_0 is an arbitrary constant. Equivalently, writing out the constraint in terms of its linear dependence on velocities,

$$\phi(q, \dot{q}, t) = \sum_{i=1}^N a_i \dot{q}_i \quad (2.74)$$

the coefficients a_i may depend on q but not on \dot{q} . The constraint is then holonomic provided the set of coefficients is a gradient,

$$a_i = \frac{\partial \psi}{\partial q_i} \quad (2.75)$$

Alternatively, the differential form $a = \sum_{i=1}^N a_i dq_i$ must be closed for holonomic constraints, while for non-holonomic constraints, it will not be closed.

In view of the above definition and caveat, let us reconsider the examples given earlier. The constraint of example 1 is holonomic. The constraint of example 3 is non-holonomic. The constraint of example 2 is non-holonomic if the angles θ, φ are both dynamical variables, since the constraints in (2.70) cannot be transformed into holonomic ones. On the other hand, suppose we kept the angle θ fixed and removed the axle. The wheel is then allowed to roll in a straight line on the plane. In that case, the constraints are equivalent to holonomic ones,

$$\begin{aligned}\psi_1 &= x - R\varphi \cos \theta \\ \psi_2 &= y - R\varphi \sin \theta\end{aligned}\tag{2.76}$$

where the values of ψ_1 and ψ_2 are allowed to be any real constants.

2.12 Lagrangian formulation for holonomic constraints

Consider a system with coordinates $q_i(t)$, for $i = 1, \dots, N$, Lagrangian $L(q; \dot{q}; t)$, and subject to a set of A holonomic constraints $\phi_\alpha(q; t) = 0$ for $\alpha = 1, \dots, A < N$. The action principle still instructs us to extremize the action, but we must do so now subject to the constraints. This can be done through the use of *extra dynamical degrees of freedom*, referred to as *Lagrange multipliers* $\lambda_\alpha(t)$, again with $\alpha = 1, \dots, A$, namely precisely as many as the are holonomic constraints. The λ_α are independent of the variables q_i , and may be considered more or less on the same footing with them. Instead of extremizing the action $S[q]$ introduced earlier with respect to q , we extremize a new action, given by,

$$S[q; \lambda] = \int_{t_1}^{t_2} dt \left(L(q; \dot{q}; t) + \sum_{\alpha=1}^A \lambda_\alpha(t) \phi_\alpha(q; t) \right)\tag{2.77}$$

with respect to *both* q_i and λ_α for the ranges $i = 1, \dots, N$ and $\alpha = 1, \dots, A$. The role of the variables λ_α is as follows. Extremizing $S[q; \lambda]$ with respect to λ_α , keeping the independent variables q_i fixed, we recover precisely all the holonomic constraints, while extremizing with respect to q_i , keeping the independent variable λ_α fixed, gives the Euler-Lagrange equations,

$$\begin{aligned}0 &= \phi_\alpha(q; t) \\ 0 &= \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} - \sum_{\alpha=1}^A \lambda_\alpha \frac{\partial \phi_\alpha}{\partial q_i}\end{aligned}\tag{2.78}$$

A few comments are in order. First, we see that the constraints themselves result from a variational principle, which is very convenient. Second, note that the counting of equations works out: we have $N + A$ independent variables q_i and λ_α , and we have A constraint

equations, and N Euler-Lagrange equations. Third, notice that the λ_α never enter with time derivatives, so that they are non-dynamical variables, whose sole role is to provide the compensating force needed to keep the system satisfying the constraint.

Finally, and perhaps the most important remark is that, in principle, holonomic constraints can always be eliminated. This can be seen directly from the Lagrangian formulation. Since we are free to choose any set of generalized coordinates, we choose to make all the constraint functions into new coordinates,

$$\begin{aligned} q'_i(q; t) &= q_i(t) & i &= 1, \dots, N - A \\ q'_i(q; t) &= \phi_\alpha(q; t) & \alpha &= 1, \dots, A, \quad i = \alpha + N - A \end{aligned} \quad (2.79)$$

The way this works is that given the new coordinates of the first line, the coordinates q_i with $i = \alpha + N - A$ then vary with q'_i so as to satisfy (2.79).

The Lagrangian in the new coordinates q' is related to the original Lagrangian by,

$$L(q; \dot{q}; t) = L'(q'; \dot{q}'; t) \quad (2.80)$$

But now we see that in this new coordinate system, the equations of motion for λ_α simply set the corresponding coordinates $q'_i(t) = 0$ for $i = N - A + 1, \dots, N$. The Euler-Lagrange equations also simplify. This may be seen by treating the cases $i \leq N - A$ and $i > N - A$ separately. For $i \leq N - A$, we have $\partial\phi_\alpha/\partial q'_i = 0$, since ϕ_α and q'_i are independent variables. Thus the corresponding Euler-Lagrange equations are,

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0 \quad i = 1, \dots, N - A \quad (2.81)$$

For $i > N - A$, we have instead,

$$\frac{\partial \phi_\alpha}{\partial q'_i} = \delta_{\alpha+N-A,i} \quad (2.82)$$

so that the Euler-Lagrange equations become,

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = \lambda_{i-N+A} \quad i = N - A + 1, \dots, N \quad (2.83)$$

We see that the role of this last set of A equations is only to yield the values of λ_α . Since these variables were auxiliary throughout and not of direct physical relevance, this last set of equations may be ignored altogether.

2.13 Lagrangian formulation for some non-holonomic constraints

Non-holonomic constraints are more delicate. There appears to be no systematic treatment available for the general case, so we shall present here a well-known and well-understood important special case, when the constraint functions are all linear in the velocities. The proof will be postponed until later in this course. Consider a Lagrangian $L(q, \dot{q}; t)$ subject to a set of non-holonomic constraints $\phi_\alpha = 0$ with,

$$\phi_\alpha(q; \dot{q}; t) = \sum_{i=1}^N C_\alpha^i(q; t) \dot{q}_i \quad (2.84)$$

The Euler-Lagrange equations are then given by,

$$0 = \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} - \sum_{\alpha=1}^A \lambda_\alpha C_\alpha^i \quad (2.85)$$

A special case is when the constraint is reducible to a holonomic one, so that

$$C_\alpha^i = \frac{\partial \psi_\alpha}{\partial q_i} \quad (2.86)$$

When this is the case, the Euler-Lagrange equations of (2.85) reduce to the ones for holonomic constraints derived in (2.78).

2.14 Examples

The options of expressing the equations of mechanics in any coordinate system, and of solving (some, or all of) the holonomic constraints proves to be tremendous assets of the Lagrangian formulation. We shall provide two illustrations here in terms of some standard mechanical problems (see for example Landau and Lifshitz).

The first example is the system of a double pendulum, with two massive particles of masses m_1 and m_2 suspended by weightless stiff rods of lengths ℓ_1 and ℓ_2 respectively, as illustrated in figure 6 (a). The motion is restricted to the vertical plane, where the angles φ and ψ completely parametrize the positions of both particles. The whole is moving in the presence of the standard uniform gravitational field with acceleration g .

To obtain the equations of motion, we first obtain the Lagrangian. We use the fact that this system is subject only to holonomic constraints, which may be solved for completely in terms of the generalized coordinates φ and ψ , so that the Cartesian coordinates for the positions of the masses may be obtained as follows,

$$\begin{aligned} x_1 &= +\ell_1 \sin \varphi & x_2 &= \ell_1 \sin \varphi + \ell_2 \sin \psi \\ z_1 &= -\ell_1 \cos \varphi & z_2 &= -\ell_1 \cos \varphi - \ell_2 \cos \psi \end{aligned} \quad (2.87)$$

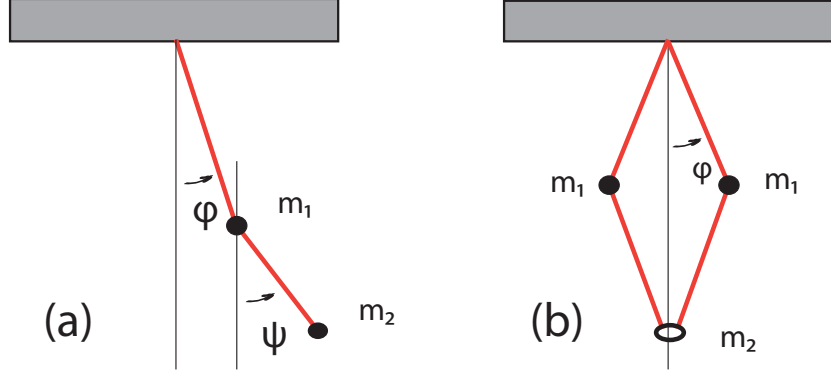


Figure 6: Motion of a double pendulum in the plane in (a), and of two coupled double pendulums subject to rotation around its axis by angle ψ in (b).

The kinetic energy is given by the sum of the kinetic energies for the two particles,

$$\begin{aligned}
 T &= \frac{1}{2}m_1(\dot{x}_1^2 + \dot{z}_1^2) + \frac{1}{2}m_2(\dot{x}_2^2 + \dot{z}_2^2) \\
 &= \frac{1}{2}(m_1 + m_2)\ell_1^2\dot{\varphi}^2 + \frac{1}{2}m_2\ell_2^2\dot{\psi}^2 + \frac{1}{2}m_2\ell_1\ell_2\dot{\varphi}\dot{\psi}\cos(\varphi - \psi)
 \end{aligned} \tag{2.88}$$

while the potential energy is given by,

$$V = -(m_1 + m_2)g\ell_1 \cos \varphi - m_2\ell_2 \cos \psi \tag{2.89}$$

The Lagrangian is given by $L = T - V$.

In the second example, illustrated in figure 6, we have two double pendulums, with segments of equal length ℓ , coupled by a common mass m_2 which is free to slide, without friction, on the vertical axis. The whole is allowed to rotate in three dimensions around the vertical axis, with positions described by the angle of aperture φ and the angle ψ of rotation around the vertical axis. The positions of the masses m_1 in Cartesian coordinates are,

$$\begin{aligned}
 x_1 &= \pm \ell \sin \varphi \cos \psi \\
 y_1 &= \pm \ell \sin \varphi \sin \psi \\
 z_1 &= -\ell \cos \varphi
 \end{aligned} \tag{2.90}$$

where the correlated \pm signs distinguish the two particles with mass m_1 . The position of the particle with mass m_2 is given by,

$$(x_2, y_2, z_2) = (0, 0, -2\ell \cos \varphi) \tag{2.91}$$

The kinetic energy of both masses m_1 coincide, and the total kinetic energy takes the form,

$$\begin{aligned} T &= m_1(\dot{x}_1^2 + \dot{y}_1^2 + \dot{z}_1^2) + \frac{1}{2}m_2\dot{z}_2^2 \\ &= m_1\ell^2(\dot{\varphi}^2 + \dot{\psi}^2 \sin^2 \varphi) + 2m_2\dot{\varphi}^2 \sin^2 \varphi \end{aligned} \tag{2.92}$$

while the potential energy is given by,

$$V = -2(m_1 + m_2)g \cos \varphi \tag{2.93}$$

The Lagrangian is given by $L = T - V$.

3 Quadratic Systems: Small Oscillations

In many physical situations, the dynamics of a system keeps it near equilibrium, with small oscillations around the equilibrium configuration. Also, the dynamics of a given system may be too complicated for a complete analytical solution, and approximations near an analytically known solution may be sometimes the best results one can obtain. All such problems ultimately boil down to linearizing the Euler-Lagrange equations (or Hamilton's equations) or equivalently reducing the problem to an action which is at most quadratic in the positions q_i and velocities \dot{q}_i . All problems of small oscillations around an equilibrium point reduce to linear differential equations, with constant coefficients, and may thus be solved by methods of linear algebra. Small oscillations around more general solutions will in general lead to linear differential equations with time-dependent coefficients, whose resolution is in general less systematic.

3.1 Equilibrium points

A system is in *mechanical equilibrium* if the generalized forces acting on the system add up to zero. In the general class of potential-type problems, the Lagrangian is of the form,

$$L = \sum_{i,j=1}^N \frac{1}{2} M_{ij}(q) \dot{q}_i \dot{q}_j + \sum_{i=1}^N N_i(q) \dot{q}_i - V(q) \quad (3.1)$$

This form includes electro-magnetic problems, expressed in arbitrary generalized coordinates. We assume that L has no explicit time dependence, thereby precluding externally driven systems. Equilibrium points correspond to extrema q_i^0 of V , which obey

$$\left. \frac{\partial V(q)}{\partial q_i} \right|_{q_i=q_i^0} = 0 \quad (3.2)$$

If the system is prepared at an initial time t_0 in a configuration with

$$q_i(t_0) = q_i^0 \quad \dot{q}_i(t_0) = 0 \quad (3.3)$$

then evolution according to the Euler-Lagrange equations will leave the system in this configuration at all times. If the system is prepared at an initial time t_0 close to an equilibrium point then, at least for sufficiently short times past t_0 , the system will remain close to that point, and the methods of small oscillations may be used reliably to calculate the evolution of the system near equilibrium.

3.2 Mechanical stability of equilibrium points

One distinguishes three types of equilibrium configurations, namely stable, unstable, and marginally stable. To make this distinction quantitative, we expand the Lagrangian around an equilibrium point q_i^0 as follows,

$$q_i(t) = q_i^0 + \eta_i(t) + \mathcal{O}(\eta^2) \quad (3.4)$$

and retains terms up to second order in η_i and $\dot{\eta}_i$. The resulting Lagrangian L_2 is given by,

$$L_2 = \sum_{i,j=1}^N \left(\frac{1}{2} m_{ij} \dot{\eta}_i \dot{\eta}_j + n_{ij} \dot{\eta}_i \eta_j - \frac{1}{2} v_{ij} \eta_i \eta_j \right) \quad (3.5)$$

where the constant matrices m_{ij} , n_{ij} , and v_{ij} are defined by,

$$m_{ij} = M_{ij} \Big|_{q_i=q_i^0} \quad n_{ij} = \frac{\partial N_i}{\partial q_j} \Big|_{q_i=q_i^0} \quad v_{ij} = \frac{\partial^2 V}{\partial q_i \partial q_j} \Big|_{q_i=q_i^0} \quad (3.6)$$

Without loss of generality, we may assume that m_{ij} and v_{ij} are symmetric in ij , while n_{ij} is anti-symmetric, up to the addition of a total time derivative to the Lagrangian.

Let us concentrate on the most common situation where m_{ij} is positive definite as a matrix. This will always be the case if the system originates from a problem with standard kinetic energy terms, which are always positive definite in the above sense. We may then change to new variables η' in terms of which the matrix m_{ij} is just the identity matrix. The precise form of the change of variables is as follows,

$$\eta'_i = \sum_{j=1}^N \mu_{ij} \eta_j \quad m_{ij} = \sum_{k=1}^N \mu_{ik} \mu_{kj} \quad (3.7)$$

where the matrix μ_{ij} is the square root of the matrix m_{ij} , and may be chosen positive definite and symmetric. Having performed this change of variables, we may now set $m_{ij} = \delta_{ij}$ in the new variables η' , and henceforth simply omit the primes from η .

We begin by studying the special case where $n_{ij} = 0$. The Lagrangian takes the form,

$$L_2 = \sum_{i,j=1}^N \left(\frac{1}{2} \delta_{ij} \dot{\eta}_i \dot{\eta}_j - \frac{1}{2} v_{ij} \eta_i \eta_j \right) \quad (3.8)$$

The different types of equilibrium may now be characterized as follows,

- STABLE: the eigenvalues of the matrix v_{ij} are all strictly positive;

- UNSTABLE: the matrix v_{ij} has at least one negative eigenvalue;
- marginally stable: the matrix v_{ij} has at least one zero eigenvalue, all other eigenvalues being positive or zero.

The case of *unstable equilibrium* includes systems for which some of the eigenvalues of v_{ij} are positive. Note that the free particle provides an example of *marginal stability*.

The relevance of the signs of the eigenvalues becomes clear upon solving the system. The Euler-Lagrange equations are,

$$\ddot{\eta}_i + \sum_{j=1}^N v_{ij} \eta_j = 0 \quad (3.9)$$

We shall write v for the $N \times N$ matrix whose components are v_{ij} , and η for the column matrix with components η_i . Thus, in matrix notation, the Euler-Lagrange equations simply read $\ddot{\eta} + v\eta = 0$. Since the matrix v is real symmetric, it may be diagonalized by a real orthogonal matrix S , so that we have $v = S^t W S$ with $S^t S = I$ and W real and diagonal, with components,

$$W_{ij} = w_i \delta_{ij} \quad (3.10)$$

In terms of the new variables $\eta' = S\eta$, with components η'_i , defined by,

$$\eta'_i(t) = \sum_{j=1}^N S_{ij} \eta_j(t) \quad (3.11)$$

the Euler-Lagrange equations decouple into 1-dimensional oscillation equations,

$$\ddot{\eta}'_i + w_i \eta'_i = 0 \quad (3.12)$$

The stability question may now be analyzed eigenvalue by eigenvalue. The solutions are given as follows,

$$\begin{aligned} w_i = +\omega_i^2 > 0 & \quad \eta'_i(t) = \gamma_i^+ e^{i\omega_i t} + \gamma_i^- e^{-i\omega_i t} \\ w_i = -\omega_i^2 < 0 & \quad \eta'_i(t) = \gamma_i^+ e^{\omega_i t} + \gamma_i^- e^{-\omega_i t} \\ w_i = 0 & \quad \eta'_i(t) = \gamma_i^1 t + \gamma_i^0 \end{aligned} \quad (3.13)$$

where in each case ω_i is real and positive, and $\gamma_i^\pm, \gamma_i^1, \gamma_i^0$ are constants to be determined by initial conditions. For generic assignments of initial conditions $\gamma_i^\pm, \gamma_i^1, \gamma_i^0$, the amplitudes of

oscillation will remain bounded if and only if all eigenvalues $w_i > 0$, which is the stable case. If at least one eigenvalue w_i is negative or zero, then the motion will become unbounded.

The general case where $n_{ij} \neq 0$ is more complicated, but often very interesting. It includes systems of charged particles in a magnetic field, or in the presence of a Coriolis force. The corresponding Euler-Lagrange equations are,

$$\ddot{\eta}_i + \sum_{j=1}^N (2n_{ij}\dot{\eta}_j + v_{ij}\eta_j) = 0 \quad (3.14)$$

A *stable system* is one in which all solutions remain bounded in time, and must be oscillatory, and of the form,

$$\eta_i(t) = \gamma_i e^{i\omega t} \quad (3.15)$$

with ω real, and satisfying the equation,

$$\sum_{j=1}^N \left(-\omega^2 \delta_{ij} + 2i\omega n_{ij} + v_{ij} \right) \gamma_j = 0 \quad (3.16)$$

This is not quite the standard form of a characteristic equation, but it may be analyzed and solved along parallel lines. In the section on Lagrange points, we shall encounter a real mechanical system with stable equilibrium points, but at which the potential V is actually a (local) maximum !

3.3 Small oscillations near a general solution

The above discussion may be generalized by considering small oscillations not around an equilibrium point, but rather around a non-trivial classical trajectory $q_i^0(t)$. Small fluctuations away from this trajectory may be parametrized as before,

$$q_i(t) = q_i^0(t) + \eta_i(t) + \mathcal{O}(\eta^2) \quad (3.17)$$

Without going through all the details, the corresponding quadratic Lagrangian for $\eta_i(t)$ will be of the form,

$$L_2 = \sum_{i,j=1}^N \left(\frac{1}{2} m_{ij}(t) \dot{\eta}_i \dot{\eta}_j + n_{ij}(t) \dot{\eta}_i \eta_j - \frac{1}{2} v_{ij}(t) \eta_i \eta_j \right) \quad (3.18)$$

leading again to linear Euler-Lagrange equations, but this time with time-dependent coefficients. The corresponding Euler-Lagrange equations are given by,

$$\sum_{j=1}^N \left(\frac{d}{dt} (m_{ij} \dot{\eta}_j + n_{ij} \eta_j) - n_{ji} \dot{\eta}_j + v_{ij} \eta_j \right) = 0 \quad (3.19)$$

where m_{ij} , n_{ij} , and v_{ij} now all depend on time. Such equations cannot, in general, be solved analytically, but must be handled numerically.

3.4 Magnetic stabilization

An first example of a system with non-trivial stability problems is provided by a particle, with mass m and electric charge e moving in a two-dimensional plane under the influence of a constant uniform magnetic field pointing in the z -direction, $\mathbf{B} = (0, 0, B)$, and an inverted harmonic oscillator potential,

$$V(x, y) = -\frac{1}{2} m \omega^2 (x^2 + y^2) \quad (3.20)$$

At face value, the system would appear unstable as the inverted harmonic potential drives the system away from any finite configuration. The stability analysis is, however, more complicated, and more interesting. The Lagrangian for the system is given by the coupling of a particle of mass m and electric charge e to a constant magnetic field B perpendicular to the xy -plane, in a potential V ,

$$L = \frac{1}{2} m (\dot{x}^2 + \dot{y}^2) + e A_x \dot{x} + e A_y \dot{y} - V(x, y) \quad (3.21)$$

We choose the gauge $A_x = -\frac{1}{2} B y$ and $A_y = +\frac{1}{2} B x$, so that the Lagrangian is given by,

$$L = \frac{1}{2} m (\dot{x}^2 + \dot{y}^2) + \frac{1}{2} e B (x \dot{y} - y \dot{x}) + \frac{1}{2} m \omega^2 (x^2 + y^2) \quad (3.22)$$

The corresponding Euler-Lagrange equations are given by,

$$\begin{aligned} m \ddot{x} - e B \dot{y} - m \omega^2 x &= 0 \\ m \ddot{y} + e B \dot{x} - m \omega^2 y &= 0 \end{aligned} \quad (3.23)$$

This system of equations is linear with constant coefficients, so the solutions are of the form,

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} e^{\lambda t} \quad \begin{pmatrix} m(\lambda^2 - \omega^2) & -eB\lambda \\ +eB\lambda & m(\lambda^2 - \omega^2) \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = 0$$

where x_0 and y_0 are initial conditions, independent of t . The four eigenvalues λ solve the quartic equation,

$$\left((2m\lambda - ieB)^2 + e^2B^2 - 4m^2\omega^2 \right) \left((2m\lambda + ieB)^2 + e^2B^2 - 4m^2\omega^2 \right) = 0 \quad (3.24)$$

When $4m^2\omega^2 < e^2B^2$, the four eigenvalues are purely imaginary, all solutions are oscillatory and thus give rise to “stable” motion around $x = y = 0$. When $4m^2\omega^2 \geq e^2B^2$ there are two a run-away modes which drive the particle indefinitely off the potential hill, as would be the case for zero magnetic field, and the system is “unstable”. In summary, a sufficiently strong magnetic field has a stabilizing effect.

3.5 Lagrange Points

The general 3-body problem for masses subject to gravitational attraction cannot be solved exactly. When one of the masses is much smaller than both of the other masses, the problem becomes more tractable. This problem also has considerable practical significance for the orbits of satellites moving in a region of space whose gravitational field is dominated by two astronomical masses, such as the Earth and the Moon. Some remarkable phenomena occur, which we shall now study.

Let m_1 and m_2 be the masses of the astronomical bodies, and $m \ll m_1, m_2$ the mass of the satellite. Let $\mathbf{x}_1, \mathbf{x}_2$ and \mathbf{x} the corresponding Cartesian positions. For simplicity, we shall assume that the mutual orbit of m_1 and m_2 around one another is circular, and we choose an inertial frame whose center is the center of mass of m_1 and m_2 . Thus, we have the relation, $m_1\mathbf{x}_1 + m_2\mathbf{x}_2 = 0$. Given the circular orbits, we have,

$$\begin{aligned} \mathbf{x}_1 &= (r_1 \cos \omega t, r_1 \sin \omega t, 0) \\ \mathbf{x}_2 &= (-r_2 \cos \omega t, -r_2 \sin \omega t, 0) \end{aligned} \quad (3.25)$$

as well as,

$$\begin{aligned} r_1 &= \frac{m_2 d}{M} & d &= r_1 + r_2 \\ r_2 &= \frac{m_1 d}{M} & M &= m_1 + m_2 \end{aligned} \quad (3.26)$$

subject to Kepler’s law

$$GM = d^3 \omega^2 \quad (3.27)$$

In the above inertial frame, the Lagrangian for the satellite is given by,

$$L = \frac{1}{2}m\dot{\mathbf{x}}^2 - \frac{Gm_1m}{|\mathbf{x} - \mathbf{x}_1|} - \frac{Gm_2m}{|\mathbf{x} - \mathbf{x}_2|} \quad (3.28)$$

In the co-moving frame, which rotates at angular frequency ω with respect to the inertial frame, the satellite position may be parametrized as follows,

$$\mathbf{x} = (x \cos \omega t - y \sin \omega t, x \sin \omega t + y \cos \omega t, z) \quad (3.29)$$

In the co-moving frame, the kinetic energy T and the potential energy V are time-independent, and take the form,

$$\begin{aligned} T &= \frac{1}{2}m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2 + \omega^2(x^2 + y^2) + 2\omega(xy - yx)) \\ V &= -\frac{Gm_1m}{[(x - r_1)^2 + y^2 + z^2]^{1/2}} - \frac{Gm_2m}{[(x + r_2)^2 + y^2 + z^2]^{1/2}} \end{aligned} \quad (3.30)$$

The value $z = 0$ is a minimum of the potential for any fixed values of x, y . Thus, motion in the plane $z = 0$ is stable, and we shall restrict to this case. Even with this simplification, the problem is too hard to solve exactly.

Notice that the change of frame has provided a centrifugal force acting outward on the satellite. So, let us see whether there are any equilibrium positions in the x, y plane, balancing the attractive gravitational potentials against the repulsive centrifugal force. To this end, we extremize the *effective potential*

$$V_{\text{eff}} = -\frac{1}{2}m\omega^2(x^2 + y^2) - \frac{Gm_1m}{[(x - r_1)^2 + y^2]^{1/2}} - \frac{Gm_2m}{[(x + r_2)^2 + y^2]^{1/2}} \quad (3.31)$$

The equations for an extremum are as follows,

$$\begin{aligned} \omega^2 x &= \frac{Gm_1(x - r_1)}{[(x - r_1)^2 + y^2]^{3/2}} + \frac{Gm_2(x + r_2)}{[(x + r_2)^2 + y^2]^{3/2}} \\ \omega^2 y &= \frac{Gm_1 y}{[(x - r_1)^2 + y^2]^{3/2}} + \frac{Gm_2 y}{[(x + r_2)^2 + y^2]^{3/2}} \end{aligned} \quad (3.32)$$

The solution has two branches. The first has $y = 0$, so that the equilibrium point is co-linear with the masses m_1 and m_2 . The equation governing the position x ,

$$\omega^2 x = \frac{Gm_1(x - r_1)}{|x - r_1|^3} + \frac{Gm_2(x + r_2)}{|x + r_2|^3} \quad (3.33)$$

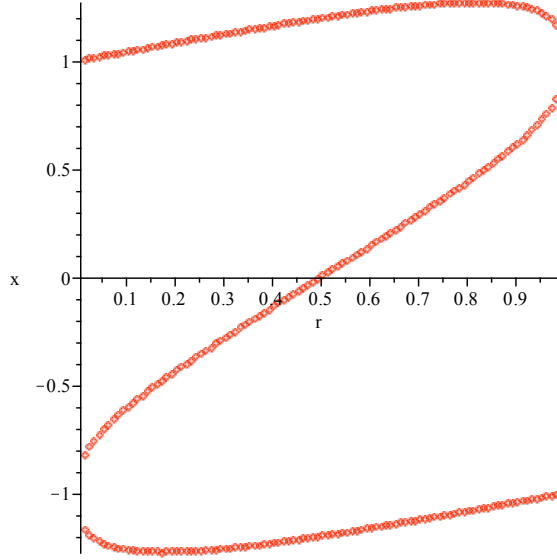


Figure 7: Plot of the positions x of the colinear Lagrange points, as a function of $r = r_1$ or $r = r_2$, in units of d . The Lagrange point of the center branch lies on the interval between masses m_1 and m_2 , while the outer branches lie outside the interval.

is equivalent to a polynomial in x of degree 5 and is solved numerically in figure 7.

For the branch $y \neq 0$, we may simplify the last equation by dividing by a factor of y . Multiplying the resulting equation by x and subtracting it from the first equation gives,

$$\begin{aligned} 0 &= \frac{-Gm_1r_1}{[(x-r_1)^2+y^2]^{3/2}} + \frac{Gm_2r_2}{[(x+r_2)^2+y^2]^{3/2}} \\ \omega^2 &= \frac{Gm_1}{[(x-r_1)^2+y^2]^{3/2}} + \frac{Gm_2}{[(x+r_2)^2+y^2]^{3/2}} \end{aligned} \quad (3.34)$$

This is a system of linear equations for the two root factors. Solving for the root factors, and then simplifying the result gives the following equations,

$$\begin{aligned} (x+r_2)^2+y^2 &= d^2 \\ (x-r_1)^2+y^2 &= d^2 \end{aligned} \quad (3.35)$$

Thus, the two equilibrium points $(x_L, \pm y_L)$ form an *equilateral triangle* with the masses m_1, m_2 . These points are referred to as the *Lagrange points*. From looking at the potential, it becomes clear that the Lagrange points are *maxima* of the effective potential V_{eff} .

3.6 Stability near the non-colinear Lagrange points

We shall focus on the non-colinear Lagrange points, since their positions are known analytically. The analysis for the co-linear Lagrange points is analogous. Linearizing the Euler-Lagrange equations around the Lagrange points,

$$\begin{aligned} x &= x_L + X & x_L &= \frac{1}{2}(r_1 - r_2) \\ y &= y_L + Y & y_L &= \frac{\sqrt{3}}{2}d \end{aligned} \quad (3.36)$$

we find the following equations of motion,

$$\begin{aligned} \ddot{X} - 2\omega\dot{Y} - \frac{3}{4}\omega^2 X + \lambda\omega^2 Y &= 0 \\ \ddot{Y} + 2\omega\dot{X} - \frac{9}{4}\omega^2 Y + \lambda\omega^2 X &= 0 \end{aligned} \quad (3.37)$$

with

$$\lambda = \frac{3\sqrt{3}}{4} \frac{m_1 - m_2}{m_1 + m_2} \quad (3.38)$$

Since the coefficients in this differential equation are constant, the solutions are of the form,

$$\begin{aligned} X &= X_0 e^{\kappa t} \\ Y &= Y_0 e^{\kappa t} \end{aligned} \quad (3.39)$$

where κ is given by,

$$\kappa^2 = \left(-\frac{1}{2} \pm i\sqrt{\frac{23}{16} - \lambda^2} \right) \omega^2 \quad (3.40)$$

Given the definition of λ , we have,

$$\lambda^2 = \frac{27}{16} - \varepsilon \quad 0 < \varepsilon < \frac{27}{16} \quad (3.41)$$

In the case where one of the astronomical bodies is much heavier than the other, the parameter ε will be small. In this case, the solutions are as follows

$$\begin{aligned} \kappa_{\pm}^{(1)} &\sim \pm i\omega \\ \kappa_{\pm}^{(2)} &\sim \pm i\omega\varepsilon \end{aligned} \quad (3.42)$$

To this order, the time dependence is purely oscillatory, with no runaway solutions ! The Lagrange points, even though they occur at the *maximum* of the potential, are actually stable due to the Coriolis forces. These properties will continue to hold as long as $\lambda^2 > 23/16$, or $0 \leq m_1^2 + m_2^2 - 25m_1m_2$. Applying these calculations to possible Lagrange points in the Earth-Moon system, we use the values of their masses,

$$\begin{aligned} m_1^{(\text{earth})} &= 5.97 \times 10^{24} \text{kg} \\ m_2^{(\text{Moon})} &= 7.36 \times 10^{22} \text{kg} \end{aligned} \tag{3.43}$$

Since $\alpha = m_2/m_1 \sim 0.0123$, and $1 - 25\alpha + \alpha^2 \sim 0.6926 > 0$, we see that the Earth-Moon system does have stable Lagrange points.

4 Hamiltonian Formulation of Mechanics

Having transformed Newton's equations into the Lagrangian formulation of mechanics, one may wonder why yet a further different formulation is worth pursuing. The Hamiltonian formulation is based on a change of variables from velocities to associated momenta, and as a result converts the (typically) second order equations of motion of Newton and Lagrange into first order equations only. Of course, such a change can be effected on any set of second order equations, but the Hamiltonian formulation does this in a particularly elegant and useful manner. As a result, the Hamiltonian formulation has compelling geometric and algebraic structures, from which powerful methods emerge for the study of symmetries, and the formulations of both quantum and statistical physics.

In this chapter we shall present the basic construction of the Hamiltonian formulation, and postpone more advance geometric and algebraic considerations to later.

4.1 Canonical position and momentum variables

The starting point will be a Lagrangian formulation and we shall assume, for the sake of simplicity, that all holonomic constraints have been eliminated. The remaining canonical position variables are $q_i(t)$ with $i = 1, \dots, N$. We now introduce generalized momenta through the definition,

$$p_i \equiv \frac{\partial L}{\partial \dot{q}_i}(q, \dot{q}; t) \quad (4.1)$$

The canonical momentum enters directly in the Euler-Lagrange equations, which may be recast in terms of p_i by,

$$\frac{dp_i}{dt} = \frac{\partial L}{\partial q_i} \quad (4.2)$$

The Hamiltonian formulation is derived from the Lagrangian by performing a *Legendre transform*. We begin by performing the following change of variables,

$$(q_i, \dot{q}_i; t) \rightarrow (q_i, p_i; t) \quad (4.3)$$

Thus, it will be required to invert the relation (4.1), and to obtain the function,

$$\dot{q}_i(q, p; t) \quad (4.4)$$

For the simplest cases, the relation (4.1) admits an inverse and \dot{q}_i may be uniquely obtained in terms of q and \dot{q} . For certain systems, however, the relation (4.1) is not invertible, and the system is subject to *constraints*.

The space (or manifold) locally parametrized by the coordinates q_i, p_i for $i = 1, \dots, N$ is referred to as *phase space*. This concept plays a fundamental role in the global analysis of classical mechanical dynamics, as well as in statistical and quantum physics.

4.2 Derivation of the Hamilton's equations

Assuming invertibility of (4.1), the function p_i is uniquely defined via (4.4), and we wish to change variables to q_i, p_i . To this end, consider the total differential of L ,

$$dL = \frac{\partial L}{\partial t} dt + \sum_{i=1}^N \left(\frac{\partial L}{\partial q_i} dq_i + \frac{\partial L}{\partial \dot{q}_i} d\dot{q}_i \right) \quad (4.5)$$

By the definition of the momenta p_i , the last term may be partially recast in terms of the momenta p_i , so that we have equivalently,

$$dL = \frac{\partial L}{\partial t} dt + \sum_{i=1}^N \left(\frac{\partial L}{\partial q_i} dq_i + p_i d\dot{q}_i \right) \quad (4.6)$$

This expression is not yet satisfactory since the differential still appeals to the velocity. To eliminate this dependence, we use instead the Legendre transform of L , which is the Hamiltonian H , defined by,

$$H = \sum_{i=1}^N p_i \dot{q}_i - L \quad (4.7)$$

As it stands, this quantity has a mixed dependence on q_i, \dot{q}_i , and p_i . Later on we shall eliminate its dependence on \dot{q}_i in favor of p_i . Its total differential is given by,

$$dH = -dL + \sum_{i=1}^N (p_i d\dot{q}_i + \dot{q}_i dp_i) \quad (4.8)$$

Eliminating dL between (4.6) and (4.8) on the one hand, we obtain,

$$dH = -\frac{\partial L}{\partial t} dt + \sum_{i=1}^N \left(-\frac{\partial L}{\partial q_i} dq_i + \dot{q}_i dp_i \right) \quad (4.9)$$

On the other hand, eliminating now the velocity in favor of positions q_i and momenta in p_i in H , using (4.4), H becomes a function of q_i, p_i , and t , and its total differential is given by,

$$dH = \frac{\partial H}{\partial t} dt + \sum_{i=1}^N \left(\frac{\partial H}{\partial q_i} dq_i + \frac{\partial H}{\partial p_i} dp_i \right) \quad (4.10)$$

Comparison of (4.9) and (4.10) reveals the following relations,

$$\dot{q}_i = \frac{\partial H}{\partial p_i} \quad \dot{p}_i = -\frac{\partial H}{\partial q_i} \quad (4.11)$$

These are Hamilton's equations of mechanics. The relation between Lagrangian and Hamiltonian mechanics, contained in the subsidiary relations,

$$\frac{\partial L}{\partial t} = -\frac{\partial H}{\partial t} \quad (4.12)$$

may be omitted altogether. Note that if the Lagrangian has no explicit time-dependence, which is equivalent to saying that the system is invariant under time-translations, then also the Hamiltonian has no explicit time-dependence, and vice-versa. Thus, in the Hamiltonian formulation, time-translation invariance is guaranteed by the fact that the Hamiltonian has no explicit time-dependence.

4.3 Some Examples of Hamiltonian formulation

Example 1. Consider a system with N degrees of freedom q_i , and Lagrangian given by,

$$L = \sum_{i,j=1}^N \frac{1}{2} M_{ij}(q) \dot{q}_i \dot{q}_j - V(q) \quad (4.13)$$

We assume the matrix $M_{ij}(q)$ to be real symmetric and invertible for all q . The associated canonical momenta are given by,

$$p_i = \frac{\partial L}{\partial \dot{q}_i} = \sum_{j=1}^N M_{ij}(q) \dot{q}_j \quad (4.14)$$

Inverting this relation allows us to express the velocities \dot{q}_i in terms of the momenta p_i ,

$$\dot{q}_i = \sum_{j=1}^N \left(M(q)^{-1} \right)_{ij} p_j \quad (4.15)$$

where $M(q)^{-1}$ denotes the inverse matrix of $M(q)$. Using the definition of the Hamiltonian, and eliminating \dot{q}_i it is found to be given by,

$$H = \sum_{i,j=1}^N \frac{1}{2} \left(M(q)^{-1} \right)_{ij} p_i p_j + V(q) \quad (4.16)$$

An immediate application of this example is a particle in spherical coordinates, r, θ, ϕ , given by the standard Lagrangian with potential,

$$L = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2\theta\dot{\phi}^2) - V(r, \theta, \phi) \quad (4.17)$$

Its canonical momenta are,

$$p_r = m\dot{r} \quad p_\theta = mr^2\dot{\theta} \quad p_\phi = mr^2\sin^2\theta\dot{\phi} \quad (4.18)$$

The associated Hamiltonian is readily found with the help of the above general formulas,

$$H = \frac{1}{2m} \left(p_r^2 + \frac{p_\theta^2}{r^2} + \frac{p_\phi^2}{r^2\sin^2\theta} \right) + V(r, \theta, \phi) \quad (4.19)$$

Example 2. A particle with mass m and electric charge e in the presence of an electro-magnetic field with scalar potential Φ and vector potential \mathbf{A} , has Lagrangian,

$$L = \frac{1}{2}m\mathbf{v}^2 - e\Phi + e\mathbf{v} \cdot \mathbf{A} \quad \mathbf{v} = \dot{\mathbf{x}} \quad (4.20)$$

The canonical momentum is found to be,

$$\mathbf{p} = m\mathbf{v} + e\mathbf{A} \quad (4.21)$$

and thus receives an electro-magnetic contribution. The Hamiltonian is found to be,

$$H = \frac{1}{2m} (\mathbf{p} - e\mathbf{A})^2 + e\Phi \quad (4.22)$$

4.4 Variational Formulation of Hamilton's equations

The action $S[q]$ associated with the Lagrangian $L(q, \dot{q}; t)$,

$$S[q] = \int_{t_1}^{t_2} dt L(q(t); \dot{q}(t); t) \quad (4.23)$$

may be recast in terms of the Hamiltonian using the Legendre transformation formula (4.7), to get a form which now depends not just on positions $q_i(t)$ but also on momenta $p_i(t)$, and is given by,

$$I[q, p] = \int_{t_1}^{t_2} dt \left(\sum_{i=1}^N p_i(t) \dot{q}_i(t) - H(q(t); p(t); t) \right) \quad (4.24)$$

The functional derivatives of I with respect to q and p are given by,

$$\begin{aligned}\frac{\delta I[q, p]}{\delta q_i(t)} &= -\dot{p}_i(t) - \frac{\partial H}{\partial p_i} \\ \frac{\delta I[q, p]}{\delta p_i(t)} &= +\dot{q}_i(t) - \frac{\partial H}{\partial q_i}\end{aligned}\tag{4.25}$$

Setting these functional derivatives to zero, as is appropriate for enforcing the action principle, is readily seen to be equivalent to the Hamilton equations of (4.11).

4.5 Poisson Brackets and symplectic structure

There is a powerful geometrical structure on phase space, the *Poisson bracket* which is very useful for a more mathematical formulation of classical mechanics, as well as for its relation with quantum mechanics. For any two functions $A(q, p)$ and $B(q, p)$, the Poisson bracket is denoted by $\{A, B\}$, whose definition is as follows,²

$$\{A, B\} = \{A, B\}_{q,p} \equiv \sum_{i=1}^N \left(\frac{\partial A}{\partial q_i} \frac{\partial B}{\partial p_i} - \frac{\partial A}{\partial p_i} \frac{\partial B}{\partial q_i} \right)\tag{4.26}$$

The Poisson bracket is linear in A and B , anti-symmetric under interchange of its arguments, acts as a derivative in each argument, and satisfies the Jacobi identity,

$$\begin{aligned}0 &= \{A, B\} + \{B, A\} \\ \{A, BC\} &= \{A, B\}C + \{A, C\}B \\ 0 &= \{\{A, B\}, C\} + \{\{B, C\}, A\} + \{\{C, A\}, B\}\end{aligned}\tag{4.27}$$

These relations follow from the application of basic chain rule, and Schwarz's identity.

Interestingly, we may proceed more algebraically, and more globally in terms of phase space. It follows from the definition of ([4.26]) that,

$$\begin{aligned}\{q_i, q_j\} = \{p_i, p_j\} &= 0 \\ \{q_i, p_j\} = -\{p_i, q_j\} &= \delta_{i,j}\end{aligned}\tag{4.28}$$

²We use the sign convention for Poisson brackets of the classic book by Goldstein, which is the same as the conventions used by José and Saletan, but opposite to the convention used by Landau and Lifschitz. When exhibiting the canonical variables q and p in terms of which the Poisson bracket is expressed locally is useful, we shall denote the Poisson bracket instead by $\{A, B\}_{q,p}$.

Suppose we now postulate a pairing $\{a, b\}$ which is linear in a , and linear in b , satisfies the properties (4.27), as well as the relations (4.28). Then it follows that the structure $\{A, B\}$ coincides with the Poisson bracket of (4.26). The novelty of our new approach is that linearity together with relations (4.27) are extremely natural and general equations which arise in many branches of mathematics, and which are related to the theory of *Lie algebras*. The *canonical Poisson brackets* of (4.28) merely express a normalization of a privileged set of coordinates (q, p) on phase space.

In fact, let us unify the position and momentum coordinates q_i and p_i with $i = 1, \dots, N$ into a single set of coordinates ϕ_α of phase space, where $\alpha = 1, \dots, 2N$, and

$$\begin{aligned}\phi_i &= q_i & i = 1, \dots, N \\ \phi_{i+N} &= p_i\end{aligned}\tag{4.29}$$

The canonical Poisson brackets of (4.28) now take the form,

$$\{\phi_\alpha, \phi_\beta\} = \mathcal{J}_{\alpha\beta}\tag{4.30}$$

where we define the matrix \mathcal{J} as follows,

$$\mathcal{J} = \begin{pmatrix} 0 & I_N \\ -I_N & 0 \end{pmatrix}\tag{4.31}$$

and I_N denotes the identity matrix in N dimensions. Now the matrix \mathcal{J} is very special, and corresponds to a *canonical symplectic pairing*. In particular, we have $\mathcal{J}^2 = -I_{2N}$. More generally, if the Poisson brackets on phase space are given by the relation (4.30) for a general (real anti-symmetric) matrix \mathcal{J} , and the matrix \mathcal{J} is invertible, then we actually have a *symplectic structure* on phase space. More formally, the phase space *manifold* is a Poisson manifold if it carries relation (4.30), and a symplectic manifold if \mathcal{J} is invertible.

4.6 Time evolution in terms of Poisson brackets

The time derivative of any function $A(q, p; t)$, along a trajectory which satisfies Hamilton's equations, may be expressed simply via Poisson brackets,

$$\frac{d}{dt}A = \frac{\partial}{\partial t}A + \sum_i \left(\frac{\partial A}{\partial q_i} \dot{q}_i + \frac{\partial A}{\partial p_i} \dot{p}_i \right)\tag{4.32}$$

Using Hamilton's equations we eliminate \dot{q}_i and \dot{p}_i , and we find,

$$\frac{dA}{dt} = \frac{\partial A}{\partial t} + \{A, H\}\tag{4.33}$$

In particular, for the canonical variables q_i and p_i , and for the phase space variables ϕ_α , Hamilton's equations may be recast in the following form,

$$\begin{aligned}\dot{q}_i &= \{q_i, H\} & \dot{\phi}_\alpha &= \{\phi_\alpha, H\} \\ \dot{p}_i &= \{p_i, H\}\end{aligned}\tag{4.34}$$

We shall soon interpret these equations even more geometrically.

4.7 Canonical transformations

The idea is to look for a set of generalized coordinates in terms of which the system simplifies as much as possible. In the Lagrangian formulation, we allowed for a change of coordinates $q_i \rightarrow q'_i(q; t)$ accompanied by a redefinition of the Lagrangian $L'(q', \dot{q}'; t) = L(q, \dot{q}; t)$ maps the Euler-Lagrange equations for L into the Euler-Lagrange equations for L' . A change of coordinates is particularly useful if one of the new coordinates q'_i is *cyclic*, since then the system may be partially integrated.

In the Hamiltonian formulation, positions and momenta are on the same footing, so we expect to be able to make changes of coordinates on the full phase space. Let us denote the coordinate transformations as follows,

$$q'_i = q'_i(q, p; t) \qquad p'_i = p'_i(q, p; t)\tag{4.35}$$

The transformation is canonical provided Hamilton's equations in the new coordinates take on the same form as in the old coordinates,

$$\dot{q}'_i = \frac{\partial H'}{\partial p'_i} \qquad \dot{p}'_i = -\frac{\partial H'}{\partial q'_i}\tag{4.36}$$

where H' is the new Hamiltonian in the coordinates q'_i, p'_i . Not all transformations of the form (4.35) are canonical. Rather than trying to work out directly the conditions (4.36) for a transformation to be canonical, it is easier to go back to the action principle in the Hamiltonian form, which is where Hamilton's equations derived from. The integrand of the action must be the same in both coordinates, up to a total time derivative,

$$\sum_{i=1}^N p_i dq_i - H(p, q; t)dt = \sum_{i=1}^N p'_i dq'_i - H'(p', q'; t)dt + dF\tag{4.37}$$

where F is a scalar function. Now we could continue to take q_i, p_i as independent phase space variables. Actually, since the differentials appearing in (4.37) are dt, dq_i, dq'_i , it turns

out to be simpler to use q_i and q'_i as independent variables, and we shall thus assume that F is a function of these, so that $F = F(q_i, q'_i; t)$. Identifying independent differentials gives,

$$\frac{\partial F}{\partial t} = H'(q', p'; t) - H(q, p; t) \quad \frac{\partial F}{\partial q_i} = p_i \quad \frac{\partial F}{\partial q'_i} = -p'_i \quad (4.38)$$

The function F may be viewed as the *generating function* for the canonical transformation.

There is another, but equivalent, way of looking at canonical transformations. Using the relations of (4.38), one shows that canonical transformations preserve the Poisson bracket, in the following sense

$$\{A', B'\}_{q', p'} = \{A, B\}_{q, p} \quad (4.39)$$

Here, we denote Poisson brackets evaluated with respect to the coordinates q, p by $\{\cdot, \cdot\}_{q, p}$ to render this dependence explicit. The functions A', B' are related to A, B by,

$$A'(q', p') = A(q, p) \quad B'(q', p') = B(q, p) \quad (4.40)$$

Often one does not even make the distinction between primed and unprimed functions (see e.g. Landau and Lifshitz). To prove this result, we compute both side in terms of a set of independent variables which are well-adapted to the problem, and these are q_i, Q_i . The result is,

$$\begin{aligned} \{A, B\}_{q, p} &= \sum_{i, j} \frac{\partial q'_j}{\partial p_i} \left(\frac{\partial A}{\partial q_i} \frac{\partial B}{\partial q'_j} - \frac{\partial B}{\partial q_i} \frac{\partial A}{\partial q'_j} \right) \\ \{A', B'\}_{q', p'} &= \sum_{i, j} \frac{\partial q_j}{\partial p'_i} \left(\frac{\partial A}{\partial q'_i} \frac{\partial B}{\partial q_j} - \frac{\partial B}{\partial q'_i} \frac{\partial A}{\partial q_j} \right) \end{aligned} \quad (4.41)$$

To obtain these result, we have made use of the fact that the following matrices are symmetric under interchange of i, j ,

$$\frac{\partial p_i}{\partial q_j} = \frac{\partial^2 F}{\partial q_i \partial q_j} \quad \frac{\partial p'_i}{\partial q'_j} = \frac{\partial^2 F}{\partial q'_i \partial q'_j} \quad (4.42)$$

Next, to compare both lines of (4.41), we evaluate the inverses of the prefactors,

$$\frac{\partial p_j}{\partial q'_i} = \frac{\partial^2 F}{\partial q_j \partial q'_i} \quad \frac{\partial p'_j}{\partial q_i} = -\frac{\partial^2 F}{\partial q'_j \partial q_i} \quad (4.43)$$

It follows that (4.39) holds true.

4.8 Symmetries and Noether's Theorem

Noether's theorem, which we derived in the Lagrangian formulation, has an important extension and reformulation in the Hamiltonian formulation. Recall that an infinitesimal transformation δq_i on the dynamical variables q_i is a symmetry of the Euler-Lagrange equations provided the corresponding infinitesimal change in the Lagrangian obeys,

$$\delta L = \frac{d\Lambda}{dt} \quad (4.44)$$

where Λ is a function of $q(t)$ and $\dot{q}(t)$ which is local in t . The Noether charge associated with this symmetry transformation is given by,

$$Q = \sum_{i=1}^N \frac{\partial L}{\partial \dot{q}_i} \delta q_i - \Lambda \quad (4.45)$$

By construction in the Lagrangian formulation, Q is a function of q and \dot{q} .

In order to make use of Q in the Hamiltonian formulation, we begin by changing variables from $(q, \dot{q}; t) \rightarrow (q, p; t)$ as in (4.3), so that Q is now a function of $(q, p; t)$. Its conservation may be re-expressed using the Poisson bracket and formula (4.30), which applies to *any* function of $(q, p; t)$,

$$0 = \frac{dQ}{dt} = \frac{\partial Q}{\partial t} + \{Q, H\} \quad (4.46)$$

There is a powerful result, referred to as *Poisson's theorem*, which provides relations between different continuous symmetries, and Noether charges, of a given mechanical system.

4.9 Poisson's Theorem

Poisson's Theorem states that if Q_1 and Q_2 are two conserved Noether charges, then their Poisson bracket $\{Q_1, Q_2\}$ is also a conserved Noether charge. It is worth proving this theorem, for the sake of practice with Poisson brackets. To prove the Theorem, it will suffice to prove that the total time derivative of $\{Q_1, Q_2\}$ vanishes. Thus, we calculate,

$$\begin{aligned} \frac{d}{dt}\{Q_1, Q_2\} &= \frac{\partial}{\partial t}\{Q_1, Q_2\} + \{\{Q_1, Q_2\}, H\} \\ &= \left\{\frac{\partial Q_1}{\partial t}, Q_2\right\} + \left\{Q_1, \frac{\partial Q_2}{\partial t}\right\} + \{\{Q_1, H\}, Q_2\} + \{Q_1, \{Q_2, H\}\} \end{aligned} \quad (4.47)$$

where we have used the Jacobi identity to recast the last term of the first line into the last two terms of the second line. Finally, it suffices to combine the first and third term, and the second and fourth term, to obtain,

$$\frac{d}{dt}\{Q_1, Q_2\} = \{\dot{Q}_1, Q_2\} + \{Q_1, \dot{Q}_2\} \quad (4.48)$$

which must indeed vanish in view of the assumption that both Q_1 and Q_2 are conserved, so that $\dot{Q}_1 = \dot{Q}_2 = 0$. Thus, $\{Q_1, Q_2\}$ is conserved. We shall later give a general definition of a Lie algebra, and see that the relations established already above are precisely such that the conserved Noether charges form a Lie algebra.

Poisson's theorem shows that if we know two infinitesimal symmetry transformations $\delta_1 q$ and $\delta_2 q$, with associated Noether charges Q_1 and Q_2 , then $\{Q_1, Q_2\}$ is conserved. It may of course be the case that $\{Q_1, Q_2\} = 0$, or more generally that $\{Q_1, Q_2\}$ is a function of Q_1 and Q_2 , so that we do not get any new charge from this procedure. But if $\{Q_1, Q_2\}$ is functionally independent of Q_1 and Q_2 , then we obtain a new Noether charge, and we would expect a new infinitesimal symmetry transformation $\delta_{12} q$. Thus, the question arises as to how to compute the infinitesimal symmetry transformation from the Noether charge.

4.10 Noether charge reproduces the symmetry transformation

The fundamental result is as follows. If Q is a conserved Noether charge, given by (4.45), and expressed in terms of the variables $(q, p; t)$,

$$Q(q, p; t) = \sum_{i=1}^N p_i \delta q_i - \Lambda(q, p; t) \quad (4.49)$$

then the associated infinitesimal transformation is given by,

$$\begin{aligned} \delta q_i &= \{q_i, Q\} \\ \delta p_i &= \{p_i, Q\} \end{aligned} \quad (4.50)$$

We shall prove the first relation, the second one being analogous. First, using (4.49), we compute,

$$\{q_i, Q\} = \frac{\partial Q}{\partial p_i} = \delta q_i + \sum_{j=1}^N p_j \frac{\partial \delta q_j}{\partial p_i} - \frac{\partial \Lambda}{\partial p_i} \quad (4.51)$$

To obtain the last term, we use its defining relation, now expressed in terms of variables $(q, p; t)$, and by expanding this relation we find,

$$\sum_{i=1}^N \left(\frac{\partial L}{\partial q_i} \delta q_i + \frac{\partial L}{\partial \dot{q}_i} \delta \dot{q}_i \right) = \frac{\partial \Lambda}{\partial t} + \sum_{i=1}^N \left(\frac{\partial \Lambda}{\partial q_i} \dot{q}_i + \frac{\partial \Lambda}{\partial \dot{q}_i} \ddot{q}_i \right) \quad (4.52)$$

Expressing $\delta\dot{q}_i = d(\delta q_i)/dt$ in terms of its independent variables, we obtain,

$$\delta\dot{q}_i = \frac{\partial\delta q_i}{\partial t} + \sum_{j=1}^N \left(\frac{\partial\delta q_i}{\partial q_j} \dot{q}_j + \frac{\partial\delta q_i}{\partial \dot{q}_j} \ddot{q}_j \right) \quad (4.53)$$

Recalling that this condition for a symmetry must hold for all q_i , whether or not they satisfy the Euler-Lagrange equations, we may identify both sides of (4.52) using (4.53), and we find for the term proportional to \ddot{q}_i ,

$$\sum_{i=1}^N \frac{\partial L}{\partial \dot{q}_i} \frac{\partial\delta q_i}{\partial \dot{q}_j} = \frac{\partial\Lambda}{\partial \dot{q}_j} \quad (4.54)$$

Changing variables to $(q, p; t)$, we find,

$$\frac{\partial\Lambda}{\partial p_i} = \sum_{j=1}^N \frac{\partial\Lambda}{\partial \dot{q}_j} \frac{\partial \dot{q}_j}{\partial p_i} \quad (4.55)$$

Using now relation (4.54), we find,

$$\frac{\partial\Lambda}{\partial p_i} = \sum_{j,k=1}^N p_j \frac{\partial \dot{q}_k}{\partial p_i} \frac{\partial\delta q_j}{\partial \dot{q}_k} \quad (4.56)$$

Putting all together, we have,

$$\{q_i, Q\} = \delta q_i + \sum_{j=1}^N p_j \left(\frac{\partial\delta q_j}{\partial p_i} - \sum_{k=1}^N \frac{\partial \dot{q}_k}{\partial p_i} \frac{\partial\delta q_j}{\partial \dot{q}_k} \right) \quad (4.57)$$

By the same argument that we derived (4.55), we also find that

$$\frac{\partial\delta q_j}{\partial p_i} = \sum_{k=1}^N \frac{\partial \dot{q}_k}{\partial p_i} \frac{\partial\delta q_j}{\partial \dot{q}_k} \quad (4.58)$$

so that we indeed recover the first line of (4.50), and prove our theorem.

5 Lie groups and Lie algebras

In this section, we shall discuss the role of Lie groups and Lie algebras in classical mechanics. We begin by defining the algebraic concept of a group, then produce some of the groups most frequently used in physics within the context of the multi-dimensional harmonic oscillator system, and finally give a general definition of a Lie group, and of a Lie algebra.

5.1 Definition of a group

A group G, \star consists of a set G and a *multiplication* law \star , defined between any pair of elements $g_1, g_2 \in G$. It is denoted by $g_1 \star g_2$, and is subject to the following axioms,

1. Multiplication is closed in the set G : the product $g_1 \star g_2 \in G$ for all $g_1, g_2 \in G$;
2. Associativity: the triple product obeys $(g_1 \star g_2) \star g_3 = g_1 \star (g_2 \star g_3)$ for all $g_1, g_2, g_3 \in G$;
3. G contains a unit element e such that $e \star g = g \star e = g$ for all $g \in G$;
4. Every $g \in G$ has an inverse, denoted g^{-1} , so that $g \star g^{-1} = g^{-1} \star g = e$.

A few comments are in order. (a) When a multiplication law is associative, the parentheses in the triple product may be dropped altogether, $g_1 \star g_2 \star g_3 = (g_1 \star g_2) \star g_3 = g_1 \star (g_2 \star g_3)$ since no ambiguity can arise. (b) The unit element unique. (c) The inverse of each $g \in G$ is unique. These properties define a group G, \star , or often simply denoted by G when the multiplication law is clearly specified.

- Direct product of groups: If G' and G'' are groups, respectively with operations \star' and \star'' , then the direct product $G = G' \times G''$ also forms a group with the natural combined operation $g_1 \star g_2 = (g'_1, g''_1) \star (g'_2, g''_2) = (g'_1 \star' g'_2, g''_1 \star'' g''_2)$ for all $g'_1, g'_2 \in G'$ and $g''_1, g''_2 \in G''$.
- Subgroups: A subset $H \subset G$ is referred to as a *subgroup* of G provided H, \star is a group.
 - H is an *invariant subgroup* of G , if $g \star h \star g^{-1} \in H$, for all $g \in G$ and $h \in H$. For example, if G is a product group, with $G = G' \times G''$, then both G' and G'' are invariant subgroups of G . But this is not the most general situation; the product may also be semi-direct, as it is in the case of the group of Euclidean motions (translations and rotations) in 3-space. The group of translations is an invariant subgroup, but the product is not direct.
 - A group G is defined to be *simple* if the only invariant subgroups of G are trivial, namely the subgroup $\{e\}$, and G itself. Simple groups form building blocks into which general groups may be decomposed, as direct (or semi-direct) products.

- Abelian versus non-Abelian groups: If, for all pairs $g_1, g_2 \in G$, we have

$$g_1 \star g_2 = g_2 \star g_1 \quad (5.1)$$

then the group G, \star is said to be Abelian (or commutative). In the contrary case (when there exists at least one pair $g_1, g_2 \in G$ such that $g_1 \star g_2 \neq g_2 \star g_1$) then the group is said to be non-Abelian (or non-commutative). Both Abelian and non-Abelian groups play fundamental roles throughout physics.

- There is an important topological distinction between different types of groups. A group is said to be *discrete* if all of its points are isolated; otherwise, it is *continuous*.
- Elementary examples of Abelian groups are as follows,

$$\mathbf{Z}_{,+} \subset \mathbf{Q}_{,+} \subset \mathbf{R}_{,+} \subset \mathbf{C}_{,+} \quad \mathbf{Q}^0_{, \times} \subset \mathbf{R}^0_{, \times} \subset \mathbf{C}^0_{, \times} \quad (5.2)$$

where the superscript 0 indicates that the zero element is to be removed from the set. The inclusion relations indicated in (5.2) correspond to subgroups. The groups \mathbf{Z}, \mathbf{Q} are discrete, while \mathbf{R}, \mathbf{C} are continuous.

- The set of all $M \times N$ matrices with addition, either with integer, rational, real, or complex entries, also forms an Abelian group, for all values of M, N . Note that addition of matrices is a special case of multiplication of matrices. Indeed, let A and B be two $M \times N$ matrices, and introduce the following matrices of dimension $(M+N) \times (M+N)$,

$$\mathcal{A} = \begin{pmatrix} I_M & A \\ 0 & I_N \end{pmatrix} \quad \mathcal{B} = \begin{pmatrix} I_M & B \\ 0 & I_N \end{pmatrix} \quad (5.3)$$

Then the product of the matrices \mathcal{A} and \mathcal{B} corresponds to the sum of A and B ,

$$\mathcal{A}\mathcal{B} = \begin{pmatrix} I_M & A+B \\ 0 & I_N \end{pmatrix} \quad (5.4)$$

Henceforth, we shall view matrix addition as a special case of matrix multiplication.

5.2 Matrix multiplication groups

Matrix multiplication groups are the most important groups used in physics. Square matrices, of size $N \times N$ for any positive integer $N = 1, 2, 3, \dots$, close under matrix multiplication, and matrix multiplication is *always associative*. The identity element is the identity $N \times N$ matrix, and will be denoted either by I_N , or simply by I when no confusion is expected to

arise. Not every $N \times N$ matrix has an inverse though. To be invertible, a matrix M must have non-zero determinant.³ This leads us to introduce the *general linear groups* of $N \times N$ invertible matrices (with non-zero determinant),

$$Gl(N, \mathbf{Q}) \subset Gl(N, \mathbf{R}) \subset Gl(N, \mathbf{C}) \quad (5.5)$$

respectively with entries taking values in \mathbf{Q} , \mathbf{R} , and \mathbf{C} . Rescaling any invertible matrix M by a non-zero number (corresponding to the number field in which the matrix entries take values) yields again an invertible matrix. This makes it natural and useful to introduce groups for which all matrices have determinant 1, a requirement that is more or less systematically indicated with the letter S preceding the group name. This leads us to introduce the *special linear groups* of $N \times N$ matrices with unit determinant,

$$Sl(N, \mathbf{Z}) \subset Sl(N, \mathbf{Q}) \subset Sl(N, \mathbf{R}) \subset Sl(N, \mathbf{C}) \quad (5.6)$$

The discrete groups built on \mathbf{Z} and \mathbf{Q} do sometimes enter into physics, but only in rather special settings. Henceforth, we shall concentrate on the continuous groups with either real or complex entries.

There are three series of *classical matrix groups*, each of which is defined through a quadratic relation on its elements.

- The *orthogonal group* $O(N)$ consists of $N \times N$ real matrices M which obey the relation,

$$M^t I_N M = M^t M = I_N \quad (5.7)$$

Considering the relation (5.7) on complex matrices M , the group is denoted by $O(N, \mathbf{C})$ instead. To verify that the set of matrices M obeying (5.7) forms a group (either with real or complex-valued entries), it suffices to check that matrix multiplication preserves (5.7), and that each M has an inverse. To check the first, we use $(M_1 M_2)^t (M_1 M_2) = M_2^t M_1^t M_1 M_2 = M_2^t M_2 = I_N$. To check invertibility, we take the determinant of relation (5.7), and find $(\det M)^2 = 1$, proving invertibility of each element.

– The group $SO(N)$ is defined with the additional requirement $\det M = 1$.

- The *unitary group* $U(N)$ consists of $N \times N$ complex matrices M which obey the relation

$$M^\dagger I_N M = M^\dagger M = I_N \quad M^\dagger \equiv (M^*)^t \quad (5.8)$$

The proof is elementary. Taking again the determinant, we have $|\det M|^2 = 1$, so that M is invertible. To check closure, we compute $(M_1 M_2)^\dagger M_1 M_2 = M_2^\dagger M_1^\dagger M_1 M_2 = I_N$.

³In these notes, N will almost always be taken to be finite; for definitions, problems, and properties involving infinite N , see my notes on Quantum Mechanics.

- The group $SU(N)$ is defined with the additional requirement $\det M = 1$.
- These groups are of fundamental importance in physics and enter frequently. For example the Standard Model of particle physics is based on a Yang-Mills theory with gauge group $SU(3)_c \times SU(2)_L \times U(1)_Y$.
- The *symplectic group* $Sp(2N)$ consists of $2N \times 2N$ matrices M which obey the relation,

$$M^t \mathcal{J} M = \mathcal{J} \quad \mathcal{J} = \mathcal{J}_{2N} = \begin{pmatrix} 0 & -I_N \\ I_N & o \end{pmatrix} \quad (5.9)$$

where \mathcal{J} is the *symplectic matrix*, encountered already as the Poisson bracket on all of phase space. To prove invertibility, we take the determinant of (5.9), and find $(\det M)^2 = 1$. We also have, $(M_1 M_2)^t \mathcal{J} (M_1 M_2) = M_2^t M_1^t \mathcal{J} M_1 M_2 = M_2^t \mathcal{J} M_2 = \mathcal{J}$. The symplectic group considered over the complex is denoted by $Sp(2N, \mathbf{C})$.

- One also encounters the group $USp(2N) = Sp(2N, \mathbf{C}) \cap SU(2N)$.

5.3 Orthonormal frames and parametrization of $SO(N)$

The importance of the group $SO(N)$ derives from the fact that it preserves the Euclidean inner product and distance in the N -dimensional vector space \mathbf{R}^N . Parametrizing vectors in \mathbf{R}^N by column matrices, and the inner product by matrix products,

$$(X, Y) = X^t Y = \sum_{i=1}^N x_i y_i \quad X = \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_N \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_N \end{pmatrix} \quad (5.10)$$

relation (5.7) is equivalent to the invariance on the inner product,

$$(MX, MY) = (X, Y) \quad \text{for all } X, Y \quad (5.11)$$

The Euclidean distance between two points defined by $d(X, Y) = \sqrt{(X - Y, X - Y)}$ is then also invariant, $d(MX, MY) = d(X, Y)$. In particular, a pair of orthogonal vectors is mapped into a pair of orthogonal vectors. This last fact gives us a very natural way of viewing orthogonal transformations physically, and parametrizing them explicitly. To see this, introduce the following basis for \mathbf{R}^N ,

$$V_1 = \begin{pmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix} \quad V_2 = \begin{pmatrix} 0 \\ 1 \\ \cdot \\ \cdot \\ 0 \end{pmatrix} \quad \cdots \quad V_N = \begin{pmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ 1 \end{pmatrix} \quad (5.12)$$

This basis is orthonormal, namely we have $(V_i, V_j) = \delta_{ij}$.

We now view an orthogonal matrix $M \in SO(N)$ as a *transformation* on vectors in \mathbf{R}^N . In particular, M will transform each basis vector V_i into a new vector $V'_i = MV_i$. As a result of (5.11), we have $(V'_i, V'_j) = \delta_{ij}$. Thus, the transformed basis vectors V'_i also form an orthonormal basis. In fact, the components of the matrices V'_1, V'_2, \dots, V'_N are precisely the entries of the matrix M ,

$$V'_i = MV_i = \begin{pmatrix} m_{1i} \\ m_{2i} \\ \vdots \\ m_{Ni} \end{pmatrix} \quad M = \begin{pmatrix} m_{11} & m_{12} & \cdots & m_{1N} \\ m_{21} & m_{22} & \cdots & m_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ m_{N1} & m_{N2} & \cdots & m_{NN} \end{pmatrix} \quad (5.13)$$

For $\det M = 1$ the orientation of the new frame is preserved, while for $\det M = -1$ the orientation is reversed. The group $SO(N)$ maps an oriented orthonormal frame into an orthonormal frames with the same orientation. Specifically, this map is one-to-one and onto.

Conversely, viewing $M \in SO(N)$ as a transformation on the unit vectors of an orthonormal frame gives us a natural parametrization of all orthogonal matrices. The parametrization proceeds as follows. We begin by parametrizing a general N -dimensional unit vector V in a recursive manner,

$$V = \begin{pmatrix} v \sin \phi \\ \cos \phi \end{pmatrix} \quad (5.14)$$

where ϕ is an angle taking values in the interval $\phi \in [0, \pi]$, and v is a unit vector in dimension $N - 1$, represented here by an $N - 1$ -dimensional column matrix. The vector v may in turn be parametrized in this recursive manner, and so on. In all, the parametrization of an N -dimensional unit vector requires $N - 1$ angles. Note, however, that a two-dimensional unit vector, which is the last step in the recursion, is parametrized by,

$$v = \begin{pmatrix} \sin \psi \\ \cos \psi \end{pmatrix} \quad (5.15)$$

but where the angle $\psi \in [0, 2\pi]$. To parametrize the full matrix $M \in SO(N)$, we begin by parametrizing the last column V'_N with the help of the above recursive method. Next, we parametrize V'_{N-1} . This is also a unit vector, which may be parametrized recursively by the above method as well, but we must now insist on one relation, namely the orthogonality with V'_N , namely $(V'_{N-1}, V'_N) = 0$. But this relation may be easily enforced by fixing one of the $N - 1$ angles in the general unit vector parametrization of V'_{N-1} . Next, V'_{N-2} is a unit vector,

but it must be orthogonal to both V'_{N-1} and V'_N , and so on. In total, the parametrization of $SO(N)$ requires

$$(N-1) + (N-2) + (N-3) + \cdots + 1 = \frac{1}{2}N(N-1) \quad (5.16)$$

angles, with ranges as indicated by the recursive construction of each unit vector.

The total number of independent angles needed to parametrize $SO(N)$ is the *dimension* of the group. This number could also have been derived directly, by looking at the defining relation $M^t M = I$. The matrix M has N^2 real entries. The equation $M^t M = I$ imposes N^2 relations, which are not all independent. Since the matrix $M^t M$ is automatically symmetric, it imposes only $N(N+1)/2$ relations, leaving $N^2 - N(N+1)/2 = N(N-1)/2$ independent parameters.

5.4 Three-dimensional rotations and Euler angles

For the case $N = 3$, corresponding to rotations of customary 3-dimensional space, the resulting parametrization is equivalent to Euler angles. To see this, we introduce rotations around the coordinate axes x and z , as follows,

$$R_x(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & \sin \theta \\ 0 & -\sin \theta & \cos \theta \end{pmatrix} \quad R_z(\phi) = \begin{pmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (5.17)$$

The parametrization of the most general rotation of $SO(3)$ is then obtained as follows,

$$M = M(\phi, \theta, \psi) = R_z(\psi)R_x(\theta)R_z(\phi) \quad (5.18)$$

and is given by

$$M = \begin{pmatrix} \cos \psi \cos \phi - \cos \theta \sin \phi \sin \psi & \cos \psi \sin \phi + \cos \theta \cos \phi \sin \psi & \sin \psi \sin \theta \\ -\sin \psi \cos \phi - \cos \theta \sin \phi \cos \psi & -\sin \psi \sin \phi + \cos \theta \cos \phi \cos \psi & \cos \psi \sin \theta \\ \sin \theta \sin \phi & -\sin \theta \cos \phi & \cos \theta \end{pmatrix}$$

Notice that indeed, the last column is the most general unit 3-vector, parametrized by two independent angles, with ranges $\psi \in [0, 2\pi]$ and $\theta \in [0, \pi]$. The first two columns are unit vectors orthogonal to the last column.

5.5 Definition of a Lie group

All of the above matrix groups whose entries are real or complex numbers, are examples of Lie groups. Thus, it is appropriate to give a more general definition of a Lie group, to later expound its connection with Lie algebras.

A Lie group G is a group whose elements may be parametrized (at least locally, in the neighborhood of every point of G) by a set of real parameters (s_1, s_2, \dots, s_d) ,

$$g(s) = g(s_1, s_2, \dots, s_d) \quad (5.19)$$

in such a way that the product and inverse functions, conveniently combined as follows,

$$g(s) \star g^{-1}(t) = g(\sigma(s, t)) \quad \sigma(s, t) = (\sigma_1(s, t), \sigma_2(s, t), \dots, \sigma_d(s, t)) \quad (5.20)$$

is governed by *real analytic* composition functions $\sigma(s, t)$.

A powerful theorem of Sophus Lie (Norway, 1842-1899) states that if the functions σ are merely assumed to be continuous, then they are automatically real analytic.

5.6 Definition of a Lie algebra

A Lie algebra \mathcal{G} is a vector space endowed with a pairing, usually denoted by the bracket $[\cdot, \cdot]$ (the same symbol as the commutator). The defining properties that make this vector space with the pairing a Lie algebra are as follows. For all $X, Y, Z \in \mathcal{G}$, we must have,

1. The pairing takes values in \mathcal{G} , namely $[X, Y] \in \mathcal{G}$;
2. The pairing is *anti-symmetric*, $[X, Y] = -[Y, X]$;
3. The pairing is bilinear, $[(\lambda X + \mu Y), Z] = \lambda[X, Z] + \mu[Y, Z]$ for all λ, μ in the field of the vector space \mathcal{G} . Linearity in the second entry then follows from anti-symmetry;
4. The Jacobi identity $[[X, Y], Z] + [[Y, Z], X] + [[Z, X], Y] = 0$ holds.

Two fundamental examples of Lie algebra pairings are given by the commutator $[\cdot, \cdot]$ on square matrices, in which case the Jacobi identity holds trivially; the other is the Poisson bracket, for which we had established the Jacobi identity earlier.

5.7 Relating Lie groups and Lie algebras

Given a Lie group G , it is straightforward to construct the associated Lie algebra \mathcal{G} , by expanding the group multiplication law in the neighborhood of the identity element of G . The vector space of the Lie algebra is then the *tangent space* to G at the identity, whose elements are tangent vectors. We shall use parameters such that the identity of G corresponds to $s = (s_1, s_2, \dots, s_d) = 0$. For matrix groups, one may simply expand the matrices around the identity, and one has,⁴

$$g(s) = I + \sum_{i=1}^d s_i X_i + \mathcal{O}(s^2) \quad (5.21)$$

The tangent vectors X_i form a basis for the Lie algebra \mathcal{G} . To first order in parameters, the product in G amounts to addition in the Lie algebra,

$$g(s)g(t) = I + \sum_{i=1}^d (s_i + t_i) X_i + \mathcal{O}(s^2, st, t^2) \quad (5.22)$$

so that to linear order, the inverse is obtained by reversing the sign of the parameters. To second order, the commutator is produced. It is convenient to consider the combination,

$$g(s)g(t)g(s)^{-1}g(t)^{-1} = I + \sum_{i,j=1}^d s_i t_j [X_i, X_j] + \mathcal{O}(s^3, s^2 t, st^2, t^3) \quad (5.23)$$

But now, for all infinitesimal values of the parameters s, t , this combination must again be an element of the Lie algebra. Thus, the commutator must take values in \mathcal{G} , as required by the first axiom of the Lie algebra structure. In particular, we may decompose the commutator itself onto the basis elements X_i by,

$$[X_i, X_j] = \sum_{k=1}^d f_{ijk} X_k \quad (5.24)$$

The coefficients f_{ijk} are the *structure constants* of the Lie algebra. By construction, they are anti-symmetric in ij . The Jacobi identity imposes the following condition,

$$\sum_{m=1}^d (f_{ijm} f_{mkn} + f_{jkm} f_{min} + f_{kim} f_{mjn}) = 0 \quad (5.25)$$

⁴When the Lie group is defined more abstractly, one needs to define the tangent space more carefully, customarily in terms of one-parameter subgroups.

for all $i, j, n = 1, \dots, d$. Associativity of the group multiplication law implies the Jacobi identity at the level of the algebra. In conclusion, to any given Lie group G , there is associated a unique Lie algebra \mathcal{G} obtained by expanding around the identity element of G .

A powerful theorem of Sophus Lie informs on the converse. Given a Lie algebra \mathcal{G} , there exists a unique *simply connected* Lie group of which \mathcal{G} is the Lie algebra in the sense defined above.⁵ For example, you may check that the Lie groups $SU(2)$ and $SO(3)$ are different, and yet their Lie algebras is common and is just the angular momentum algebra; the same holds for $SU(4)$ and $SO(6)$.

The power of the theory of Lie groups and Lie algebras lies in the fact that continuous symmetries in physics form Lie groups, but these can be studied, up to global topological issues such as simple-connectedness, by studying the “linearize problem” of Lie algebras.

5.8 Symmetries of the degenerate harmonic oscillator

Consider the N -dimensional *degenerate* harmonic oscillator, governed by the Lagrangian,

$$L = \sum_{i=1}^N \frac{1}{2} m (\dot{p}_i^2 - \omega^2 q_i^2) \quad (5.26)$$

Note that this is a quadratic problem, but a very special one, since the frequencies of the N oscillators coincide. A special case is when $N = 3$, and the generalized coordinates coincide with the Cartesian coordinates of a single particle, with Lagrangian,

$$L = \frac{1}{2} m \dot{\mathbf{x}}^2 - \frac{1}{2} m \omega^2 \mathbf{x}^2 \quad (5.27)$$

The potential term is invariant under 3-dimensional rotations. By analogy, the Lagrangian of (5.19) is invariant under rotations in N -dimensional space.

• Orthogonal symmetry

To see how this works, it will be convenient to recast the Lagrangian of (5.1), and the associated Hamiltonian in matrix form, by introducing,

$$Q = \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_N \end{pmatrix} \quad P = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_N \end{pmatrix} \quad (5.28)$$

⁵Simple-connectedness is a topological concept. A space G is simply connected if every closed path in G can be shrunk to a point through a continuous sequence of closed paths which lie entirely inside G .

so that

$$\begin{aligned} L &= \frac{1}{2}m\dot{Q}^t\dot{Q} - \frac{1}{2}m\omega^2 Q^t Q \\ H &= \frac{1}{2m}P^t P + \frac{1}{2}m\omega^2 Q^t Q \end{aligned} \quad (5.29)$$

An orthogonal transformation $R \in O(N)$ satisfies $T^t R = I$ and acts as follows on Q and P ,

$$\begin{aligned} Q &\rightarrow Q' = RQ & (Q')^t(Q') &= Q^t Q \\ P &\rightarrow P' = RP & (P')^t(P') &= P^t P \end{aligned} \quad (5.30)$$

Thus, the kinetic and potential parts of L and H are separately invariant. The Poisson bracket $\{q_i, p_j\} = \delta_{ij}$ is also invariant. The Noether charge associated with the infinitesimal rotations takes the form of generalized angular momentum,

$$L_{ij} = q_i p_j - q_j p_i \quad (5.31)$$

and satisfies the Lie algebra relations of $SO(N)$, namely,

$$\{L_{ij}, L_{kl}\} = \delta_{ik}L_{jl} - \delta_{jk}L_{il} - \delta_{il}L_{jk} + \delta_{jl}L_{ik} \quad (5.32)$$

The infinitesimal transformation is again generated by the Poisson bracket,

$$\delta q_i = \{q_i, L\} = - \sum_{k=1}^N \omega_{ik} q_k \quad L = \frac{1}{2} \sum_{k,l=1}^N \omega_{kl} L_{kl} \quad (5.33)$$

where the anti-symmetric coefficients $\omega_{ji} = -\omega_{ij}$ parametrize the infinitesimal rotation.

• Unitary symmetry

By a method familiar from quantum mechanics, we can recast the Hamiltonian in terms of “raising and lowering” variables, defined by,

$$\begin{aligned} A &= \frac{1}{\sqrt{2m\omega}}(iP + m\omega Q) \\ A^* &= \frac{1}{\sqrt{2m\omega}}(-iP + m\omega Q) \end{aligned} \quad (5.34)$$

In terms of A and A^* , and their matrix entries a_i and a_i^* , the Hamiltonian and Poisson brackets become,

$$\begin{aligned} H &= \omega A^\dagger A & \{a_i, a_j\} &= 0 \\ & & \{a_i, a_j^*\} &= -i\delta_{ij} \end{aligned} \quad (5.35)$$

In this form, it is clear that we may make on A an arbitrary *unitary* transformation $g \in U(N)$,

$$\begin{aligned} A &\rightarrow A' = gA \\ A^\dagger &\rightarrow (A')^\dagger = A^\dagger g^\dagger \end{aligned} \quad (5.36)$$

and H as well as the Poisson bracket are invariant under this transformation. The corresponding Noether charges are given by,

$$Q_T = A^\dagger T A \quad (5.37)$$

where T is any Hermitian $N \times N$ matrix (namely the generators of the Lie algebra of $U(N)$). These charges obey the following composition law under Poisson brackets,

$$\{Q_{T_1}, Q_{T_2}\} = iQ_{[T_1, T_2]} \quad (5.38)$$

The invariance of the harmonic oscillator under $SO(N)$ is contained in the $U(N)$ symmetry of the oscillator, which is consistent with the fact that $SO(N)$ is a subgroup of $U(N)$.

• Symplectic Symmetry

To conclude, we point out the role of *symplectic* transformations. Suppose we introduced, as we had done already earlier, the combined coordinates of phase space, in the form of a column matrix of height $2N$ by,

$$X = \begin{pmatrix} P \\ m\omega Q \end{pmatrix} \quad (5.39)$$

whose entries will be denoted collectively by x_α with $\alpha = 1, \dots, 2N$. The Hamiltonian and Poisson brackets may then be recast as follows,

$$H = \frac{1}{2m} X^t X \quad \{x_\alpha, x_\beta\} = m\omega \mathcal{J}_{\alpha\beta} \quad (5.40)$$

where \mathcal{J} is the symplectic matrix introduced in (5.9) when discussing the symplectic group. Now, there is no doubt that H is invariant under the group $SO(2N)$ which is larger than the group $U(N)$ (their dimensions are respectively $N(2N-1)$ and N^2). But the Poisson bracket is naturally invariant under another group: $Sp(2N)$. So, the combined Hamiltonian structure of the Hamiltonian itself and the Poisson bracket are invariant under the intersection $SO(2N) \cap Sp(2N)$. The intersection is easily worked out: a matrix satisfying both $M^t M = I_{2N}$ and $M^t \mathcal{J}_{2N} M = \mathcal{J}_{2N}$ must commute with \mathcal{J}_{2N} . As a result, it must be of the form

$$\begin{aligned} M &= \begin{pmatrix} M_1 & M_2 \\ -M_2 & M_1 \end{pmatrix} & M_1^t M_1 + M_2^t M_2 &= I_N \\ & & M_1^t M_2 - M_2^t M_1 &= 0 \end{aligned} \quad (5.41)$$

But these are precisely the conditions for required to make the matrix $M_1 + iM_2$ to be unitary. Thus, we recover $SO(2N) \cap Sp(2N) = U(N)$, recovering our previous result.

5.9 The relation between the Lie groups $SU(2)$ and $SO(3)$

It is well-known from the angular momentum algebra of quantum mechanics that the Lie algebras of $SU(2)$ and $SO(3)$ coincide. And yet, the Lie groups are different, and the corresponding representation theories of both Lie groups are also different, namely $SU(2)$ has unitary representations for all non-negative half-integers $j = 0, \frac{1}{2}, 1, \frac{3}{2}, \dots$ while the (single-valued) representations of $SO(3)$ are required to have integer j , as is the case in orbital angular momentum.

In this subsection, we shall describe the precise map between the Lie groups, and this will allow us to see how the groups differ from one another. An arbitrary element $g \in SU(2)$ may be parametrized by two complex numbers $u, v \in \mathbf{C}$, as follows,

$$g = \begin{pmatrix} u & -v^* \\ v & u^* \end{pmatrix} \quad |u|^2 + |v|^2 = 1 \quad (5.42)$$

The pair (u, v) , subject to the above constraint, precisely parametrizes the unit sphere S^3 embedded in $\mathbf{R}^4 = \mathbf{C}^2$, the sense that this map is bijective. Thus we conclude that

$$SU(2) = S^3 \quad (5.43)$$

The key ingredient to the map between $SU(2)$ and $SO(3)$ is the basis of 2×2 traceless Hermitian matrices, given by the Pauli matrices,

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (5.44)$$

Now consider the composite $g\sigma^i g^\dagger$ for $i = 1, 2, 3$. Each of these 2×2 matrices is itself traceless and Hermitian, so that they can be decomposed onto Pauli matrices as follows,

$$g\sigma^i g^\dagger = \sum_{j=1}^3 R^i_j \sigma^j \quad (5.45)$$

Using the relation $\{\sigma^i, \sigma^j\} = 2\delta^{ij}I_2$ to compute the anti-commutators $\{g\sigma^i g^\dagger, g\sigma^j g^\dagger\}$ in two different ways,

$$\{\sigma^i, \sigma^j\} = g\{\sigma^i, \sigma^j\}g^\dagger = \{g\sigma^i g^\dagger, g\sigma^j g^\dagger\} = \sum_{k,\ell=1}^3 R^i_k R^j_\ell \{\sigma^k, \sigma^\ell\} \quad (5.46)$$

we establish that R must be orthogonal,

$$R^t R = I_3 \quad (5.47)$$

Equation (5.45) provides a map from $g \in SU(2)$ to $R \in SO(3)$. Every $R \in SO(3)$ is attained in this way, but g and $-g$ map to the same R , even though g and $-g$ are distinct elements of $SU(2)$. Thus the map is 2-to-1, and we have $SO(3) = SU(2)/Z_2$ where $Z_2 = \{1, -1\}$.

6 Motion of Rigid Bodies

As much as we may understand the equations governing the motion of rigid bodies, such as tops, bicycles, and gyroscopes, the actual physical motion is not intuitive, and continues to intrigue. In the photograph below Wolfgang Pauli and Niels Bohr are no longer young men, and yet still marvel at the behavior of a toy top.



6.1 Inertial and body-fixed frames

To derive the equations for the motion of rigid bodies, we begin by defining what we mean by a rigid body, and then carefully parametrizing their motion. A rigid body is an assembly of massive points, the relative distances of all of which remain unchanged in time. Thus, we may associate with a rigid body a frame in which all points of the rigid body are static. We shall denote this frame by Γ' , and its origin by O' , as indicated in figure 8. This frame is referred to as the *body-fixed frame* for obvious reasons. Of course, this frame is not unique, and both the origin may be shifted, and a new body-fixed frame may also be rotated with respect to Γ' . So, we shall pick any one such frame, and discuss later the effect of changing frames. The massive points of the rigid body may be parametrized by the N Cartesian

coordinate matrices X'_α where the index α labels the different massive points, with masses m_α , of the body. These coordinates are independent of time.

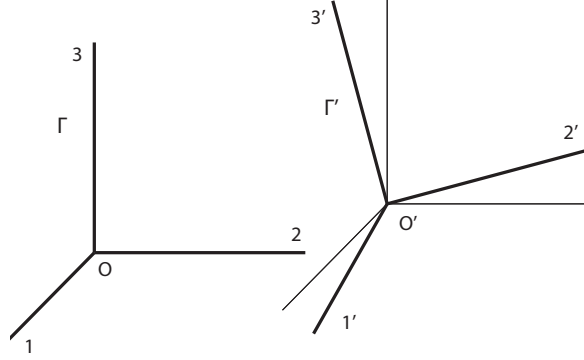


Figure 8: Inertial frame and body fixed frame

The motion of the rigid body involves two independent parts: overall translation, and overall rotation, as observed and parametrized from the point of view of an inertial frame Γ , with origin O . The translation part may be parametrized by giving the coordinates of the origin O' with respect to the inertial frame Γ ; we shall denote this by the matrix Y . We shall parametrize the rotation of Γ' with respect to Γ by the orthogonal matrix R . Together, Y and R completely parametrize the position of the body-fixed frame Γ' , and thus of the rigid body, with respect to the inertial frame Γ . In particular, a point with coordinates X'_α with respect to the body-fixed frame has coordinates X_α with respect to the inertial frame, where X_α is given by,

$$X_\alpha(t) = Y(t) + R(t)X'_\alpha \quad (6.1)$$

Here we have made time-dependence explicit, so as to clearly stress the time-independence of the body-fixed coordinates X'_α .

6.2 Kinetic energy of a rigid body

The kinetic energy, as observed from the point of view of the inertial frame Γ , of a set $\mathcal{A} = \{\alpha\}$ of massive points with masses m_α , is given by,

$$T = \sum_{\alpha \in \mathcal{A}} \frac{1}{2} m_\alpha \dot{X}_\alpha^t \dot{X}_\alpha \quad (6.2)$$

Substituting the parametrization of the rigid body motion of (6.1) into (6.2), we have,

$$T = \sum_{\alpha \in \mathcal{A}} \frac{1}{2} m_{\alpha} \left(\dot{Y}^t + X'_{\alpha}{}^t \dot{R}^t \right) \left(\dot{Y} + \dot{R} X'_{\alpha} \right) \quad (6.3)$$

The quantities \dot{Y} and \dot{R} do not depend on the particle index α , and may be brought out from under the summation. It will be natural to introduce the following quantities,

$$\begin{aligned} M &= \sum_{\alpha \in \mathcal{A}} m_{\alpha} \\ D &= \sum_{\alpha \in \mathcal{A}} m_{\alpha} X'_{\alpha} \\ Q &= \sum_{\alpha \in \mathcal{A}} m_{\alpha} X'_{\alpha} X'_{\alpha}{}^t \end{aligned} \quad (6.4)$$

Here, M is the *total mass*, which is just a number; D is the *mass dipole*, which is an N -component real vector; and Q is the *mass quadrupole*, which is an $N \times N$ real symmetric matrix. The mass quadrupole is closely related to the *inertia tensor* in 3 dimensions. In terms of the body-fixed quantities M, D, Q , the kinetic energy takes the form,

$$T = \frac{1}{2} M \dot{Y}^t \dot{Y} + \dot{Y}^t \dot{R} D + \frac{1}{2} \text{tr} \left(\dot{R} Q \dot{R}^t \right) \quad (6.5)$$

where we have used the relations, $D^t \dot{R}^t \dot{Y} = \dot{Y}^t \dot{R} D$ and $X'_{\alpha}{}^t \dot{R}^t \dot{R} X'_{\alpha} = \text{tr}(\dot{R}^t \dot{R} X'_{\alpha} X'_{\alpha}{}^t)$.

6.3 Angular momentum of a rigid body

Angular momentum, in the inertial frame, of each particle α in the rigid body is given by the familiar expression as an anti-symmetric matrix,

$$L_{\alpha} = X_{\alpha} P_{\alpha}^t - P_{\alpha} X_{\alpha}^t \quad (6.6)$$

as is easily established from the standard Noether formula with rotation transformation $\delta X = \rho X$ with $\rho^t = -\rho$. Adding up all angular momenta in the rigid body gives,

$$L = \sum_{\alpha \in \mathcal{A}} m_{\alpha} \left(X_{\alpha} \dot{X}_{\alpha}^t - \dot{X}_{\alpha} X_{\alpha}^t \right) \quad (6.7)$$

Choosing the origin of the frame to be the center of mass, we have $D = 0$, and the total angular momentum may be decomposed into two parts, $L = L_0 + L_I$ where,

$$\begin{aligned} L_0 &= M \left(Y \dot{Y}^t - \dot{Y} Y^t \right) \\ L_I &= R Q \dot{R}^t - \dot{R} Q R^t \end{aligned} \quad (6.8)$$

Clearly, L_0 is the overall orbital angular momentum of the rigid body seen from the inertial frame Γ , while L_I is the intrinsic part seen from the inertial frame.

6.4 Changing frames

First, consider changing the frame Γ . Shifting the origin O of the frame inertial frame, while keeping the frame Γ' fixed, amounts to the transformation $Y(t) \rightarrow Y(t) + Z$ for time-independent Z , which clearly leaves $\dot{Y}, \dot{R}, M, D, Q$ unchanged. Rotating the frame Γ amounts to a transformation,

$$\begin{aligned} Y(t) &\rightarrow SY(t) \\ R(t) &\rightarrow SR(t) \end{aligned} \tag{6.9}$$

where S is an arbitrary constant rotation matrix, satisfying $S^t S = I$. This transformation also leaves the kinetic energy invariant. Translations and rotation of Γ are *symmetries* of the problem.

Second, consider changing the frame Γ' . Shifting the origin O' of the body-fixed frame, while keeping the frame Γ fixed, amounts to the transformations,

$$\begin{aligned} X'_\alpha &\rightarrow X'_\alpha + Z \\ Y(t) &\rightarrow Y(t) - R(t)Z \end{aligned} \tag{6.10}$$

While the total mass is invariant, D and Q transform, in the following manner,

$$\begin{aligned} D &\rightarrow D + MZ \\ Q &\rightarrow Q + ZD^t + DZ^t + MZZ^t \end{aligned} \tag{6.11}$$

The kinetic energy is, of course, again invariant under these combined transformations.

By suitable shift Y , it is always possible to choose a frame in which $D = 0$. The point O' is then the *center of mass* or *barycenter* of the rigid body, and the expression for the kinetic energy then simplifies.

Finally, rotating the frame Γ' , while leaving Γ unchanged, amounts to a transformation,

$$\begin{aligned} X'_\alpha &\rightarrow SX'_\alpha & S^t S &= I \\ R(t) &\rightarrow R(t)S^t \end{aligned} \tag{6.12}$$

As a result, M and X_0 are unchanged, while D and Q transform as follows,

$$\begin{aligned} D &\rightarrow SD \\ Q &\rightarrow SQS^t \end{aligned} \tag{6.13}$$

The kinetic energy is again unchanged.

6.5 Euler-Lagrange equations for a freely rotating rigid body

The Lagrangian for a rigid body freely rotating around its center of mass reduces to its kinetic energy, and the corresponding action is given by,

$$S[R] = \frac{1}{2} \int dt \operatorname{tr} (\dot{R} Q \dot{R}^t) \quad (6.14)$$

It is most convenient to derive the equations of motion directly from the variational principle, rather than working out the Euler-Lagrange equations. The most general infinitesimal variation of an orthogonal matrix R is NOT a general variation of its entries, since the matrix needs to remain orthogonal. Instead, the most general variation δR must satisfy,

$$(R + \delta R)^t (R + \delta R) = I + \mathcal{O}(\delta R^2) \quad (6.15)$$

or equivalently that the combination $\delta R R^t$ is antisymmetric. Thus, the most general variation is of the form,

$$\delta R(t) = \varepsilon(t) R(t) \quad \varepsilon(t)^t = -\varepsilon(t) \quad (6.16)$$

where ε is an $N \times N$ matrix. Carrying out the variation of the action, we obtain,

$$\delta S[R] = \int dt \operatorname{tr} \left(\frac{d\delta R}{dt} Q \dot{R}^t \right) = - \int dt \operatorname{tr} (\varepsilon R Q \ddot{R}^t) \quad (6.17)$$

Since ε is a general *anti-symmetric* variation, the action principle requires,

$$R Q \ddot{R}^t - \ddot{R} Q R^t = 0 \quad (6.18)$$

It is immediate that this equation corresponds to the conservation of the total angular momentum L_I of (6.8), seen in the inertial frame,

$$\frac{dL_I}{dt} = \frac{d}{dt} (R Q \dot{R}^t - \dot{R} Q R^t) = R Q \ddot{R}^t - \ddot{R} Q R^t = 0 \quad (6.19)$$

One may also introduce the angular momentum L_B seen from the body-fixed frame,

$$L_B = R^t L_I R \quad (6.20)$$

Conservation of L_I is equivalent to the Euler equations for L_B , given by,

$$\dot{L}_B = [L_B, R^t \dot{R}] \quad (6.21)$$

We shall shortly exhibit these equations even more explicitly in the 3-dimensional case, where they will assume a familiar form.

6.6 Relation to a problem of geodesics on $SO(N)$

In physics, it is often remarkably instructive to relate two seemingly unrelated problems. Here, we shall show that the problem of the free motion of a rigid body can be related to a problem of *geodesics* on the group $SO(N)$, with respect to a metric which depends on the mass quadrupole Q . To see how this works, we state the relevant problem of geodesics, namely of determining the curves of extremal length. Consider the infinitesimal distance function on the group $SO(N)$ given by the following metric,

$$ds^2 = \text{tr} (dR Q dR^t) \quad (6.22)$$

where the differential dR is taken along $SO(N)$, and must thus satisfy $dR R^t + R dR^t = 0$. Note that this metric is positive definite. The length of an arc \mathcal{C} in $SO(N)$ is then given by,

$$\ell_{\mathcal{C}} = \int_{\mathcal{C}} ds = \int_{\mathcal{C}} \left[\text{tr} (dR Q dR^t) \right]^{\frac{1}{2}} \quad (6.23)$$

where the integrals are taken along the arc \mathcal{C} . The arc may be parametrized by an arbitrary parameter t , so that we have equivalently,

$$\ell_{\mathcal{C}} = \int dt \left[\text{tr} (\dot{R} Q \dot{R}^t) \right]^{\frac{1}{2}} \quad (6.24)$$

where $\dot{R} = dR/dt$. Were it not for the square root, this integral would be essentially the same as the action integral for the freely rotating rigid body. It turns out that, while the actions may not be the same, the solutions to the corresponding Euler-Lagrange equations coincide. To see this, we apply the variational calculation to $\ell_{\mathcal{C}}$, with $\delta R = \varepsilon R$, $\varepsilon^t = -\varepsilon$, and we find, after an integration by parts,

$$\delta \ell_{\mathcal{C}} = - \int dt \text{tr} \left(\varepsilon R Q \frac{d}{dt} \left\{ \frac{\dot{R}^t}{\left[\text{tr} (\dot{R} Q \dot{R}^t) \right]^{\frac{1}{2}}} \right\} \right) \quad (6.25)$$

Now since the parameter t may be chosen at will, we shall now choose it to coincide with the length parameter s , defined by (6.22). With this choice, a drastic simplification occurs in the variational problem, as we have,

$$\frac{\dot{R}^t}{\left[\text{tr} (\dot{R} Q \dot{R}^t) \right]^{\frac{1}{2}}} = \frac{dR^t}{ds} \quad (6.26)$$

and the Euler-Lagrange equations become,

$$R Q \frac{d^2 R^t}{ds^2} - \frac{d^2 R}{ds^2} Q R^t = 0 \quad (6.27)$$

But this is the same equation as (6.19), except for the substitution $t \rightarrow s$.

6.7 Solution for the maximally symmetric rigid body

For a general rigid body, with general value of Q , the Euler-Lagrange equations are not so easy to solve. There is one special case which is particularly amenable to exact solution, with natural geometrical interpretation, namely the case of the maximally symmetric rigid body with $Q = \mu I_N$. Here μ is a constant, and the conservation of angular momentum L_I then requires constancy of,

$$L_I = \mu(R\dot{R}^t - \dot{R}R^t) = -2\mu\dot{R}R^t \quad (6.28)$$

It is straightforward to solve this equation, given L_I and μ , and we have,

$$R(t) = R(t_0) \exp \left\{ -\frac{L_I}{2\mu}(t - t_0) \right\} \quad (6.29)$$

The anti-symmetry of L_I guarantees that $R(t)$ remains in $SO(N)$ at all times t .

6.8 The three-dimensional rigid body in terms of Euler angles

Three dimensions is the most useful case, and here we make contact with the more familiar parametrization in terms of Euler angles. To begin, we consider the rigid body in a frame where the center of mass is at rest. The kinetic energy may be recast as follows,

$$T = -\frac{1}{2} \text{tr} (R^t \dot{R} Q R^t \dot{R}) \quad (6.30)$$

a formula which makes explicit that rotations enter through the quantity $R^t \dot{R}$. Differentiating the relation $R^t R = I$, deduce that the combination $R^t \dot{R}$ is anti-symmetric,

$$(R^t \dot{R})^t = -R^t \dot{R} \quad (6.31)$$

Introducing a basis T_1, T_2, T_3 for all anti-symmetric 3×3 matrices by,

$$T_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \quad T_2 = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad T_3 = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (6.32)$$

we may decompose as follows,

$$R^t \dot{R} = \omega_1 T_1 + \omega_2 T_2 + \omega_3 T_3 \quad (6.33)$$

In terms of the angular velocities ω_i , the kinetic energy becomes,

$$T = \frac{1}{2} \sum_{i,j=1}^3 I_{ij} \omega_i \omega_j \quad (6.34)$$

where I_{ij} is the tensor of inertia, which is found to be given by,

$$I_{ij} = \delta_{ij}(\text{tr } Q) - Q_{ij} \quad (6.35)$$

which is a familiar formula.

Next, we evaluate the components of the angular rotation velocities ω_i in terms of Euler angles. The rotation matrix R takes the form familiar for Euler angles,

$$R(\phi, \theta, \psi) = R_z(\phi) R_x(\theta) R_z(\psi) \quad (6.36)$$

To work this out in Euler angles, we compute,

$$\begin{aligned} \dot{R} &= \dot{\phi} R'_z(\phi) R_x(\theta) R_z(\psi) + \dot{\theta} R_z(\phi) R'_x(\theta) R_z(\psi) + \dot{\psi} R_z(\phi) R_x(\theta) R'_z(\psi) \\ R^t \dot{R} &= R_z(\psi)^t \left(\dot{\psi} R'_z(\psi) R_z(\psi)^t + \dot{\theta} R_x(\theta)^t R'_x(\theta) + \dot{\phi} R_x(\theta)^t R_z(\phi)^t R'_z(\phi) R_x(\theta) \right) R_z(\psi) \end{aligned} \quad (6.37)$$

A prime denotes differentiation with respect to the argument. Evaluating the derivatives,

$$\begin{aligned} R'_z(\psi) R_z(\psi)^t &= T_3 \\ R_x(\theta)^t R'_x(\theta) &= T_1 \\ R_z(\phi)^t R'_z(\phi) &= T_3 \end{aligned} \quad (6.38)$$

we obtain the following expression,

$$R^t \dot{R} = R_z(\psi)^t \left(\dot{\psi} T_3 + \dot{\theta} T_1 + \dot{\phi} R_x(\theta)^t T_3 R_x(\theta) \right) R_z(\psi) \quad (6.39)$$

Three short calculations give,

$$\begin{aligned} R_z(\psi)^t T_1 R_z(\psi) &= T_1 \cos \psi + T_2 \sin \psi \\ R_z(\psi)^t T_2 R_z(\psi) &= -T_1 \sin \psi + T_2 \cos \psi \\ R_x(\theta)^t T_3 R_x(\theta) &= T_3 \cos \theta - T_2 \sin \theta \end{aligned} \quad (6.40)$$

Putting all together, we obtain indeed formula (6.33), with the following expressions for the angular velocities,

$$\begin{aligned} \omega_1 &= \dot{\phi} \sin \theta \sin \psi + \dot{\theta} \cos \psi \\ \omega_2 &= \dot{\phi} \sin \theta \cos \psi - \dot{\theta} \sin \psi \\ \omega_3 &= \dot{\phi} \cos \theta + \dot{\psi} \end{aligned} \quad (6.41)$$

6.9 Euler equations

So far, the choice of Euler angles was uncorrelated with the properties of the rigid body. It turns out to be advantageous to make a special choice which is well-adapted to the body, by letting its origin coincide with the center of mass of the body, and the axes coincide with the three *principal axes of the rigid body*. The principal axes correspond to the three orthogonal eigenvectors of the real symmetric matrix I . With such a choice the matrix I is diagonal,

$$I_{ij} = \begin{pmatrix} I_1 & 0 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & I_3 \end{pmatrix}_{ij} \quad (6.42)$$

and the kinetic energy becomes,

$$T = \frac{1}{2}I_1\omega_1^2 + \frac{1}{2}I_2\omega_2^2 + \frac{1}{2}I_3\omega_3^2 \quad (6.43)$$

The components of angular momentum, in the body-fixed frame, are given by,

$$L'_i = I_i\omega_i \quad i = 1, 2, 3 \quad (6.44)$$

The Euler equations are,

$$\dot{L}'_i = \sum_{j,k=1}^3 \varepsilon_{ijk} L'_j \omega_k \quad (6.45)$$

or in components,

$$\begin{aligned} I_1\dot{\omega}_1 &= (I_2 - I_3)\omega_2\omega_3 \\ I_2\dot{\omega}_2 &= (I_3 - I_1)\omega_3\omega_1 \\ I_3\dot{\omega}_3 &= (I_1 - I_2)\omega_1\omega_2 \end{aligned} \quad (6.46)$$

6.10 Poinsot's solution to Euler's equations

The solutions of Euler's equations may be grouped according to the symmetry of the tensor of inertia. The cases are as follows.

- For a *spherical top*, we have $I_1 = I_2 = I_3$, so that $\dot{\omega}_i$ is constant for all $i = 1, 2, 3$. Angular velocity is constant, and parallel to angular momentum.

- For a *symmetrical top*, we have for example $I = I_1 = I_2 \neq I_3$. As a result, we have $\dot{\omega}_3$ is constant. The remaining Euler equations then reduce to,

$$\begin{aligned} I\dot{\omega}_1 &= +(I - I_3)\omega_3\omega_2 \\ I\dot{\omega}_2 &= -(I - I_3)\omega_3\omega_1 \end{aligned} \quad (6.47)$$

Introducing complex variables, these equations decouple from one another,

$$I(\dot{\omega}_1 \pm i\dot{\omega}_2) = \mp i(I - I_3)\omega_3(\omega_1 \pm i\omega_2) \quad (6.48)$$

and these equations may be solved simply by,

$$(\omega_1 \pm i\omega_2)(t) = (\omega_1 \pm i\omega_2)(t_0)e^{\mp i\Omega(t-t_0)} \quad \Omega = \frac{I - I_3}{I}\omega_3 \quad (6.49)$$

• For the *asymmetrical top*, the principal moments of inertia are all different from one another, and we may choose $I_1 < I_2 < I_3$. Conservation of kinetic energy and of total angular momentum then dictate the following relations,

$$\begin{aligned} 2T &= I_1\omega_1^2 + I_2\omega_2^2 + I_3\omega_3^2 \\ \mathbf{L}^2 &= I_1^2\omega_1^2 + I_2^2\omega_2^2 + I_3^2\omega_3^2 \end{aligned} \quad (6.50)$$

As a result, we have the inequalities,

$$2TI_1 < \mathbf{L}^2 < 2TI_3 \quad (6.51)$$

One may solve for ω_1 and ω_3 in terms of ω_2 ,

$$\begin{aligned} \omega_1^2 &= \alpha_1 - \beta_1\omega_2^2 \\ \omega_3^2 &= \alpha_3 - \beta_3\omega_2^2 \end{aligned} \quad (6.52)$$

with the following definitions of constants,

$$\begin{aligned} \alpha_1 &= \frac{2TI_3 - \mathbf{L}^2}{I_1(I_3 - I_1)} & \beta_1 &= \frac{I_2(I_3 - I_2)}{I_1(I_3 - I_1)} \\ \alpha_3 &= \frac{\mathbf{L}^2 - 2I_1T}{I_3(I_3 - I_1)} & \beta_3 &= \frac{I_2(I_2 - I_1)}{I_3(I_3 - I_1)} \end{aligned} \quad (6.53)$$

which are all positive in view of the condition $I_1 < I_2 < I_3$, and the resulting inequality (6.51). Now use the Euler equation for ω_2 , and square it,

$$I_2^2\dot{\omega}_2^2 = (I_1 - I_3)^2(\alpha_1 - \beta_1\omega_2^2)(\alpha_3 - \beta_3\omega_2^2) \quad (6.54)$$

This is an equation of the elliptic type.

7 Special Relativity

Special relativity was created to resolve a conflict between Maxwell's equations on the one hand, and Newtonian mechanics on the other hand. One way of phrasing the conflict is that Maxwell's equations predict the speed of light to be the same to observers in different inertial frames, a result which is untenable in Newtonian mechanics. The resolution of the conflict proceeds by keeping Maxwell's equations intact, but modifying the laws of mechanics. The need to revise the laws of mechanics and the principle of special relativity had been foreseen by Poincaré already in 1900. He and Lorentz had also already derived the symmetry transformations of Maxwell's equations that carry their names. But it was Einstein who dispensed with the need for an ether medium for the propagation of electro-magnetic waves, and who formulated completely the laws of relativistic mechanics, including his mass-energy relation $E = mc^2$. Here, we shall give a brief summary of the theory of special relativity.

7.1 Basic Postulates

Special relativity is based on two postulates,

1. The laws of Nature and the results of all experiments in any two inertial frames, with relative velocity \mathbf{v} , are the same.
2. The speed of light c is independent of the relative velocity \mathbf{v} .

To make the consequences of these postulates explicit, we shall now spell out the properties of inertial frames and the relations between two such frames. An inertial frame in special relativity is a coordinate system $\mathcal{R}(t, \mathbf{x})$ in which Maxwell's equations *in the absence of matter* hold true.⁶ The coordinates of two inertial frames $\mathcal{R}(t, \mathbf{x})$ and $\mathcal{R}'(t', \mathbf{x}')$ are related to one another by an affine transformation, which includes translations of time and space (the affine part), as well as rotations and boosts (the linear part).

To make this more precise, we refer to an event as the simultaneous specification of the values of time and space with respect to a certain inertial frame, say $\mathcal{R}(t, \mathbf{x})$. We define the Minkowski distance s^2 between two events (t_1, \mathbf{x}_1) and (t_2, \mathbf{x}_2) in the same inertial frame by,

$$s^2 = -c^2(t_1 - t_2)^2 + (\mathbf{x}_1 - \mathbf{x}_2)^2 \quad (7.1)$$

The physical interpretation of s^2 depends upon its sign;

⁶Note the parallel with the definition of an inertial frame in Newtonian mechanics where the motion of a point mass in the absence of external forces is a straight line at uniform velocity.

- $s^2 = 0$, the events are causally related by the propagation of a light ray between them;
- $s^2 > 0$, the events are causally unrelated;
- $s^2 < 0$, the events are causally related by the propagation of particles of any mass; one may view τ , defined by $\tau^2 = -s^2/c^2$, as the proper time between the two events.

The coordinates of two inertial frames $\mathcal{R}(t, \mathbf{x})$ and $\mathcal{R}'(t', \mathbf{x}')$ are related by an affine transformation which leaves the Minkowski distance s^2 between any two events invariant,

$$-c^2(t_1 - t_2)^2 + (\mathbf{x}_1 - \mathbf{x}_2)^2 = -c^2(t'_1 - t'_2)^2 + (\mathbf{x}'_1 - \mathbf{x}'_2)^2 \quad (7.2)$$

It is immediate that this construction automatically implies the second postulate that the speed of light is c in all inertial frames. It is also immediate that space and time translations leave the Minkowski distance invariant. Amongst the linear transformations, rotations leave s^2 invariant as well. The remaining transformations are boosts, which act linearly. Using rotation symmetry, any boost may be rotated to the x -direction, leaving y and z -coordinates untransformed. We may then parametrize a boost as follows,

$$\begin{aligned} ct' &= Act + Bx \\ x' &= Cct + Dx \\ y' &= y \\ z' &= z \end{aligned} \quad (7.3)$$

with the coefficients A, B, C, D real. Denoting the two events by $(t_1, \mathbf{x}_1) = (t, x, 0, 0)$ and $(t_2, \mathbf{x}_2) = (0, 0, 0, 0)$ with respect to the inertial frame $\mathcal{R}(t, \mathbf{x})$, and by $(t'_1, \mathbf{x}'_1) = (t', x', 0, 0)$ and $(t'_2, \mathbf{x}'_2) = (0, 0, 0, 0)$ with respect to the inertial frame $\mathcal{R}'(t', \mathbf{x}')$ then requiring invariance of the Minkowski distance between the two events, gives the following relations,

$$\begin{aligned} -c^2t^2 + x^2 &= -c^2(t')^2 + (x')^2 \\ &= -(Act + Bx)^2 + (Cct + Dx)^2 \end{aligned} \quad (7.4)$$

Since this relation must hold for all possible events, it must hold for all t, x , which requires

$$\begin{aligned} 1 &= A^2 - C^2 \\ 1 &= -B^2 + D^2 \\ 0 &= -AB + CD \end{aligned} \quad (7.5)$$

Since the first equation implies $A^2 \geq 1$, we have either $A \geq 1$ or $A \leq -1$. The latter implies a reversal of the direction of time, which should be treated separately. Similarly, the solution

$D \leq -1$ corresponds to a reflection of space, and should also be treated separately. We are left with the branch $A, D \geq 1$, which may be parametrized in terms of a single parameter,

$$A = D = \gamma \equiv \frac{1}{\sqrt{1-\beta^2}} \quad B = C = -\beta\gamma \quad (7.6)$$

Reality of the coefficients A, B, C, D requires $\beta^2 < 1$. The transformations of (7.7) take the following form in this parametrization,

$$\begin{aligned} ct' &= \gamma(ct - \beta x) \\ x' &= \gamma(x - \beta ct) \\ y' &= y \\ z' &= z \end{aligned} \quad (7.7)$$

From this it is clear that βc must be interpreted as the relative velocity between the two frames. Indeed, the point $x' = 0$, which is fixed in frame \mathcal{R}' , travels with velocity $v = x/t = \beta c$ from the point of view of frame \mathcal{R} , which is one way we may define the relative velocity between the frames.

The relativistic transformation properties of momentum, energy, mass, and of the electromagnetic fields may be derived analogously. It is much more efficient, however, to obtain such relations using the Lorentz vector and tensor notation, which we shall provide next.

7.2 Lorentz vector and tensor notation

Just as we use vector notation in 3-dimensional space to collect the three coordinates (x, y, z) into a vector \mathbf{x} , so we use also 4-vector notation to collect the four coordinates of an event $(ct, x, y, z) = (ct, \mathbf{x})$ into a 4-vector denoted x , *without bold face or arrow embellishment*. Actually, one mostly uses a slight variant of the 4-vector notation, with an index added,

$$x^\mu \equiv (x^0, x^1, x^2, x^3) = (ct, x, y, z) \quad \mu = 0, 1, 2, 3 \quad (7.8)$$

The time direction being special through its signature in the Minkowski distance, one reserves the index “0” to denote it. The Minkowski distance may be easily cast in this notation,

$$s^2 = \sum_{\mu, \nu=0,1,2,3} \eta_{\mu\nu} (x_1^\mu - x_2^\mu)(x_1^\nu - x_2^\nu) \quad (7.9)$$

where the Minkowski metric $\eta_{\mu\nu}$ is defined as follows,

$$\eta_{\mu\nu} \equiv \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}_{\mu\nu} \quad (7.10)$$

Einstein introduced the *repeated index summation convention*, which instructs one to sum over one repeated upper and one repeated lower index. With the Einstein index convention, we thus have,

$$s^2 = \eta_{\mu\nu}(x_1^\mu - x_2^\mu)(x_1^\nu - x_2^\nu) \quad (7.11)$$

The precise distinction between upper and lower indices and their separate roles will be spelled out later.

A Poincaré transformation may be expressed in 4-vector notation as well. A general affine transformation between the coordinates x^μ and x'^μ may be parametrized as follows,

$$x^\mu \rightarrow x'^\mu = \Lambda^\mu{}_\nu x^\nu + a^\mu \quad (7.12)$$

Here, Λ is a 4×4 matrix with real entries, and the repeated ν -index is to be summed over. The 4-vector a^μ is real, and parametrizes translations in time and space. Invariance of the Minkowski distance s^2 is tantamount to,

$$\begin{aligned} \eta_{\mu\nu}x^\mu x^\nu &= \eta_{\rho\sigma}x'^\rho x'^\sigma \\ &= \eta_{\rho\sigma}\Lambda^\rho{}_\mu x^\mu \Lambda^\sigma{}_\nu x^\nu \end{aligned} \quad (7.13)$$

This relation must hold for all x^μ , so that we have,

$$\eta_{\mu\nu} = \eta_{\rho\sigma}\Lambda^\rho{}_\mu \Lambda^\sigma{}_\nu \quad (7.14)$$

This relation defines all Lorentz transformations. It is instructive to write out these relations in matrix form (without indices). Interpreting the 4-vector x now as a column matrix, the Minkowski distance, and the Poincaré transformations of (7.12) may then be expressed in matrix form,

$$\begin{aligned} s^2 &= x^t \eta x \\ x &\rightarrow x' = \Lambda x + a \end{aligned} \quad (7.15)$$

The condition of invariance of the Minkowski distance, given in (7.14) becomes simply,

$$\Lambda^t \eta \Lambda = \eta \quad (7.16)$$

and defines the group $SO(1,3)$, following the nomenclature of the section on group theory. It is instructive to check the number of independent Poincaré transformation parameters. The matrix Λ has 16 real components, but obeys 16 relations expressed through the equation of two matrices which are both automatically symmetric. A real symmetric 4×4 matrix has 10 real independent components, so Λ has $16 - 10 = 6$ independent parameters, which precisely accounts for the 3 rotations and 3 boosts which make up a general Lorentz transformation.

7.3 General Lorentz vectors and tensors

The starting point for introducing 4-vector notation in the preceding section was the quantity x^μ which, under a Lorentz transformation Λ , behaves linearly in x^μ ,

$$x^\mu \rightarrow x'^\mu = \Lambda^\mu{}_\nu x^\nu \quad (7.17)$$

One refers to any object $V^\mu = (V^0, V^1, V^2, V^3)$ which transforms under Λ by

$$A^\mu \rightarrow A'^\mu = \Lambda^\mu{}_\nu A^\nu \quad (7.18)$$

as a *Lorentz vector*. To be more precise, one sometimes refers to A^μ as a *Lorentz vector with upper index*, or a *contravariant vector*.

7.3.1 Contravariant tensors

The transformation law of the product of n vectors $x_i^{\mu_i}$ with $i = 1, 2, \dots, n$ follows from the transformation law of each vector, $x_i^\mu \rightarrow x_i'^\mu = \Lambda^\mu{}_\nu x_i^\nu$, and we have,

$$x_1^{\mu_1} \cdots x_n^{\mu_n} \rightarrow x_1'^{\mu_1} \cdots x_n'^{\mu_n} = (\Lambda^{\mu_1}{}_{\nu_1} \cdots \Lambda^{\mu_n}{}_{\nu_n}) x_1^{\nu_1} \cdots x_n^{\nu_n} \quad (7.19)$$

The product $x_1^{\mu_1} \cdots x_n^{\mu_n}$ is a *tensor of rank n* . One refers to any object $A^{\mu_1 \cdots \mu_n}$ which transforms under Λ by

$$A^{\mu_1 \cdots \mu_n} \rightarrow A'^{\mu_1 \cdots \mu_n} = (\Lambda^{\mu_1}{}_{\nu_1} \cdots \Lambda^{\mu_n}{}_{\nu_n}) A^{\nu_1 \cdots \nu_n} \quad (7.20)$$

as a Lorentz tensor of rank n , or more precisely as a Lorentz tensor with upper indices or a contravariant tensor of rank n . A special case is when $n = 0$, where we obtain,

$$A \rightarrow A' = A \quad (7.21)$$

namely a tensor of rank 0, more commonly referred to as a *Lorentz scalar*.

7.3.2 Covariant tensors

Every contravariant vector and tensor naturally has an associated *covariant vector* or tensor of the same rank, which is obtained by *lowering all upper indices* using the Minkowski metric $\eta_{\mu\nu}$. The simplest case is for a contravariant vector A^μ , where we define the associated covariant vector by

$$A_\mu \equiv \eta_{\mu\nu} A^\nu \quad \Leftrightarrow \quad A^\mu = \eta^{\mu\nu} A_\nu \quad (7.22)$$

Under a Lorentz transformation Λ , the covariant vector A_μ is mapped as follows,

$$A_\mu \rightarrow A'_\mu = \eta_{\mu\nu} A'^\nu = \eta_{\mu\nu} \Lambda^\nu{}_\rho A^\rho = \eta_{\mu\nu} \Lambda^\nu{}_\rho \eta^{\rho\sigma} A_\sigma \quad (7.23)$$

By our standard conventions of raising and lowering indices, we adopt the following notation,

$$\eta_{\mu\nu} \Lambda^\nu{}_\rho \eta^{\rho\sigma} = \Lambda_\mu{}^\sigma \quad (7.24)$$

Using the defining relations of Lorentz transformations, $\eta_{\mu\nu} = \eta_{\rho\sigma} \Lambda^\rho{}_\mu \Lambda^\sigma{}_\nu$, we may reinterpret this matrix as follows. Contract the defining relation with $\eta^{\mu\tau}$ gives,

$$\delta^\tau{}_\nu = \eta_{\rho\sigma} \Lambda^\rho{}_\mu \eta^{\mu\tau} \Lambda^\sigma{}_\nu = \Lambda_\sigma{}^\tau \Lambda^\sigma{}_\nu \quad (7.25)$$

Hence, $\Lambda_\mu{}^\nu$ is the inverse of the matrix $\Lambda^\mu{}_\nu$. Thus, another way of expressing the transformation law for a covariant vector is in terms of the inverse of Λ ,

$$A_\mu \rightarrow A'_\mu = \Lambda_\mu{}^\nu A_\nu \quad (7.26)$$

Analogously, one refers to any object $A_{\mu_1 \dots \mu_n}$ which transforms under Λ by

$$A_{\mu_1 \dots \mu_n} \rightarrow A'_{\mu_1 \dots \mu_n} = (\Lambda_{\mu_1}{}^{\nu_1} \dots \Lambda_{\mu_n}{}^{\nu_n}) A_{\nu_1 \dots \nu_n} \quad (7.27)$$

as a Lorentz tensor of rank n , or more precisely as a Lorentz tensor with lower indices or a covariant tensor of rank n .

One very important example of a covariant vector is provided by the 4-derivative,

$$\partial_\mu \equiv \frac{\partial}{\partial x^\mu} \quad (7.28)$$

In view of its defining relation, $[\partial_\mu, x^\nu] = \delta_\mu{}^\nu$, the object ∂_μ transforms as a covariant vector,

$$\partial'_\mu = \frac{\partial}{\partial x'^\mu} = \Lambda_\mu{}^\nu \frac{\partial}{\partial x^\nu} = \Lambda_\mu{}^\nu \partial_\nu \quad (7.29)$$

7.3.3 Contraction and trace

Two vectors A^μ, B^μ may be *contracted* to form their Minkowski inner product,

$$A \cdot B \equiv \eta_{\mu\nu} A^\mu B^\nu = A_\mu B^\mu = A^\mu B_\mu \quad (7.30)$$

The inner product is invariant under Lorentz transformations. More generally, two tensors $A^{\mu_1 \dots \mu_n}, B^{\mu_1 \dots \mu_n}$ of the same rank n may be contracted to form a scalar,

$$A \cdot B = A_{\mu_1 \dots \mu_n} B^{\mu_1 \dots \mu_n} \quad (7.31)$$

One may also contract two tensors A , and B , of ranks $m + p$ and $n + p$ respectively over p indices to yield a tensor of rank $m + n$,

$$A_{\mu_1 \dots \mu_m \rho_1 \dots \rho_p} B^{\nu_1 \dots \nu_n \rho_1 \dots \rho_p} = C_{\mu_1 \dots \mu_m}^{\nu_1 \dots \nu_n} \quad (7.32)$$

A particularly important contraction of this type consists in taking a trace by contracting a tensor A of rank $m + 2$ with the Minkowski metric tensor η (note that the pair of indices on A to be traced has to be specified) to yield a tensor of rank m ,

$$A_{\mu_1 \dots \mu_i \dots \mu_j \dots \mu_m} \eta^{\mu_i \mu_j} = B_{\mu_1 \dots \widehat{\mu_i} \dots \widehat{\mu_j} \dots \mu_m} \quad (7.33)$$

where the indices with carets are to be omitted. Later on, we shall describe further linear operations on tensors, namely symmetrization and anti-symmetrizations, with the help of which general tensors may be decomposed into their irreducible components.

7.4 Relativistic invariance of the wave equation

In the preceding sections, we have introduced scalars, vectors and tensors under Lorentz transformations. Next we shall be interested in *scalar fields*, *vector fields*, and *tensor fields*. Before proceeding to recasting Maxwell's equations in manifestly Lorentz-invariant form, we shall treat as a warm-up the case of the relativistic wave equation for the scalar field.

A function $\phi(x)$, where $x^\mu = (ct, \mathbf{x})$, is said to be a scalar field if it behaves as follows under a Poincaré transformation,

$$\phi(x) \rightarrow \phi'(x') = \phi(x) \quad x'^\mu = \Lambda^\mu_\nu x^\nu + a^\mu \quad (7.34)$$

In English, this means that the field ϕ' in the new frame with coordinates x' equals the old field ϕ in the old frame with coordinates x . An important Lorentz-invariant equation for the propagation of a scalar field is given by the wave equation,

$$\left(-\frac{1}{c^2} \frac{\partial^2}{\partial t^2} + \Delta \right) \phi(x) = \partial^\mu \partial_\mu \phi(x) = 0 \quad (7.35)$$

Note that there is no suitable Lorentz invariant 1-st order differential equation for a scalar field, since $\partial_\mu \phi = 0$ would imply that ϕ is constant. An important generalization of the wave equation is obtained by including a *mass term*,

$$\left(\hbar^2 \partial_\mu \partial^\mu - m^2 c^2 \right) \phi(x) = 0 \quad (7.36)$$

This equation for a massive scalar field may be solved by Fourier analysis, and we have

$$\phi_k(x) = e^{ik \cdot x} \quad \hbar^2 k_\mu k^\mu + m^2 c^2 = 0 \quad (7.37)$$

Since $p^\mu = \hbar k^\mu$, we see that this equation gives the energy - momentum relation for a relativistic particle (or wave) with mass m . A further Lorentz-invariant generalization of the scalar field wave equation is obtained by adding an arbitrary function $V(\phi)$,

$$\hbar^2 \partial_\mu \partial^\mu \phi - m^2 c^2 \phi - V(\phi) = 0 \quad (7.38)$$

The latter may be derived via the action principle from the invariant Lagrangian density,

$$\mathcal{L} = -\frac{1}{2} \hbar^2 \partial_\mu \phi \partial^\mu \phi - \frac{1}{2} m^2 c^2 \phi^2 - V(\phi) \quad (7.39)$$

The Higgs particle, for example, is an elementary scalar particle, which is described by a real scalar field ϕ , with a quartic potential $V(\phi)$. Of course, the Higgs particle couples to many other particles, such as electrons, quarks and neutrinos, and this description will require additional fields in the Lagrangian density.

7.5 Relativistic invariance of Maxwell equations

The prime example of a vector field under Lorentz transformations is provided by Maxwell theory of electro-magnetism. A general (covariant) vector field $V_\mu(x)$ is a collection of 4 fields $V_0(x), V_1(x), V_2(x), V_3(x)$ which behave as follows under a Lorentz transformation Λ ,

$$V_\mu(x) \rightarrow V'_\mu(x') = \Lambda_\mu{}^\nu V_\nu(x) \quad (7.40)$$

and analogously for a contra-variant vector field.

7.5.1 The gauge field, the electric current, and the field strength

In electro-magnetism, we encounter two vector fields, the gauge potential $A_\mu = (-\Phi/c, \mathbf{A})$, and the electric current density $j^\mu = (\rho c, \mathbf{j})$, where Φ is the electric potential, \mathbf{A} the usual 3-dimensional vector potential, ρ the charge density, and \mathbf{j} the electric current density. Consistency of Maxwell's equations requires the current density to be conserved,

$$\partial_\mu j^\mu = \frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0 \quad (7.41)$$

Maxwell's equations are invariant under gauge transformations on the vector potential,

$$A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu \theta \quad (7.42)$$

for any scalar function θ . The electric and magnetic fields are first order time and space derivatives of the electric potential and of the vector potential, and both are gauge invariant.

The most systematic way to construct the electric and magnetic field is to precisely take advantage of these two properties. Computing the general first order derivatives gives $\partial_\mu A_\nu$, which behaves as follows under a gauge transformation of A_μ ,

$$\partial_\mu A_\nu \rightarrow \partial_\mu A'_\nu = \partial_\mu A_\nu + \partial_\mu \partial_\nu \theta \quad (7.43)$$

Thus, the most general first order derivative of A_μ is not gauge invariant. The gauge term $\partial_\mu \partial_\nu \theta$ is, however, always symmetric under the interchange of μ and ν . Therefore, we are guaranteed that the *anti-symmetric* part of the derivative will be gauge invariant. The corresponding tensor is referred to as the *field strength* and is defined by,

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \quad (7.44)$$

Counting the number of independent fields in the rank 2 anti-symmetric tensor $F_{\mu\nu} = -F_{\nu\mu}$ gives $4 \times 3/2 = 6$, which is precisely the correct number to incorporate the 3 components of \mathbf{E} and the three components of \mathbf{B} . Decomposing $F_{\mu\nu}$ according to its space and time indices, and using anti-symmetry, we find the following parts,

$$\begin{aligned} F_{0i} &= \partial_0 A_i - \partial_i A_0 \\ F_{ij} &= \partial_i A_j - \partial_j A_i \quad i, j = 1, 2, 3 \end{aligned} \quad (7.45)$$

The standard electric and magnetic fields may now be identified as follows,

$$\begin{aligned} F_{i0} &= E_i/c \\ F_{ij} &= \sum_{k=1}^3 \varepsilon_{ijk} B_k \end{aligned} \quad (7.46)$$

As a matrix, the field strength tensor has the following entries,

$$F_{\mu\nu} = \begin{pmatrix} 0 & -E_1/c & -E_2/c & -E_3/c \\ E_1/c & 0 & B_3 & -B_2 \\ E_2/c & -B_3 & 0 & B_1 \\ E_3/c & B_2 & -B_1 & 0 \end{pmatrix}_{\mu\nu} \quad (7.47)$$

7.5.2 Maxwell's equations in Lorentz covariant form

Lorentz invariance of the equations dictates, to a large extent, the structure of the possible equations we may have for the gauge field A_μ , and its gauge invariant field strength tensor $F_{\mu\nu}$. Maxwell's equations emerge in two groups; a first set independent of the external electric current density j^μ , and a second set which does involve j^μ .

The first group of Maxwell's equations results directly from the fact that $F_{\mu\nu}$ is a “curl” in the 4-dimensional sense. From $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, it readily follows that

$$\varepsilon^{\mu\nu\rho\sigma} \partial_\rho F_{\mu\nu} = 2\varepsilon^{\mu\nu\rho\sigma} \partial_\rho \partial_\mu A_\nu = 0 \quad (7.48)$$

since the derivative $\partial_\rho \partial_\mu A_\nu$ is symmetric in ρ, μ , while the $\varepsilon^{\mu\nu\rho\sigma}$ is antisymmetric in all its indices, and may be normalized to $\varepsilon^{0123} = 1$ (note that because of the Minkowski signature of the metric, this implies that $\varepsilon_{0123} = -1$). Expressed in terms of \mathbf{E} and \mathbf{B} , this yields the first group of Maxwell's equations, by considering separately the cases where $\sigma = 0$ and $\sigma = i$ with $i = 1, 2, 3$,

$$\begin{aligned} \varepsilon^{\mu\nu\rho 0} \partial_\rho F_{\mu\nu} &= 0 & \Leftrightarrow & \quad \nabla \cdot \mathbf{B} = 0 \\ \varepsilon^{\mu\nu\rho i} \partial_\rho F_{\mu\nu} &= 0 & \Leftrightarrow & \quad \partial_t \mathbf{B} + \nabla \times \mathbf{E} = 0 \end{aligned} \quad (7.49)$$

It is instructive to give the derivation of these formulas. On the first line, the last index on ε is a time-index, namely 0, so that the other three indices on ε must be space-indices, which we shall rebaptize i, j, k . Thus, the first line becomes $\varepsilon^{ijk0} \partial_k F_{ij} = 0$. Using the definition of the magnetic field $F_{ij} = \sum_m \varepsilon_{ijm} B_m$, the fact that, with our conventions, $\varepsilon^{ijk0} = -\varepsilon^{ijk}$, and the double contraction

$$\varepsilon^{ijk} \varepsilon_{ijm} = 2\delta_m^k \quad (7.50)$$

we find $\varepsilon^{ijk0} \partial_k F_{ij} = 2\partial_k B^k = 2\nabla \cdot \mathbf{B}$. On the second line, the last index is i , which is space-like. Thus, one and only one of the indices μ, ν, ρ must be 0. Collecting all possibilities gives $\varepsilon^{jk0i} \partial_0 F_{jk} + 2\varepsilon^{0jki} \partial_k F_{0j} = 0$. Using the definitions of the electric and magnetic fields, and the above double contraction formula, we obtain $\partial_0 B^i + \varepsilon^{ijk} \partial_j E_k = 0$, which gives the formula quoted.

The second group of Maxwell equations are also linear and involve first derivatives of the electric and magnetic fields and are sourced by the current density. There is only one Lorentz-invariant combination with these properties, up to an overall constant factor, namely

$$\partial_\mu F^{\mu\nu} = -j^\nu \quad (7.51)$$

Note that, because $F_{\mu\nu}$ is antisymmetric in μ and ν , the current density must be conserved, $\partial_\nu \partial_\mu F^{\mu\nu} = \partial_\nu j^\nu = 0$. There are actually 4 equations encoded in (7.51), one for each of the possible values of ν .

$$\begin{aligned} \partial_\mu F^{\mu 0} &= -j^0 & \Leftrightarrow & \quad \nabla \cdot \mathbf{E} = \rho \\ \partial_\mu F^{\mu i} &= -j^i & \Leftrightarrow & \quad \frac{1}{c^2} \partial_t \mathbf{E} - \nabla \times \mathbf{B} = \mathbf{j} \end{aligned} \quad (7.52)$$

In summary, the relativistic form of the two groups of Maxwell's equations reduce to,

$$\begin{aligned}\varepsilon^{\mu\nu\rho\sigma}\partial_\rho F_{\mu\nu} &= 0 \\ \partial_\mu F^{\mu\nu} &= -j^\nu\end{aligned}\tag{7.53}$$

These equations may be derived from an action principle. Considering A_μ as the fundamental variable, and defining $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, then a suitable action is given by,

$$S[A] = \int d^4x \left(-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + A_\mu j^\mu \right)\tag{7.54}$$

A variation δA_μ of A_μ then yields the following variation of the action,

$$\begin{aligned}\delta S[A] &= \int d^4x \left(-\frac{1}{2} F^{\mu\nu} \delta F_{\mu\nu} + \delta A_\mu j^\mu \right) \\ &= \int d^4x (F^{\mu\nu} \partial_\nu \delta A_\mu + \delta A_\mu j^\mu)\end{aligned}\tag{7.55}$$

Integration by part of the first term gives (surface terms vanish for variations δA_μ with compact support, and will be omitted here),

$$\delta S[A] = \int d^4x \delta A_\mu (-\partial_\nu F^{\mu\nu} + j^\mu)\tag{7.56}$$

Its vanishing for all δA_μ requires the second group of Maxwell equations.

7.6 Relativistic kinematics

We want to produce an equation which replaces $\mathbf{F} = m\mathbf{a}$ by a relativistic invariant one, namely invariant under Poincaré transformations.

We begin by producing a relativistic generalization of velocity, for a massive particle. (The massless case needs separate treatment.) As usual, we denote the infinitesimal displacement in time and space by dx^μ , with respect to an inertial frame $\mathcal{R}(t, \mathbf{x})$. The differential dx^μ transforms as a 4-vector, just as x^μ did. Notice that the translation part of the Poincaré transformation cancels out. To obtain velocity, we should divide dx^μ by the differential of time. If we divide simply by dt , then the result will not be a 4-vector, since dt is a component of a vector. But there is another notion of time (for a massive particle), given by the *proper time* τ , and defined by

$$c^2 d\tau^2 = -\eta_{\mu\nu} dx^\mu dx^\nu\tag{7.57}$$

By construction, the proper time differential $d\tau$ is Lorentz invariant. The resulting velocity u^μ , defined by,

$$u^\mu \equiv \frac{dx^\mu}{d\tau} \qquad u_\mu u^\mu = -c^2 \qquad (7.58)$$

is now automatically a 4-vector. It is useful to spell out the components of u^μ . The velocity 3-vector with respect to the frame \mathcal{R} is given by the usual formula,

$$\mathbf{v} = \frac{d\mathbf{x}}{dt} \qquad (7.59)$$

In terms of \mathbf{v} , we obtain the relation of *time-dilation*, namely,

$$d\tau^2 = dt^2 \left(1 - \frac{\mathbf{v}^2}{c^2} \right) \qquad (7.60)$$

or simply $dt = \gamma d\tau$. Thus, the components of u^μ are,

$$u^\mu = (\gamma c, \gamma \mathbf{v}) \qquad \gamma = \left(1 - \frac{\mathbf{v}^2}{c^2} \right)^{-\frac{1}{2}} \qquad (7.61)$$

Relativistic momentum for a massive particle is defined from the velocity u^μ by,

$$p^\mu \equiv m u^\mu \qquad p_\mu p^\mu = -m^2 c^2 \qquad (7.62)$$

It is assumed in modern special relativity that the mass of a particle is a relativistic invariant, so that p^μ is again a 4-vector. The components of the momentum may be derived from the components of velocity, and we find,

$$\begin{aligned} p^\mu &= (p^0, \mathbf{p}) & p^0 &= \gamma m c \\ & & \mathbf{p} &= \gamma m \mathbf{v} \end{aligned} \qquad (7.63)$$

We see that relativistic momentum \mathbf{p} is given by the non-relativistic expression $m\mathbf{v}$ only up to higher order corrections in the velocity. In fact, the expansions of the momentum components up to 4-th order in velocity are as follows,

$$\begin{aligned} \mathbf{p} &= m\mathbf{v} \left(1 + \frac{1}{2} \frac{\mathbf{v}^2}{c^2} \right) + \mathcal{O}(v^5) \\ c p^0 &= m c^2 \left(1 + \frac{1}{2} \frac{\mathbf{v}^2}{c^2} + \frac{3}{8} \left(\frac{\mathbf{v}^2}{c^2} \right)^2 \right) + \mathcal{O}(v^5) \end{aligned} \qquad (7.64)$$

We recognize the second term in cp^0 as the non-relativistic kinetic energy. Thus, it is natural to identify cp^0 with the total “kinetic energy” E of the particle, and this energy includes Einstein’s famous contribution from the rest mass,

$$E = mc^2 \quad (7.65)$$

For non-zero velocity and momentum, the generalization is obtained directly from (??),

$$E^2 = m^2c^4 + \mathbf{p}^2c^2 \quad (7.66)$$

Thus, Lorentz covariance of the 4-vector p^μ forces us to introduce this rest mass energy, lest our definition of energy not be consistent with Poincaré invariance. Energy-momentum conservation is then simply the statement that the 4-vector p^μ , as a whole, is conserved.

Another way of establishing the same result is by considering a boost in the 1-direction, by a velocity v , for which we have,

$$\begin{aligned} E' &= \gamma(E - vp^1) \\ p'^1 &= \gamma(p^1 - vE/c^2) \\ p'^2 &= p^2 \\ p'^3 &= p^3 \end{aligned} \quad (7.67)$$

If this transformation is to reproduce the non-relativistic boost $p'^1 = p^1 - mv$ in the limit where $|v| \ll c$, and $\gamma \sim 1$, then we must have $E = mc^2$ in the limit of vanishing velocity.

7.7 Relativistic dynamics

Finally, we are ready to write down the relativistic generalization of $\mathbf{F} = m\mathbf{a}$. Again, we do so while being guided by Lorentz covariance,

$$\frac{dp^\mu}{d\tau} = f^\mu \quad (7.68)$$

The time derivative is again taken with respect to proper time to guarantee Lorentz covariance of the result. The relativistic force f^μ is represented by a 4-vector. Exactly what that force is will have to be determined by physical input beyond the kinematical requirements of Poincaré symmetry. The space component of f^μ is clearly the relativistic generalization of the usual force vector. The time-component cf^0 corresponds to the time derivative of energy, so it is the power absorbed by the particle.

An important example is provided by the Lorentz force, whose non-relativistic expression is given by the famous formula,

$$\mathbf{f} = e(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad (7.69)$$

Its relativistic generalization may be obtained directly from our construction of the electromagnetic field tensor in the previous sections, and we have,

$$\frac{dp^\mu}{d\tau} = f^\mu = eF^{\mu\nu} \frac{dx^\nu}{d\tau} \quad (7.70)$$

The overall sign of the force may be verified by using the correspondence of (7.46) and the non-relativistic version of (7.69). Re-expressing this equation in terms of derivatives with respect to the time t in frame \mathcal{R} ,

$$\frac{dp^\mu}{dt} = f^\mu = eF^{\mu\nu} \frac{dx^\nu}{dt} \quad (7.71)$$

Note that by contracting by $p^\mu = mu^\mu$ on both sides, and using the anti-symmetry of $F_{\mu\nu}$, we recover the time-independence of $p^\mu p_\mu$. The space-component of this equation coincides precisely with the form of the non-relativistic equation for the Lorentz force,

$$\frac{d\mathbf{p}}{dt} = e(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad (7.72)$$

but of course the relativistic meaning of \mathbf{p} in terms of velocity is different now. The time-component of (7.71) gives the power equations,

$$\frac{dE}{dt} = e\mathbf{E} \cdot \mathbf{v} \quad (7.73)$$

confirming that the magnetic field never contributes to energy change in the particle.

7.8 Lagrangian for a massive relativistic particle

We need a relativistic invariant Lagrangian that will reduce to the Newtonian form in the limit where $|\mathbf{v}| \ll c$. It is always best to look directly for the action. The free kinetic part in the action involves the particle mass $m > 0$, the velocity \mathbf{v} , the speed of light c , and the line element of integration dt , or equivalently, its Lorentz invariant cousin $d\tau$. On purely dimensional grounds, the action must take the form $mc^2 \int d\tau \psi(\mathbf{v}/c)$ for some function ψ .

But the function $\psi(\mathbf{v}/c)$ must be a Lorentz invariant scalar, which means it must be constant. The value of the constant is determined by matching the non-relativistic limit, and we find,

$$\begin{aligned} S_0[x^\mu] &= -mc^2 \int d\tau \\ &= -mc \int dt \sqrt{-\dot{x}^\mu \dot{x}_\mu} \\ &= -mc^2 \int dt \sqrt{1 - \mathbf{v}^2/c^2} \end{aligned} \tag{7.74}$$

where $\dot{x}^\mu = dx^\mu/dt$ and $\mathbf{v} = d\mathbf{x}/dt$. It is now straightforward to add to this non-interacting action S_0 the interaction with an external (or dynamical) electro-magnetic field, giving the following total action,

$$S[x^\mu] = \int dt \left(-mc \sqrt{-\dot{x}^\mu \dot{x}_\mu} + eA_\mu(x) \dot{x}^\mu \right) \tag{7.75}$$

It is worth pointing out that, if we take all 4 components of the variable x^μ to be dynamical, then the action could have been expressed with respect to *any integration parameter*,

$$S[x^\mu] = \int d\lambda \left(-mc \sqrt{-\frac{dx^\mu}{d\lambda} \frac{dx_\mu}{d\lambda}} + eA_\mu(x) \frac{dx^\mu}{d\lambda} \right) \tag{7.76}$$

This is remarkable: we have an action which is invariant under arbitrary reparametrizations of “time λ ”, so we are basically doing General Relativity in 1 time dimension ! This fact actually makes the Hamiltonian formulation rather tricky, and I will not discuss it here. Let me close by commenting that string theory is based on a straightforward generalization of this kind of action. Instead of a point-particle sweeping out a “world-line” during its time-evolution, a string will now sweep out a 2-dimensional surface, parametrized now by two parameters instead of the single λ .

7.9 Particle collider versus fixed target experiments

Suppose we have two possible experimental set-ups for the collisions of particle of mass m onto one another; (at LHC, we have protons onto protons/anti-protons)

- *Fixed target*: one incoming particle has kinetic energy E , the other is at rest.
- *Collider*: the particles have opposite momenta, and each has kinetic energy $E/2$.

The question is now in which experiment one gets the most “bang for the buck”. Translated into scientific terms, what this amounts to is the largest total energy in the rest frame.

For the Collider, the center of mass energy available is simply E , the sum of the energies of the two particles.

To compute the center of mass energy E_{cm} for the fixed target experiment, we use the fact that the total momentum p^μ has the same square in the fixed target frame as in the center of mass frame where, by definition, its momentum vanishes. This reads as follows,

$$p_\mu p^\mu = -(E/c + mc)^2 - p^2 c^2 = -(E_{cm}/c)^2 \quad (7.77)$$

Here, p is the momentum of the incoming particle, which is related to the energy by $E^2 = p^2 c^2 + m^2 c^4$. Eliminating p^2 between these two equations gives a formula for E_{cm} directly in terms of E and m ,

$$E_{cm} = \sqrt{2m^2 c^4 + 2mc^2 E} \quad (7.78)$$

We see that for large energy ($mc^2 \ll E$), the center of mass energy in the collider experiment grows like E , while in the fixed target experiment, it grows only like $\sqrt{mc^2 E}$. Thus, the collider provides much more bang for the buck !

7.10 A physical application of time dilation

There are all the funny applications of time dilation in relativity related to space-travel and the twin paradox. But there are also very important and directly observable applications to particle physics. Here, we shall provide one such example in terms of the observed life-time of unstable particles.

For definiteness, we consider the example of a muon μ^- particle; it is very similar to the electron, but it is 200 times heavier, namely $mc^2 \sim 100 MeV$, and unstable against weak interactions, with a life-time of approximately 2×10^{-6} sec. Muons are observed in cosmic ray showers. How far can the muons travel since they were first created ? Without taking time-dilation into account, they can travel at most a distance 2×10^{-6} sec $\times c = 600 m$, certainly not enough to travel intergalactically.

But if the energy of the muon is actually $E = mc^2 \gamma$ in the frame of observation, then they can travel a distance γ times 600 . For a muon of energy $E \sim 10^6 GeV$, this comes to approximately a distance of $10^7 km$, more in the range of intergalactic traveling distances.

8 Fluid Mechanics

Fluid mechanics deals with the motion of gasses and liquids on length scales which are much larger than the typical sizes of the particles and the typical distances between the particles of the fluid (sizes for simple atoms and molecules are on the order of $10^{-9}m$ while inter-distances are on the order of $10^{-8}m$ for liquids). Fluid mechanics (or dynamics) is an effective macroscopic theory, valid for a large number of particles (on the order of 10^{26} per mole), over macroscopic distances (on the order of millimeters, meters, or kilometers), and over macroscopic time scales (on the order of seconds, hours, or years). Short distance effects, and short time scale effects, such as they occur due to interactions between individual particles of the fluid, will be taken into account only to the degree that they affect the macroscopic behavior of the system. We shall concentrate here on non-relativistic fluids.

In fluid dynamics, positions of individual particles are averaged over with a weight given by their mass. The resulting average is the *mass density* function $\rho(t, \mathbf{x})$. The significance of this function may be appreciated by evaluating its integral over a spatial volume V , at a certain time t ,

$$M(t, V) = \int_V d^3x \rho(t, \mathbf{x}) \quad (8.1)$$

The quantity $M(t, V)$ is the total mass enclosed inside volume V at time t . The volume V is macroscopic in the sense that the number of particles enclosed by V is macroscopically large. Still, the volume V may be small on the physical length scales on which we are considering the system. One millimeter cube of water is a small quantity on the length scale of a bath tub, but it is still macroscopic as it will contain on the order of 10^{19} water molecules.

Similarly, the velocities of individual particles are averaged over macroscopic volumes to give a velocity field $\mathbf{v}(t, \mathbf{x})$. The product of the mass density and velocity field gives the momentum density,

$$\rho(t, \mathbf{x})\mathbf{v}(t, \mathbf{x}) \quad (8.2)$$

of the fluid at time t and position \mathbf{x} . This quantity is also referred to as the *mass flux density*. Another important quantity is the *entropy density per unit mass* $s(t, \mathbf{x})$, and its associated *entropy flux density*, defined by,

$$\rho(t, \mathbf{x})s(t, \mathbf{x})\mathbf{v}(t, \mathbf{x}) \quad (8.3)$$

There are other physical quantities defined in an analogous manner, such as the local *pressure* $p(t, \mathbf{x})$, the local *temperature* $T(t, \mathbf{x})$, and so on.

8.1 Conservation of mass

In classical non-relativistic mechanics, and in non-relativistic fluid dynamics, mass is conserved. Thus, in a fixed volume V , the change of the total enclosed mass $M(t, V)$ as a function of time corresponds to the net mass entering the system. This balance may be expressed as a mass conservation law at fixed volume V ,

$$\frac{\partial M(t, V)}{\partial t} + \int_{\partial V} d^2\mathbf{s} \cdot \rho(t, \mathbf{x}) \mathbf{v}(t, \mathbf{x}) = 0 \quad (8.4)$$

Here, $d^2\mathbf{s}$ is the outward pointing surface element vector on the boundary surface ∂V of the spatial volume V . Note that the length of the vector $d^2\mathbf{s}$ is the surface element on ∂V , while its direction is along the outward normal to the surface at the point \mathbf{x} . Thus, a surface element $d^2\mathbf{s}$ for which $d^2\mathbf{s} \cdot \mathbf{v} > 0$, the fluid is exiting the volume V . Using Stokes's theorem to transform the surface integral into a volume integral, we have,

$$\oint_{\partial V} d^2\mathbf{s} \cdot \rho \mathbf{v} = \int_V d^3x \nabla \cdot (\rho \mathbf{v}) \quad (8.5)$$

and thus,

$$\int_V d^3x \left(\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) \right) = 0 \quad (8.6)$$

Now this equation must hold for any spatial volume element V . In particular, V can be taken to be infinitesimally small (but still of macroscopic size so that it contains a very large number of fluid particles), so that the equation translates into a local equation,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \quad (8.7)$$

This is the equation of conservation of mass, or *continuity equation*.

8.2 Conservation of momentum and Newton's third law

Next we apply Newton's third law $\dot{\mathbf{p}} = \mathbf{F}$ to a small volume V of fluid. The total momentum $\mathbf{P}(t, V)$ in this volume is given by,

$$\mathbf{P}(t, V) = \int_V d^3x \rho(t, \mathbf{x}) \mathbf{v}(t, \mathbf{x}) \quad (8.8)$$

while the total force $\mathbf{F}(t, V)$ is given by,

$$\mathbf{F}(t, V) = \int_V d^3x \mathbf{f}(t, \mathbf{x}) \quad (8.9)$$

where $\mathbf{f}(t, \mathbf{x})$ is the local force acting at time t and position \mathbf{x} .

The change in time of the total momentum $\mathbf{P}(t, V)$, calculated while keeping the volume V fixed, received two different types of contributions. The first is due to the action of the local force \mathbf{f} on the particles in V . But there is a second contribution which arises from the fact that in a time span from t to $t + dt$, some particles leave the fixed volume V . In the application of Newton's third law for this small volume V we collect these two contributions,

$$\frac{\partial}{\partial t} \mathbf{P}(t, V) + \oint_{\partial V} (d^2 \mathbf{s} \cdot \mathbf{v})(\rho \mathbf{v}) = \mathbf{F}(t, V) \quad (8.10)$$

Next, we wish to replace the surface integral by a volume integral over V , so that we can again extract a local equation. This is a bit more tricky than in the case of mass flow because of the tensorial structure of the equation. Stokes's theorem in this case tells us that we have,

$$\oint_{\partial V} (d^2 \mathbf{s} \cdot \mathbf{v})(\rho \mathbf{v}) = \int_V d^3 x (\nabla \cdot \mathbf{v})(\rho \mathbf{v}) \quad (8.11)$$

The meaning of these symbols, and their ordering, is as follows. The gradient ∇ inside $(\nabla \cdot v)$ acts on both the velocity field \mathbf{v} in the parenthesis and on whatever vector function follows. To see more explicitly how this works, it is convenient to express the equation in *components*, where the velocity components are $\mathbf{v} = (v_1, v_2, v_3)$, the components of the infinitesimal area elements are $d^2 \mathbf{s} = (d^2 s_1, d^2 s_2, d^2 s_3)$, and the components of the gradient are given by $\nabla = (\partial_1, \partial_2, \partial_3)$, where $\partial_i = \partial / \partial x_i$, and we get,

$$\oint_{\partial V} \left(\sum_{j=1}^3 d^2 s_j v_j \right) (\rho v_i) = \int_V d^3 x \sum_{j=1}^3 \partial_j (\rho v_i v_j) \quad (8.12)$$

Expressing now also \mathbf{P} and \mathbf{F} in terms of volume integrals, we obtain,

$$\int_V d^3 x \frac{\partial(\rho \mathbf{v})}{\partial t} + \int_V d^3 x (\nabla \cdot \mathbf{v})(\rho \mathbf{v}) = \int_V d^3 x \mathbf{f} \quad (8.13)$$

Since this relation holds for arbitrary volumes V , it is equivalent to the local equation,

$$\frac{\partial(\rho \mathbf{v})}{\partial t} + (\nabla \cdot \mathbf{v})(\rho \mathbf{v}) = \mathbf{f} \quad (8.14)$$

or equivalently in components,

$$\frac{\partial(\rho v_i)}{\partial t} + \sum_{j=1}^3 \partial_j (\rho v_i v_j) = f_i \quad (8.15)$$

where we have also expressed the local force vector in components $\mathbf{f} = (f_1, f_2, f_3)$. This is sometimes referred to as the *Cauchy momentum equation*. Note that whenever one wants to be sure of one's tensor calculations, it is best to use the component form, which tends to be safer and more explicit than the vector form.

8.3 Euler's equations

The derivatives may be partially worked out with the help of Leibnitz's rule, and we find,

$$v_i \left(\frac{\partial \rho}{\partial t} + \sum_{j=1}^3 \partial_j (\rho v_j) \right) + \rho \left(\frac{\partial v_i}{\partial t} + \sum_{j=1}^3 (v_j \partial_j) v_i \right) = f_i \quad (8.16)$$

The first term on the left side is proportional to the mass conservation equation, and vanishes. As a result, we are left with the simplified *Euler equations*, recast here in vector notation,

$$\rho \left(\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} \right) = \mathbf{f} \quad (8.17)$$

To make progress towards understanding these equations, we need to obtain a more detailed description of the possible local forces acting on the fluid. A list of simplifying assumptions on \mathbf{f} tends to be helpful.

8.4 Ideal fluids

The motion of a physical fluid is generally subject to processes in which energy dissipates as a result of internal friction forces and the exchange of heat between different parts of the fluid. Internal friction forces produce *viscosity*. Ideal fluids are defined by the assumption that the effects of thermal conductivity and viscosity can be neglected. The inclusion of viscosity effects will be postponed until the subsequent section.

In an ideal fluid one contribution to the force $\mathbf{F}(t, V)$ applied to an arbitrary small volume V at time t is the result of local pressure $p(t, \mathbf{x})$ integrated over the boundary ∂V ,

$$\mathbf{F}(t, V) = - \oint_{\partial V} d^2 \mathbf{s} p(t, \mathbf{x}) \quad (8.18)$$

where $d^2 \mathbf{s}$ is again the outward pointing surface element. By Stokes's theorem, the surface integral may be transformed into a volume integral, whose form will be convenient for the derivation of Euler's equation,

$$\oint_{\partial V} d^2 \mathbf{s} p(t, \mathbf{x}) = \int_V d^3 x \nabla p(t, \mathbf{x}) \quad (8.19)$$

The minus sign arises from the convention that the forces exerted on the particles of the system inside volume V is opposite to the gradient of the pressure. Therefore, in an ideal fluid, the contribution of the pressure to the local force is given by $\mathbf{f} = -\nabla p$.

Another contribution to the force $\mathbf{F}(t, V)$ arises from external forces on the fluid, such as the force of gravity. These forces act on each particle separately, and are bulk effects. We shall denote the corresponding local force schematically by $\rho \mathbf{g}$, keeping in mind that \mathbf{g} is the gravitational acceleration to which the fluid is subject. Combining these two contributions, in the case of an ideal fluid, the Euler equations simplify as follows,

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = \mathbf{g} - \frac{1}{\rho} \nabla p \quad (8.20)$$

Other forces, such as electric and magnetic may be included as well.

8.4.1 Isentropic flows

Further restrictions may sometimes be imposed to approximate the dynamics of the fluid. A first is adiabatic behavior of the fluid, namely when the entropy per unit mass $s(t, \mathbf{x})$ remains constant during the flow. This condition reads,

$$\frac{ds}{dt} = \frac{\partial s}{\partial t} + \mathbf{v} \cdot \nabla s = 0 \quad (8.21)$$

Equivalently, with the help of the mass conservation equation, one may write a continuity equation for the entropy flux density $\rho s \mathbf{v}$,

$$\frac{\partial(\rho s)}{\partial t} + \nabla(\rho s \mathbf{v}) = 0 \quad (8.22)$$

An even further restriction is when s is actually constant in space and time, namely when the flow is *isentropic*. One may then use the equations of thermodynamics to obtain relations on the small macroscopic equilibrium system of volume V . The thermodynamic function appropriate to the independent thermodynamic variables of entropy and pressure is the *enthalpy*. Considering the quantities per unit mass, we have the relation,

$$dw = Tds + Vdp \quad (8.23)$$

where w is the enthalpy per unit mass, and V is the volume per unit mass, namely the inverse density, $1/\rho$. As we keep s constant, the equation reduces to $dw = dp/\rho$, or in the present context,

$$\nabla w = \frac{1}{\rho} \nabla p \quad (8.24)$$

With this extra assumption, the Euler equation reduces to,

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = \mathbf{g} - \nabla w \quad (8.25)$$

Upon taking a curl of both side, the dependence on w cancels, and we obtain an equation involving only the velocity.

8.4.2 Bernoulli equation

Steady state flow occurs when $\partial \mathbf{v} / \partial t = 0$. The Euler equation simplifies and, with the help of the following identity,

$$\frac{1}{2} \nabla \mathbf{v}^2 = (\mathbf{v} \cdot \nabla) \mathbf{v} + \mathbf{v} \times (\nabla \times \mathbf{v}) \quad (8.26)$$

may be re-expressed as follows,

$$\frac{1}{2} \nabla \mathbf{v}^2 - \mathbf{v} \times (\nabla \times \mathbf{v}) = \mathbf{g} - \nabla w \quad (8.27)$$

Steady state flows form an important class of special cases. In particular, in steady state the fluid particles follow trajectories which are fixed in time. These trajectories are often referred to as *streamlines*, and are defined as curves which are everywhere parallel to the velocity field,

$$\frac{d\mathbf{x}}{dt} = \lambda(\mathbf{x}) \mathbf{v}(\mathbf{x}) \quad (8.28)$$

where $d\mathbf{x}/dt$ is being evaluated at the point \mathbf{x} , and for some non-zero function λ . Along a streamline, we may derive a conserved quantity, by taking the inner product of (8.27) with $d\mathbf{x}/dt$ and using (8.28). We shall assume that the external force per unit mass \mathbf{g} derives from a potential Φ by $\mathbf{g} = -\nabla \Phi$, as is the case for the non-relativistic gravitational potential. The second term on the right in (8.27) cancels out since it is orthogonal to \mathbf{v} , and the remaining two terms combine into total time-derivatives along the streamline,

$$\frac{d\mathbf{x}}{dt} \cdot \nabla \left(\frac{1}{2} \mathbf{v}^2 + w + \Phi \right) = \frac{d}{dt} \left(\frac{1}{2} \mathbf{v}^2 + w + \Phi \right) = 0 \quad (8.29)$$

Hence along each streamline, the following combination is constant,

$$\frac{1}{2} \mathbf{v}^2 + w + \Phi = \text{constant} \quad (8.30)$$

where the value of the constant depends on the particular streamline under consideration. This result is *Bernoulli's equation*.

8.4.3 Conservation of circulation

The *velocity circulation* of the velocity \mathbf{v} along a contour \mathcal{C} is defined by,

$$\oint_{\mathcal{C}} \delta \mathbf{x} \cdot \mathbf{v} \quad (8.31)$$

where the contour \mathcal{C} is a fluid contour following the particles of the fluid. The symbol $\delta \mathbf{x}$ is used here for the infinitesimal in order to distinguish it from the derivative with respect to time, which will also be used. The time-derivative of the circulation will receive contributions from the time-dependence of the velocity \mathbf{v} as well as from the time-dependence of the contour \mathcal{C} , and may be evaluated (still for isentropic flows) as follows,

$$\frac{d}{dt} \oint_{\mathcal{C}} \delta \mathbf{x} \cdot \mathbf{v} = \oint_{\mathcal{C}} \delta \mathbf{x} \cdot \frac{d\mathbf{v}}{dt} + \oint_{\mathcal{C}} \frac{d\delta \mathbf{x}}{dt} \cdot \mathbf{v} \quad (8.32)$$

For isentropic flows, we have,

$$\frac{d\mathbf{v}}{dt} = \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\nabla(w + \Phi) \quad (8.33)$$

Since the integral of a gradient along any closed contour vanishes, the first term on the right side of (8.32) actually vanishes, and we are left with,

$$\frac{d}{dt} \oint_{\mathcal{C}} \delta \mathbf{x} \cdot \mathbf{v} = \oint_{\mathcal{C}} \frac{d\delta \mathbf{x}}{dt} \cdot \mathbf{v} \quad (8.34)$$

Now along the fluid contour \mathcal{C} , the time derivative of \mathbf{x} is just the velocity, so that we have,

$$\frac{d\delta \mathbf{x}}{dt} \cdot \mathbf{v} = \delta \frac{d\mathbf{x}}{dt} \cdot \mathbf{v} = (\delta \mathbf{v}) \cdot \mathbf{v} = \delta \left(\frac{1}{2} \mathbf{v}^2 \right) \quad (8.35)$$

But this object is an exact differential, whose integral along a closed contour again vanishes. We conclude that the velocity circulation along a closed fluid contour vanishes

$$\frac{d}{dt} \oint_{\mathcal{C}} \delta \mathbf{x} \cdot \mathbf{v} = 0 \quad (8.36)$$

also referred to as *Kelvin's theorem*. Note that, to prove this result, we have explicitly used the fact that the fluid is isentropic, so that the local forces per unit mass is a gradient of the enthalpy. The result does not hold for flows which are not isentropic.

8.4.4 Potential flow

Potential flow is defined to be a flow for which the velocity field is a gradient, everywhere in space and at all times,

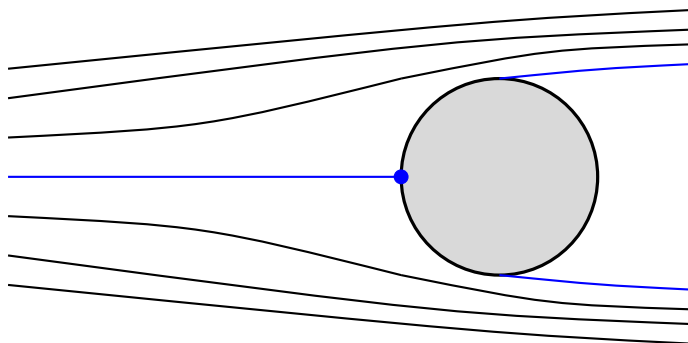
$$\mathbf{v} = \nabla\phi \quad (8.37)$$

and where ϕ is a well-defined single-valued function, referred to as the velocity potential. Clearly then, the velocity circulation around any closed path vanishes,

$$\oint_C \delta\mathbf{x} \cdot \mathbf{v} = \oint_C \delta\mathbf{x} \cdot \nabla\phi = 0 \quad (8.38)$$

One may ask whether a flow for which the velocity circulation vanishes at all points in space must be a potential flow, and the answer is yes, as long as the flow is everywhere regular.

A more subtle dynamical question is whether a flow which starts out uniform in the far past, and is thus circulation-free, will remain so at all later times, and here the answer is not always affirmative. One source of obstruction is the emergence of singularities in a flow for which viscosity has been ignored. This can happen for a flow past a fixed body, as explained in Landau and Lifshitz with the help of the figure below.



The streamline indicated in blue is singular, as by symmetry half of the fluid must flow over the disc and half of it under the disc. This makes the dot indicated in blue a singular point of the flow. Such a singular point is actually not physical. Singularities produce large gradients (in principle infinite), but with large gradients the fluid will now be far from ideal, and physical effects due to friction, and thus viscosity, need to be taken into account. These effects will smoothen the singularities exhibited by the flow of an ideal fluid, with the effect that velocity circulation may be generated locally.

Euler's equations simplify considerably for potential flow. Starting from their form in (8.25), and using the formula (8.26), we find,

$$\frac{\partial \mathbf{v}}{\partial t} + \frac{1}{2} \nabla v^2 - \mathbf{v} \times (\nabla \times \mathbf{v}) = -\nabla w \quad (8.39)$$

where we have absorbed the contribution from \mathbf{g} into w . Substituting $\mathbf{v} = \nabla \phi$, the curl term cancels out, and the remaining equation can be recast as follows,

$$\nabla \left(\frac{\partial \phi}{\partial t} + \frac{1}{2} (\nabla \phi)^2 + w \right) = 0 \quad (8.40)$$

This equation may be integrated, and results in,

$$\frac{\partial \phi}{\partial t} + \frac{1}{2} (\nabla \phi)^2 + w = f(t) \quad (8.41)$$

for some function f which depends on t only. Actually, the function f may be absorbed into the potential ϕ , since ϕ was defined only up to a function of t only.

8.4.5 Incompressible fluids

In many practical applications, the density of a fluid may be well-approximated by a density ρ which is constant in space and time. The flow is then referred to as *incompressible*, and this approximation leads to considerable simplifications. In particular, the flow is then always isentropic, and divergence-free,

$$w = \frac{p}{\rho} \quad \nabla \cdot \mathbf{v} = 0 \quad (8.42)$$

Potential flow in an incompressible fluid is particularly simple, since it is governed by a linear equation for the velocity potential ϕ . Indeed, the velocity \mathbf{v} obeys $\nabla \cdot \mathbf{v} = 0$ as well as $\mathbf{v} = \nabla \phi$, so that we must have,

$$\Delta \phi = 0 \quad \frac{\partial \phi}{\partial t} + \frac{1}{2} (\nabla \phi)^2 + \frac{p}{\rho} = 0 \quad (8.43)$$

8.4.6 Some examples of problems on incompressible flows

It will probably be useful to present a few examples. To keep things simple and practical, we concentrate on ideal fluids which are incompressible.

(a) *Flow in a rotating cylinder in a gravitational field*

We determine the shape of the surface of an incompressible ideal fluid subject to a uniform gravitational field contained in a cylinder rotating about the vertical axis with constant angular velocity ω . Taking the z -axis to be vertical along the gravitational force, the components of the velocity field are given by,

$$v_x = +\omega y \quad v_y = -\omega x \quad (8.44)$$

The vector field is automatically divergence-free, but it is not a potential flow. The Euler equations become,

$$\omega^2 x = \frac{1}{\rho} \frac{\partial p}{\partial x} \quad \omega^2 y = \frac{1}{\rho} \frac{\partial p}{\partial y} \quad 0 = g + \frac{1}{\rho} \frac{\partial p}{\partial z} \quad (8.45)$$

The equation may be integrated for the pressure as a function of x, y, z , and we find,

$$p(x, y, z) = \rho \left(-gz + \frac{1}{2} \omega^2 (x^2 + y^2) \right) + p_0 \quad (8.46)$$

where p_0 is an arbitrary integration constant. The pressure of the fluid must compensate the constant pressure of the air in the cylinder at the interface with the air, and so the shape of the surface is a paraboloid given by,

$$p_{\text{air}} - p_0 = \rho \left(-gz + \frac{1}{2} \omega^2 (x^2 + y^2) \right) \quad (8.47)$$

(b) Motion of a sphere through an incompressible fluid

A sphere of radius R moves with velocity \mathbf{u} through an incompressible fluid. The velocity \mathbf{u} may depend on time. We determine the potential flow of the fluid past the sphere. The vector field $\mathbf{v} = \nabla \phi$ derives from a flow potential ϕ , which in turn satisfies $\Delta \phi = 0$. This problem has rotational symmetry around \mathbf{u} , so that ϕ itself is rotation invariant around \mathbf{u} . Also, ϕ must vanish at ∞ . Given the linearity of the equation for ϕ and its boundary conditions, ϕ must depend linearly on \mathbf{u} , so that it must be a dipole solution,

$$\phi = c \mathbf{u} \cdot \nabla \frac{1}{r} = -c \frac{\mathbf{u} \cdot \mathbf{r}}{r^3} \quad (8.48)$$

where r is the distance from the center of the sphere, and c is a constant to be determined. Since the fluid is in contact with the sphere, the components normal to the sphere of the fluid velocity and the sphere must be equal. To obtain this relation, we compute \mathbf{v} from ϕ ,

$$\mathbf{v} = \frac{c}{r^5} \left(3(\mathbf{u} \cdot \mathbf{r})\mathbf{r} - r^2 \mathbf{u} \right) \quad (8.49)$$

Requiring equality of $\mathbf{u} \cdot \mathbf{r}$ and $\mathbf{v} \cdot \mathbf{r}$ on the sphere where $r = R$ gives, $c = \frac{1}{2}R^3$. The associated pressure p is as follows,

$$p = p_0 - \frac{1}{2}\rho\mathbf{v}^2 - \rho\partial\phi/\partial t \quad (8.50)$$

The time derivative is calculated by noting that the center of the sphere moves with velocity \mathbf{u} , so that

$$\frac{\partial\phi}{\partial t} = -\mathbf{u} \cdot \nabla\phi + \dot{\mathbf{u}} \cdot \nabla_{\mathbf{u}}\phi \quad (8.51)$$

One may for example evaluate the pressure on the surface of the sphere as a function of the angle θ between \mathbf{u} and \mathbf{r} ,

$$p = p_0 + \frac{\rho u^2}{8}(9\cos^2\theta - 5) + \frac{1}{2}\rho\mathbf{r} \cdot \dot{\mathbf{u}} \quad (8.52)$$

obtained by substituting the above expressions for ϕ and \mathbf{v} .

8.5 Viscous fluids

The mass continuity equation of (8.7) is easily seen to be invariant under time-reversal, governed by the transformations,

$$T : (t, \mathbf{x}) \rightarrow (-t, \mathbf{x}) \quad (\rho, \mathbf{v}) \rightarrow (\rho, -\mathbf{v}) \quad (8.53)$$

Under this transformation the Euler equations of (8.17) are invariant provided the force field \mathbf{f} is invariant under time reversal. This will be the case, for example, when \mathbf{f} is independent of time t and independent of velocity \mathbf{v} (or more generally even in \mathbf{v}). Ideal fluids give an important special case. Whenever the system is invariant under time-reversal, its dynamics is reversible.

8.5.1 The momentum flux density tensor

Friction forces are proportional to velocity, and thus odd under time-reversal. As a result, the equations of fluid dynamics in the presence of friction cannot be time-reversal invariant. Instead the flow must be irreversible. The irreversibility is a reflection of the physical fact that the motion of the fluid produces heat and this heat is dissipated. To incorporate the effects of friction and thus of viscosity, we need to modify Euler's equations. The easiest way

to do this is through the use of the *momentum flux density tensor* Π_{ij} (or simple the *stress tensor*. For any fluid, the dynamical equations may be expressed as follows,

$$\partial_t(\rho v_i) + \sum_{j=1}^3 \partial_j \Pi_{ij} = 0 \quad (8.54)$$

For an ideal fluid, Π_{ij} takes the form,

$$\Pi_{ij} = \rho v_i v_j + p \delta_{ij} \quad (8.55)$$

and we recover Euler's equations. More generally, we may always express the force \mathbf{f} with components f_i in terms of the stress tensor by the following equations,

$$f_i = - \sum_{j=1}^3 \partial_j (\Pi_{ij} - \rho v_i v_j) \quad (8.56)$$

Note that we are not requiring f_i to be the gradient of a scalar (which is not always the case), but instead we are requiring f_i to be the divergence of a symmetric rank-two tensor (which is always true, as the integrability conditions are trivial). Thus, our starting point for the study of viscous fluids will be the momentum conservation equation (8.54), and it remains to find a suitable correction to the ideal fluid stress tensor to incorporate the effects of friction and viscosity. The effect is customarily parametrized as follows,

$$\Pi_{ij} = \rho v_i v_j + p \delta_{ij} - \sigma'_{ij} \quad (8.57)$$

where σ'_{ij} gives the viscous irreversible contributions to the stress tensor.

To determine σ'_{ij} , we use symmetry arguments. Assuming the fluid is *isotropic*, namely that the equations which govern it are invariant under rotations, then σ'_{ij} must be a tensor under space rotations. Next, we know from physical principles that friction and viscosity arise only when fluid parts of different velocities move adjacent to one another, but that there is no friction when velocity is uniform. Hence, σ'_{ij} must involve the gradient of the velocity $\nabla_i v_j$. We will assume that this gradient is small, so that only first order terms in the gradient must be retained. Finally, since the stress tensor is symmetric under interchange of i and j , it must be of the following form,

$$\sigma'_{ij} = \eta \left(\partial_i v_j + \partial_j v_i - \frac{2}{3} \delta_{ij} \sum_{k=1}^3 \partial_k v_k \right) + \zeta \delta_{ij} \sum_{k=1}^3 \partial_k v_k \quad (8.58)$$

The tensor multiplying the *coefficient of viscosity* η has been arranged so as to be *traceless*, namely $\sum_{i=1}^3 \sigma'_{ii} = 0$. the remaining term proportional to the *second viscosity* is proportional

to the identity tensor δ_{ij} . Note however that this term is not simply a contribution to the pressure, since it is odd in velocities and thus odd under time-reversal, while the pressure is even. The viscosity coefficients η, ζ will in general depend upon the pressure and the temperature, and are therefore generally dependent on space \mathbf{x} and possibly even on time t .

8.5.2 The Navier-Stokes equation

Under the simplifying assumption that η and ζ are actually constant throughout the fluid, the equations of momentum conservation simplify considerably, and reduce to,

$$\rho \left(\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} \right) = -\nabla p + \eta \Delta \mathbf{v} + \left(\zeta + \frac{1}{3} \eta \right) \nabla (\nabla \cdot \mathbf{v}) \quad (8.59)$$

This equation is usually referred to as the *Navier-Stokes equation*. The symbol Δ refers to the Laplace operator given by $\Delta = (\nabla \cdot \nabla)$.

For an incompressible fluid we have $\nabla \cdot \mathbf{v} = 0$, and the equation simplifies further, since the last term above then cancels,

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\nabla \left(\frac{p}{\rho} \right) + \frac{\eta}{\rho} \Delta \mathbf{v} \quad (8.60)$$

Landau and Lifshitz give some experimental values for $\nu = \eta/\rho$ for various common substances at 68 degrees Fahrenheit, in units of cm^2/s , namely Water 0.01; Air 0.15; Alcohol 0.02; Glycerine 6.8; Mercury 0.001.

8.5.3 Viscous flow in a long pipe

Consider the stationary flow of an incompressible viscous fluid inside a cylinder of length L and radius R with $R \ll L$, and pressure difference ΔP between both ends of the cylinder. Compute the profile of the pressure and the velocity, and evaluate the net flow of mass per unit time across the cylinder.

Take the z -axis to be the axis of symmetry of the cylinder, and x, y perpendicular to the z -axis. We introduce polar coordinates (r, ϕ) with $r^2 = x^2 + y^2$. Since the fluid is incompressible, ρ is constant and the mass continuity equation implies that the velocity field satisfies $\nabla \cdot \mathbf{v} = 0$. Restricting to flows for which the velocity is at all points in the z -direction $\mathbf{v} = (0, 0, v_z)$, the mass continuity equation then implies that v_z is independent of z . Rotational symmetry of the cylinder about the z -axis requires v_z to be independent of ϕ , so that we have $v_z(r)$. The Navier-Stokes equation greatly simplifies. First of all $\partial_t \mathbf{v} = 0$ for

a stationary flow, while we also have $(\mathbf{v} \cdot \nabla)\mathbf{v} = (0, 0, v_z \partial_z v_z) = 0$. Hence the Navier-Stokes equation reduces to,

$$\eta \Delta \mathbf{v} = \nabla p \quad (8.61)$$

Since $v_x = v_y = 0$, we must have $\partial_x p = \partial_y p = 0$, so that the pressure field only depends on z . The component along z of the Navier-Stokes equation then becomes,

$$\eta \Delta v_z = \partial_z p \quad (8.62)$$

Since v_z is independent of z , it follows that $\partial_z p$ must be constant, and since we know the total pressure difference P across the cylinder of length L , we have,

$$\partial_z p = -\frac{P}{L} \quad p(z) = p_0 - \frac{\Delta P}{L} z \quad (8.63)$$

Since v_z depends only on r , the Laplacian acting on v_z is conveniently expressed in polar coordinates, and we have,

$$\eta \frac{1}{r} \partial_r (r \partial_r v_z) = -\frac{\Delta P}{L} \quad (8.64)$$

integrating once gives,

$$\eta r \partial_r v_z = -\frac{P}{2L} r^2 + \alpha \quad (8.65)$$

where α is an integration constant. Integrating once more, we find,

$$v_z(r) = -\frac{P}{4\eta L} r^2 + \alpha \ln r + v_0 \quad (8.66)$$

Now the $\ln r$ term must be absent since its presence would imply that the velocity field diverges at the center of the cylinder, and thus we have $\alpha = 0$. Furthermore, when $r = R$, the velocity must vanish at the cylinder, which determines v_0 . Putting all together, we have,

$$v_z(r) = \frac{P}{4\eta L} (R^2 - r^2) \quad (8.67)$$

The total mass flow per unit time Q is given by the integral,

$$Q = 2\pi \int_0^R r dr \rho v_z(r) \quad (8.68)$$

where the factor of 2π arises from the angular integration over ϕ . The integral evaluates to,

$$Q = \frac{\pi P}{8\eta L} R^4 \quad (8.69)$$

Applying this calculation to the problem of evaluating blood flow in arteries for fixed arterial pressure, we see that a mere 10% decrease in the diameter of the blood vessel brings about a 40% reduction in the blood flow !

9 Completely integrable systems

From experience, we know that certain simple mechanical systems have a large enough number of conserved charges, or *first integrals*, so that the mechanical degrees of freedom may be separated and the system may be completely solved for, or *completely integrated*. This raises the question as to whether there is a general criterion for determining when a system is completely integrable.

9.1 Criteria for integrability, and action-angle variables

Liouville's Theorem provides a powerful result on this problem. Consider a system with N degrees of freedom q_i, p_i , for $i = 1, \dots, N$, with Hamiltonian H , and obeying canonical Poisson brackets. If the system possesses precisely N functionally independent first integrals $I_i(q, p)$ for $i = 1, \dots, N$, all of which mutually commute, namely,

$$\{I_i, I_j\} = 0 \quad \text{for all} \quad i, j = 1, \dots, N \quad (9.1)$$

then the system is completely integrable (in the sense of Liouville). As a result, there exist action variables $J_i(I_1, \dots, I_N)$, which are functions only of I_i , and conjugate angle variables ϕ_i , such that

$$\begin{aligned} \{J_i, \phi_j\} &= \delta_{ij} \\ \{J_i, J_j\} &= 0 \\ \{\phi_i, \phi_j\} &= 0 \end{aligned} \quad (9.2)$$

If the Hamiltonian is conservative, H itself will be a function of the first integrals only.

A few comments are in order. First, *functional independence* of the first integrals may be cast most conveniently in the differential form notation,

$$dI_1 \wedge dI_2 \wedge \dots \wedge dI_N \neq 0 \quad (9.3)$$

and is easy to check. Secondly, equation $\{J_i, J_j\} = 0$ follows from (9.1), while the first equation in (9.2) provides a normalization for the angle ϕ_i .

9.2 Standard examples of completely integrable systems

First, we provide a (incomplete) list of the standard completely integrable systems, and postpone the more sophisticated Toda lattice, Sutherland-Calogero-Moser system, Korteweg de Vries equation, and their Lax pair treatment, until the subsequent section. Standard integrable systems are as follows,

1. The free particle, and the harmonic oscillator in any dimension, since their Euler-Lagrange equations are linear.
2. The free spinning top in any dimension (as we shall show later).
3. Spherically symmetric Lagrangians with standard kinetic energy, in any dimension

$$L = \frac{1}{2}m\dot{\mathbf{r}}^2 - V(r) \qquad \mathbf{r} = (x_1, \dots, x_N), \quad r = |\mathbf{r}| \qquad (9.4)$$

in any dimension. Angular momentum $L_{ij} = x_i p_j - x_j p_i$ is conserved in any dimension, and thus is L^2 . Therefore the radial and angular motions decouple. Radial motion can be solved by energy conservation. Angular motion can be solved because the free spinning top in any dimension is completely integrable.

4. The spinning top in the presence of a uniform gravitational field, or a spinning magnetic moment in a uniform magnetic field.

9.3 More sophisticated integrable systems

Next, there are also more sophisticated integrable systems, the integrability of many of which has been shown only over the past few decades.

1. Generalizations of the spinning top to any coset space G/H where G and H are Lie groups, and H is a subgroup of G .
2. The Toda Lattice, with exponential nearest-neighbor interactions between particles on a line, is characterized by the following Hamiltonian,

$$H = \sum_i \left(\frac{1}{2} p_i^2 + \omega^2 e^{q_i - q_{i-1}} \right) \qquad (9.5)$$

Periodic Toda closes the cycle of pairings with $q_{N+1} \equiv q_1$, while *ordinary Toda* ends. These systems are intimately connected with the Cartan catalog of finite-dimensional simple Lie algebras for ordinary Toda, and with infinite-dimensional Kac-Moody or current algebras for periodic Toda. The Toda system given above corresponds to the Lie algebra $SU(N)$ or equivalently $SL(N)$. The Toda lattice systems are integrable for any Lie algebra of this type.

3. The Sutherland-Calogero-Moser systems, with two-body interactions between particles on a line, are given by the following Hamiltonian,

$$H = \sum_{i=1}^N \frac{1}{2} p_i^2 + \sum_{i \neq j}^N V(q_i - q_j) \quad (9.6)$$

where the potential may take one of the following three forms,

$$V(q) \sim \frac{1}{q^2} \quad V(q) \sim \frac{1}{\sin^2 q} \quad V(q) \sim \wp(q) \quad (9.7)$$

where \wp is the Weierstrass elliptic function.

9.4 Elliptic functions

As a meromorphic function of its complex argument z , the Weierstrass function is an elliptic or *doubly periodic* function. Denoting the periods by ω_1, ω_2 , double periodicity requires,

$$\wp(z + m\omega_1 + n\omega_2) = \wp(z) \quad m, n \in \mathbf{Z} \quad (9.8)$$

By Liouville's theorem, such a function can never be holomorphic, and in fact the sum of the residues of its poles must vanish. The Weierstrass function $\wp(z)$ has one double pole (per lattice cell), and is normalized as follows,

$$\wp(z) = \frac{1}{z^2} + \sum_{(m,n) \neq (0,0)} \left(\frac{1}{(z + m\omega_1 + n\omega_2)^2} - \frac{1}{(m\omega_1 + n\omega_2)^2} \right) \quad (9.9)$$

where the sum runs over $m, n \in \mathbf{Z}$. The function $\wp(z)$ satisfies the differential equation,

$$(\wp(z)')^2 = 4\wp(z)^3 - g_2\wp(z) - g_3 = 0 \quad (9.10)$$

where g_2 and g_3 are independent of z , but do depend on ω_1, ω_2 . Its inverse gives the solution to various elliptic integrals, the simplest one of which is given by,

$$\int_{\wp(z_0)}^{\wp(z)} \frac{dw}{\sqrt{4w^3 - g_2w - g_3}} = z - z_0 \quad (9.11)$$

9.5 Solution by Lax pair

In general, it is quite difficult to find all the conservation laws of a given system, or to establish whether it is completely integrable along the lines of the above definition. A dramatic advance was made by Peter Lax (1925-present, third Abel Prize recipient in 2005) by reformulating the problem in an unexpected way.

Consider the Sutherland-Calogero-Moser system for the $1/q^2$ potential. Its Hamilton equations,

$$\dot{q}_i = p_i \quad \dot{p}_i = - \sum_{j \neq i} \frac{2\omega^2}{(q_i - q_j)^3} \quad (9.12)$$

may be obtained as follows. There exist two $N \times N$ matrix-valued functions L, M of p_i and q_i , whose components are given as follows,

$$\begin{aligned} L_{ij} &= p_i \delta_{ij} + \frac{\omega(1 - \delta_{ij})}{q_i - q_j} \\ M_{ij} &= -\frac{\omega(1 - \delta_{ij})}{(q_i - q_j)^2} + \delta_{ij} \sum_{k \neq i} \frac{\omega}{(x_i - x_k)^2} \end{aligned} \quad (9.13)$$

Now, if L and M satisfy the Lax equation,

$$\dot{L} = [L, M] \quad (9.14)$$

then it follows that q_i and p_i must satisfy the Hamilton equations for the system. When this is the case, the pair L, M is referred to as a *Lax pair*. How does one find a Lax pair? Some guessing and some luck are helpful, since there is no general method known, although Estabrook and Wahlquist's *prolongation structure method* claims to transform the problem into an algebraic one. Lax pairs may be derived for each one of the more sophisticated problems, namely spinning top of G/H (although this system can be solved by simpler methods), the Toda Lattice, and the trigonometric and elliptic SCM systems. The matrices in the Lax pair always obey the same Lax equation.

What good does a Lax pair do? To see this, it suffices to look at the Lax equation. In fact, the Lax equation is essentially the Heisenberg equations of motion in quantum mechanics, and we know from quantum mechanics that it is solved by the evolution matrix,

$$L(t) = S(t)L(0)S(t)^{-1} \quad (9.15)$$

except that here, since no factor of i occurs, the evolution matrix S is not unitary. The matrix $S(t)$ itself satisfies the equation,

$$\dot{S} = -MS \quad (9.16)$$

Now the key observation is that as a result of the solution (9.15) of the Lax equation, the eigenvalues of $L(t)$ are time-independent. One refers to this type of time-evolution as *isospectral deformation*. In other words, the eigenvalues are all constants of motion ! While the eigenvalues may be hard to find explicitly (when N is larger than 2 or 3), one may equivalently characterize the constants of motion by the successive powers of $L(t)$,

$$I_n = \text{tr}(L(t)^n) \quad n = 1, \dots, N \quad (9.17)$$

In fact, the time-independence of these quantities may be seen directly from the Lax equation,

$$\dot{I}_n = \sum_{k=0}^{n-1} \text{tr}(L^k \dot{L} L^{n-k-1}) = n \text{tr}([L, M] L^{n-1}) = 0 \quad (9.18)$$

The fact that the integrals mutually Poisson commute may be easily checked by explicit calculation with the help of the following partial calculations,

$$\frac{\partial L_{ij}}{\partial p_\alpha} = \delta_{i,\alpha} \delta_{j,\alpha} \quad \frac{\partial L_{ij}}{\partial q_\alpha} = \frac{\omega(1 - \delta_{ij})}{(q_i - q_j)^2} (\delta_{j,\alpha} - \delta_{i,\alpha}) \quad (9.19)$$

The first two integrals,

$$I_1 = \sum_{i=1}^N p_i \quad \frac{1}{2} I_2 = H = \frac{1}{2} \sum_{i=1}^N p_i^2 - \frac{1}{2} \sum_{i \neq j}^N \frac{\omega^2}{(q_i - q_j)^2} \quad (9.20)$$

are just total momentum and total energy.

Finally, it is worth clarifying how the Lax equation can be computed efficiently. One starts from the Lax operator L , given in the first line of (9.13). Knowing that $\text{tr} L^2$ must be conserved informs us that it must equal $2H$, since the normalizations of the kinetic terms must agree. By the Hamilton equations, we then get the equations of motion of (9.12). It remains to determine M , and work out the Lax equation. To do so, we split L into two parts, L^S which is symmetric, and L^A which is anti-symmetric,

$$L_{ij} = L_{ij}^S + L_{ij}^A \quad L_{ij}^S = p_i \delta_{ij} \quad L_{ij}^S = \frac{\omega(1 - \delta_{ij})}{q_i - q_j} \quad (9.21)$$

in terms of which the Lax equation splits into two,

$$\begin{aligned} \dot{L}^S &= [L^A, M] \\ \dot{L}^A &= [L^S, M] \end{aligned} \quad (9.22)$$

Starting with the last equation, we compute separately,

$$\dot{L}_{ij}^A = -\frac{\omega(1 - \delta_{ij})}{(q_i - q_j)^2}(\dot{q}_i - \dot{q}_j) \quad [L^S, M]_{ij} = M_{ij}(p_i - p_j) \quad (9.23)$$

Equating both tells us that M must be a diagonal matrix $A_i \delta_{ij}$ (which commutes with the general diagonal matrix L^S) plus the first term on the rhs of (9.13). Using now also the first equation of (9.22), we determine also A_i to be given by the second term in M in formula (9.13). The whole procedure involves a certain amount of guess work, but it is not intrinsically difficult to work out.

9.6 The Korteweg de Vries (or KdV) equation

The KdV equation is a one-dimensional reduction of the Navier-Stokes equation. It was invented to describe the long-distance motion of water in long narrow channels, where transverse motion is of no interest (very appropriately carried out by two Dutchmen). Denote by x the coordinate for the space dimension of interest, by t the time, and let ∂_x and ∂_t denote the partial derivatives with respect to these variables. The height of the water (i.e. the amplitude of the wave) will be denoted by the field $u(t, x)$. The KdV equation is then,

$$\partial_t u + \partial_x^3 u - 3\partial_x u^2 = 0 \quad (9.24)$$

This is a non-linear partial differential equation; NEVER easy to solve !

Before investigating the existence of a Lax pair and the integrability of the system, we point out some of its most useful properties. (1) The KdV equation may be recast as a conservation equation,

$$\partial_t u + \partial_x(\partial_x^2 u - 3u^2) = 0 \quad (9.25)$$

(2) KdV derives from a Lagrangian. To see this, we introduce a new field ϕ by $u = \partial_x \phi$. The corresponding field equation for ϕ is then the Euler-Lagrange equation of the action,

$$S[\phi] = \int dx dt \left(\partial_x \phi \partial_t \phi - (\partial_x \phi)^3 - \frac{1}{2}(\partial_x^2 \phi)^2 \right) \quad (9.26)$$

KdV has simple and very important solutions, to be discussed in the next section.

9.7 The KdV soliton

The KdV equation admits solitary waves, or *soliton* solutions. There are various definitions of what a soliton is. Here, it will suffice to view a soliton as a solution to a non-linear wave

equation which suffers no dispersion in its time-evolution. The soliton solution for the KdV equation can be written down explicitly, by looking for solutions of the following form,

$$u(t, x) = f(x - vt) \quad (9.27)$$

where the constant v represents the velocity of the wave profile. Substituting the Ansatz of (9.27) into (9.24) reduces to an ordinary, but non-linear, equation for f ,

$$-vf' + f''' - 3(f^2)' = 0 \quad (9.28)$$

where $'$ denotes the derivative with respect to x . It is straightforward to integrate this equation once, and we find,

$$-vf + f'' - 3f^2 - c = 0 \quad (9.29)$$

where c is an integration constant. Less obviously, this equation can also be integrated, as may be seen most easily by multiplying (9.29) by a factor of f' . We find,

$$(f')^2 - 2f^3 - vf^2 - 2cf + \tilde{c} = 0 \quad (9.30)$$

where \tilde{c} is a new integration constant.

Concentrating on solutions which tend to zero at infinity, namely $f(x), f'(x) \rightarrow 0$ as $x \rightarrow \pm\infty$, we must have $\tilde{c} = 0$. Next, we write $f(x) = -g(x)^2$, so that the equation for g becomes,

$$4(g')^2 + 2g^4 - vg^2 + 2c = 0 \quad (9.31)$$

Requiring again that $g(x), g'(x) \rightarrow 0$ as $x \rightarrow \pm\infty$, we must have $c = 0$, which results in,

$$4(g')^2 + 2g^4 - vg^2 = 0 \quad (9.32)$$

For $v < 0$, this equation has no (real) solutions. For $v > 0$, it may easily be integrated in terms of hyperbolic functions, and we find,

$$u(t, x) = \frac{v/2}{\text{ch}^2\left(\frac{\sqrt{v}}{2}(x - vt - x_0)\right)} \quad (9.33)$$

This is the KdV soliton solution. Note that if we analytically continue v from positive to negative, the hyperbolic function ch becomes a trigonometric \cos function, but now the solution no longer vanishes at $x = \pm\infty$, and it is also singular. So, the soliton is uni-directional. Note that this lack of time-reversal of the solution is not surprising, as the equation itself lacks time-reversal invariance and parity invariance, though it is invariant under combined parity and time-reversal.

9.8 Integrability of KdV by Lax pair

The KdV equation also admits a Lax pair representation. In fact, Lax made his original discovery within the setting of the KdV system. The Lax pair now consists of a pair of differential operators instead of finite-dimensional matrices,

$$\begin{aligned} L &= -\partial_x^2 + u \\ M &= \partial_x^3 - 6u\partial_x - 3(\partial_x u) \end{aligned} \quad (9.34)$$

where $u(t, x)$ is a function of both t and x . Notice that L is essentially a Schrödinger operator with potential u . The KdV equation follows from the Lax equation,

$$\partial_t L = [L, M] \quad (9.35)$$

and as a result, when u obeys the KdV equation, then the time evolution of the operator L preserves the spectrum of L . It is easy to set up the scattering problem, and define the scattering data. To do so, we consider the associated wave equation,

$$L\psi(\lambda, x) = \lambda\psi(\lambda, x) \quad (9.36)$$

Assuming that the field $u(t, x)$ falls off to zero as $x \rightarrow \pm\infty$, a wave incoming from $x = -\infty$ will be take the following asymptotic form, at fixed time,

$$\psi(\lambda, x) = e^{ikx} \quad \text{as } x \rightarrow -\infty \quad (9.37)$$

with $\lambda = k^2$. Crossing the region where the profile of u is non-trivial, the wave will be scattered producing transmitted and reflected components. Thus, the asymptotics of the full solution will take on the following form,

$$\begin{aligned} x \rightarrow -\infty & \quad \psi(\lambda, x) \sim e^{ikx} + r(\lambda)e^{-ikx} \\ x \rightarrow +\infty & \quad \psi(\lambda, x) \sim t(\lambda)e^{ikx} \end{aligned} \quad (9.38)$$

where, as usual, the coefficients $r(\lambda)$ and $t(\lambda)$ are referred to as the reflection and transmission coefficients. They satisfy the unitarity relations,

$$|r(\lambda)|^2 + |t(\lambda)|^2 = 1 \quad (9.39)$$

Of course, for a general solution of the KdV equation, it may be quite difficult or plain impossible to compute $r(\lambda)$ and $t(\lambda)$. Still, the following general results of *inverse scattering theory* prove to be incredibly powerful, and relevant.

The existence of solitons, and the possibility that the KdV equation may be solvable was first proposed by Martin Kruskal in 1967, largely based on numerical work. In the same year, Gardner, Greene, Kruskal and Miura developed methods to solve the KdV equation (at least partially) by analytical methods. Lax introduced his pair of operators for KdV in 1968, and the above scattering method was used to exhibit explicitly the action-angle variables for the system by Zakharov and Faddeev in 1971. The result is as follows. The action variables are all contained in the function $t(\lambda)$, while the angle variables are $\xi(\lambda) = \arg(r(\lambda))$. They satisfy the following relations with respect to the Poisson brackets associated with the original KdV fields,

$$\begin{aligned}\{t(\lambda), t(\mu)\} &= 0 \\ \{t(\lambda), \xi(\mu)\} &= \delta(\lambda - \mu) \\ \{\xi(\lambda), \xi(\mu)\} &= 0\end{aligned}\tag{9.40}$$

The system possesses an infinite number of conserved charges, which may be obtained quite explicitly by expanding the transmission coefficient as follows,

$$t(\lambda) = \sum_{n=1}^{\infty} I_n \lambda^{-n}\tag{9.41}$$

The Hamiltonian is the first one of these conserved charges. The one-soliton solution corresponds to a transmission coefficient with a single pole in the (lower) complex plane. By generalizing the form of $t(\lambda)$ and including N poles (in the lower half plane), one can construct, quite explicitly, solutions with N fully interacting solitons.

Other systems with just 1 space dimension may be solved by the inverse scattering method as well. They include the *non-linear Schrödinger equation*, the *sine-Gordon equation*, and its generalization the *Toda field theories*.

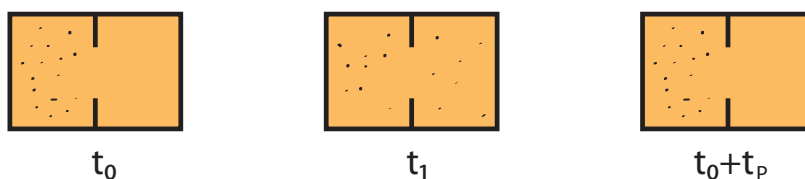
10 Global properties of Hamiltonian mechanics

In this section, we shall present a synopsis of the mathematical apparatus of manifolds, vector field, exterior differential forms, and exterior differential calculus. We shall then reformulate the equations of mechanics in terms of this language. The ultimate purpose of such a mathematical formulation of mechanics is two fold. First, the geometrical structure of Hamilton's equations, as flows on phase space is natural and beautiful. Second, these mathematical tools are what is needed to address global and long-time-evolution problems in mechanics. Recall that so far, given a physical mechanical problem, our main concerns have been with introducing appropriate sets of generalized, deriving the Lagrangian, the constraints and the Hamiltonian, and then to solve exactly wherever possible. We have introduced the mathematical tools of group theory to help in this respect. But what is to be done when a system does not lend itself to exact solution ?

10.1 The Poincaré Recurrence Theorem

To illustrate the power of abstract mathematical results when dealing with the long-time-evolution of mechanical systems, we state here one of the most famous, and most counterintuitive such results, the Poincaré recurrence theorem. Essentially, the Theorem states that a Hamiltonian system will, after a *sufficiently long time*, return to a point in phase space very close to its point of departure.

More precisely, time-evolution of a Hamiltonian system will define a flow $\phi(t) = (p_i(t), q_i(t))$ on phase space. If the flow is confined to a bounded subspace D of phase space (such as when $V(q_i) \geq 0$ and $V(q_i) \rightarrow +\infty$ when $q_i \rightarrow \infty$, and the total energy of the system is fixed), then in any neighborhood U of any point of D there is a point $x \in U$ such that $\phi(t_0 + t_P) = \phi(t_0)$. The time t_P is referred to as the *Poincaré recurrence time*.



This theorem leads to counterintuitive applications, for example in statistical mechanics. Suppose we prepare a system of molecules in a container so that at initial time t_0 all molecules are on one side of a partition of the container. After time $t_1 > t_0$, the molecules tend to occupy all available space in both partitions. The theorem tells us that there will be some time t_P after which all the molecules are essentially back in only one partition. When the number of molecules is very large, the Poincaré recurrence time t_P will be very large.

10.2 Definition of a manifold

In a nutshell, a manifold is a space that *locally* looks like \mathbf{R}^n . A Riemannian manifold is a space with a metric that locally looks like \mathbf{R}^n with the Euclidean metric. Now, let's exhibit more precise definitions. I will assume that you are familiar with some basic topological concepts, like a topological space, open sets, neighborhoods, and continuity.

A (topological) manifold M is a topological space such that every point $x \in M$ has an open neighborhood which is *homeomorphic* to an open set in \mathbf{R}^n . (Recall that a homeomorphism φ is a map which is bijective (one-to-one and onto), continuous, and whose inverse map is also continuous.)

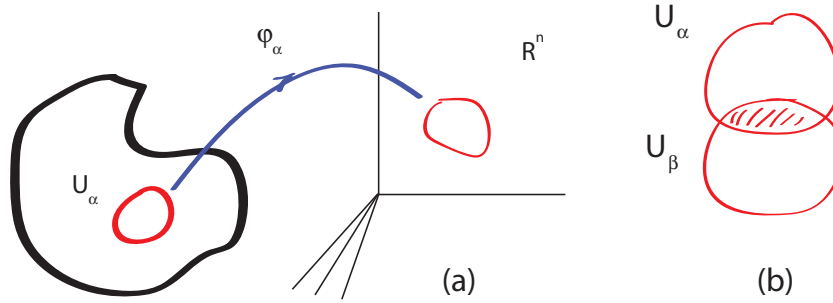


Figure 9: (a) Map φ_α from an open set $\mathcal{U}_\alpha \subset M$ to an open set of \mathbf{R}^n ; (b) Two intersecting open sets $\mathcal{U}_\alpha, \mathcal{U}_\beta$ of M .

Some associated definitions, notations, and terminology is in order.

- By continuity, the value of n must be the same throughout the manifold. The number n is referred to as the *dimension of the manifold* M .
- The map $\varphi_\alpha : x \rightarrow \varphi_\alpha(x) \in \mathbf{R}^n$ for $x \in \mathcal{U}_\alpha$ provides a set of *local coordinates for* M in the neighborhood \mathcal{U}_α of M .
- The pair $(\mathcal{U}_\alpha, \varphi_\alpha)$ is referred to as a *chart* (think of a geographical map).
- An *atlas* $\{(\mathcal{U}_\alpha, \varphi_\alpha), \alpha \in \mathcal{A}\}$ is a set of charts such that we have

$$\bigcup_{\alpha \in \mathcal{A}} \mathcal{U}_\alpha = M \quad (10.1)$$

Thus, an atlas provides a tool by which each point in M is in at least one chart. In other words, an atlas covers the entire manifold with open neighborhoods, each of which may be described in a set of local coordinates.

- Since each \mathcal{U}_α is an open set, the only way the \mathcal{U}_α can cover all points in M is by having non-empty intersections (see figure 9(b)). Points belonging to an intersection of two neighborhoods are described by at least two different local coordinate charts. All the structure of the manifold, and in particular how it may differ from \mathbf{R}^n globally, is contained in these intersections. For each points $x \in \mathcal{U}_\alpha \cap \mathcal{U}_\beta$, we now have two different maps to \mathbf{R}^n , one given by $\varphi_\alpha(x)$ and one given by $\varphi_\beta(x)$. Since the maps $\varphi_\alpha, \varphi_\beta$ are homeomorphisms, it makes sense to consider the composed map,

$$\varphi_\beta \circ \varphi_\alpha^{-1} : \mathbf{R}^n \rightarrow \mathbf{R}^n \quad (10.2)$$

on the open set of \mathbf{R}^n corresponding to $\mathcal{U}_\alpha \cup \mathcal{U}_\beta$. By construction, the map $\varphi_\beta \circ \varphi_\alpha$ is a local homeomorphism. Generally, one distinguished the following types of manifolds,

1. $\varphi_\beta \circ \varphi_\alpha^{-1}$ is continuous (with continuous inverse): M is a *topological manifold*.
2. $\varphi_\beta \circ \varphi_\alpha^{-1}$ is \mathcal{C}^k -differentiable: M is a \mathcal{C}^k -*differentiable manifold*.
3. $\varphi_\beta \circ \varphi_\alpha^{-1}$ is real analytic: M is a *real analytic manifold*.
(Recall that real analytic = completely described by its Taylor series.)
4. $\varphi_\beta \circ \varphi_\alpha^{-1}$ is complex analytic (holomorphic): M is a *complex manifold*.

In physics, especially in non-quantum physics, we are mostly interested in quantities that are differentiable at least once, or often many times. Thus, for example, in mechanics, we shall mostly be interested in *differentiable manifolds*.

10.3 Examples

The following is a list of examples of manifolds, which turn out to all be real analytic,

1. \mathbf{R}^n itself.
2. The direct product of two real analytic manifolds $M_1 \times M_2$ is a real analytic manifold.
3. Spheres S^n may be embedded in \mathbf{R}^{n+1} by coordinates $x_1^2 + x_2^2 + \cdots + x_n^2 + x_{n+1}^2 = R^2$.
4. Hyperboloids $H_{p,q}$ may be embedded in \mathbf{R}^{n+1} , $x_1^2 + x_2^2 + \cdots + x_p^2 - x_{p+1}^2 + \cdots + x_{n+1}^2 = R^2$.
5. Lie groups
6. The annulus, the torus, Möbius strip, the Klein bottle.

The following is a list of spaces that are *not manifolds*,

1. a cone, or generally spaces with singularities;
2. a flag, or generally spaces whose dimension varies;

10.4 Maps between manifolds

Let f be a map from a manifold M with dimension m to a manifold N of dimension n . To study such maps, we use the definition of a manifold for both M and N , and reduce the problem to a study of maps from (open neighborhoods of) \mathbf{R}^m into (open neighborhoods of) \mathbf{R}^n , as exhibited in figure 10.

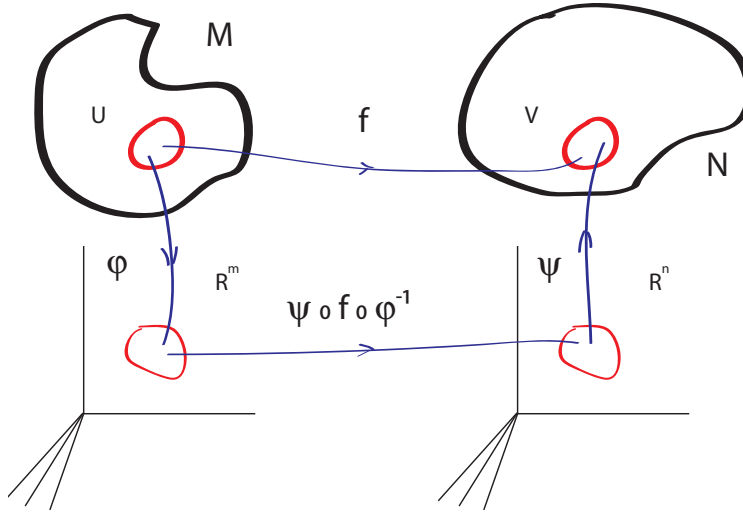


Figure 10: Map f from manifold M into manifold N corresponds to a map from a neighborhood of \mathbf{R}^m to a neighborhood of \mathbf{R}^n .

The map f is \mathcal{C}^k -differentiable if φ, ψ and $\psi \circ f \circ \varphi^{-1}$ are \mathcal{C}^k -differentiable. The map f is complex analytic (or holomorphic) if M and N are complex manifolds, and $\varphi, \psi, \psi \circ f \circ \varphi^{-1}$ are all complex analytic, and so on.

There is an important special case when $M = N$. Maps from M into itself, which are bijective (one-to-one and onto), and \mathcal{C}^k -differentiable are referred to as *diffeomorphisms*. They are sometimes also referred to as *coordinate transformations*, especially in general relativity, but in doing so one may lose sight of the differentiable nature of the transformations. In fact, the principle of general relativity may be stated concisely by saying that the laws of Nature must be invariant under all *diffeomorphisms of space-time*.

10.5 Vector fields and tangent space

Vector fields at a point x in a manifold M span the tangent space to M at the point x . The tangent space is a linear space, i.e. a vector space, which gives a linear model of the manifold in the infinitesimal neighborhood of x . To give a mathematical definition of these objects, we begin by defining vector fields.

Consider a 1-parameter curve $x(\lambda)$ in M , where λ is a real parameter, taking values in some interval of \mathbf{R} . Now consider a real-valued function f on M . These functions fit into our construction of maps between manifolds by taking $N = \mathbf{R}$. In particular, we can evaluate the function f on the curve $x(\lambda)$ in M , which gives us a map from an interval of the real parameter λ into \mathbf{R} . We define *the tangent vector field X at the point x_0* by its action on real functions f as follows,

$$Xf = X(f) \equiv \left. \frac{df(x(\lambda))}{d\lambda} \right|_{\lambda=\lambda_0} \quad x_0 = x(\lambda_0) \quad (10.3)$$

X is a *directional derivative*, associated with the behavior of the curve $x(\lambda)$ near x_0 . It has the following properties,

1. X is linear: $X(f + \alpha g) = X(f) + \alpha X(g)$, for $\alpha \in \mathbf{R}$;
2. X acts as a derivative, i.e. it satisfies Leibnitz's rule: $X(fg) = fX(g) + gX(f)$.

In any local coordinate system on M around the point x_0 , given by (u^1, \dots, u^m) , we have

$$X(f) = \sum_{i=1}^m \frac{\partial f}{\partial u^i} \frac{du^i}{d\lambda} \Big|_{\lambda=\lambda_0} \quad (10.4)$$

an expression that is sometimes written as $\vec{\nabla} f \cdot d\vec{u}/d\lambda$. We may extract from this a more formal expression for what X is, independently of the function f it is being applied to, and we have,

$$X = \sum_{i=1}^m \frac{du^i}{d\lambda} \Big|_{\lambda=\lambda_0} \frac{\partial}{\partial u^i} \quad (10.5)$$

We see that X is a partial differential operator. Clearly, the space of all vector fields X at a point x_0 forms a vector space, whose dimension n coincides with that of the manifold M . A basis in this vector space may be chosen as follows,

$$\partial_i = \frac{\partial}{\partial u^i} \quad i = 1, \dots, n \quad (10.6)$$

This defines the so-called *coordinate basis*. Conversely, any linear combination

$$X = \sum_{i=1}^n \xi^i(u) \Big|_{u(x_0)} \frac{\partial}{\partial u^i} \quad (10.7)$$

evaluated at the point $u_0^i = u^i(x_0)$, produces an infinitesimal curve through x_0 , with tangent vector $\xi^i(u_0)$. The vector space of all tangent vectors at a point $x_0 \in M$ is referred to as the *tangent space*, and denoted $T_{x_0}(M)$.

Having defined vector fields at a single point $x_0 \in M$, we may extend this definition throughout the manifold, and associate with each point $x \in M$, its tangent space $T_x(M)$. In a local coordinate patch, a general vector field then takes the form,

$$X = \sum_{i=1}^n \xi^i(u) \frac{\partial}{\partial u^i} \quad (10.8)$$

In the overlap $\mathcal{U}_\alpha \cap \mathcal{U}_\beta$ between two charts \mathcal{U}_α and \mathcal{U}_β , the same vector field X is expressed in terms of two different coordinate systems, u_α^i and u_β^i . This requires the tangent vectors in the coordinate systems to be related as follows,

$$\sum_{i=1}^n \xi_\alpha^i(u_\alpha) \frac{\partial}{\partial u_\alpha^i} = \sum_{i=1}^n \xi_\beta^i(u_\beta) \frac{\partial}{\partial u_\beta^i} \quad (10.9)$$

Applying both sides successively to all the u_β^i , we get,

$$\xi_\beta^i(u_\beta) = \sum_{j=1}^n \xi_\alpha^j(u_\alpha) \frac{\partial u_\beta^i}{\partial u_\alpha^j} \quad (10.10)$$

Vector fields form a Lie algebra under the operation of commutator of differential operators. Two vector fields X_1, X_2 with component fields ξ_1^i and ξ_2^i have the following commutator,

$$[X_1, X_2] = \sum_{i=1}^n \eta^i \partial_i \quad (10.11)$$

where the composite vector field η is given by,

$$\eta^i = \sum_{j=1}^n \left(\xi_1^j \partial_j \xi_2^i - \xi_2^j \partial_j \xi_1^i \right) \quad (10.12)$$

The Jacobi identity may be checked by explicit computation. The corresponding Lie algebra is the affine version of $GL(n, \mathbf{R})$.

10.6 Poisson brackets and Hamiltonian flows

We can immediately apply the concepts developed in the previous section to Hamiltonian mechanics. Earlier, we have introduced the notion of *phase space* in a somewhat casual manner, namely as the space with coordinates p^i, q^i for $i = 1, 2, \dots, N$. We shall now assume that p^i, q^i provide local coordinates on the manifold of phase space, which we denote by \mathcal{P} . The dimension of \mathcal{P} is clearly $2N$. The Hamiltonian $H(p, q)$ is a real-valued function on \mathcal{P} . Now the Hamiltonian produces a unique vector field X_H on phase space, which is defined by the Poisson bracket. To see this, we consider its action on any other real-valued function $f(p, q)$ on \mathcal{P} by,

$$X_H(f) = \{H, f\} = \sum_{i=1}^N \left(\frac{\partial H}{\partial q^i} \frac{\partial f}{\partial p^i} - \frac{\partial H}{\partial p^i} \frac{\partial f}{\partial q^i} \right) \quad (10.13)$$

In fact, we see that pretty much any scalar function $G(p, q)$ on phase space will generate a vector field, which we may represent somewhat abstractly by,

$$X_G = \{G, \cdot\} = \sum_{i=1}^N \left(\frac{\partial G}{\partial q^i} \frac{\partial}{\partial p^i} - \frac{\partial G}{\partial p^i} \frac{\partial}{\partial q^i} \right) \quad (10.14)$$

where the center dot leaves space for the function to which the vector field is to be applied. The commutator of the vector fields X_{G_1} and X_{G_2} gives a new vector field $X_{\{G_1, G_2\}}$ obtained from the Poisson bracket. To see this, we start from the definitions,

$$\begin{aligned} [X_{G_1}, X_{G_2}]f &= X_{G_1}(X_{G_2}(f)) - X_{G_2}(X_{G_1}(f)) \\ &= \{G_1, \{G_2, f\}\} - \{G_2, \{G_1, f\}\} \\ &= \{\{G_1, G_2\}, f\} \end{aligned} \quad (10.15)$$

where we have use the Jacobi identity in going from line 2 to line 3. Thus, the Lie algebras of vector fields and of Poisson brackets are isomorphic to one another.

A vector field defines a *flow* on the manifold. In mechanics, the Hamiltonian vector field generates the flow of time-evolution,

$$\frac{df}{dt} = \{H, f\} = X_H(f) \quad (10.16)$$

which in turn makes contact with our earliest definition of a vector field in terms of a curve through a point on the manifold.

10.7 Stokes's theorem and grad-curl-div formulas

To motivate the introduction of differential forms, we begin by reviewing some standard stuff from a new perspective. Some of the most basic ingredients in the calculus of several variables are the following notions. We begin by working in flat \mathbf{R}^3 , to be generalized to arbitrary dimension later,

gradient on functions	$\nabla f = \text{grad} f$	
curl on vector fields	$\nabla \times \mathbf{A} = \text{curl} \mathbf{A}$	
divergence on vector fields	$\nabla \cdot \mathbf{j} = \text{div} \mathbf{j}$	(10.17)

These differential operators are encountered in virtually all areas of physics, including fluid dynamics, thermodynamics, and of course Maxwell's theory of electro-magnetism. Associated with these differential operators are integration formulas, which are related by Stokes's theorem, as follows,

$$\begin{aligned}
 \int_a^b d\mathbf{x} \cdot \nabla f &= f(b) - f(a) \\
 \int_{\Sigma} d^2\mathbf{s} \cdot (\nabla \times \mathbf{A}) &= \oint_{\partial\Sigma} d\mathbf{x} \cdot \mathbf{A} \\
 \int_D d^3x (\nabla \cdot \mathbf{j}) &= \oint_{\partial D} d^2\mathbf{s} \cdot \mathbf{j}
 \end{aligned} \tag{10.18}$$

Here, the symbols $\partial\Sigma$ and ∂D denote the boundaries of the 2- and 3-dimensional domains Σ and D respectively. The integration domains, and their boundaries, are represented schematically in figure 11. To any one who has ever worked with Maxwell theory, it is quite familiar

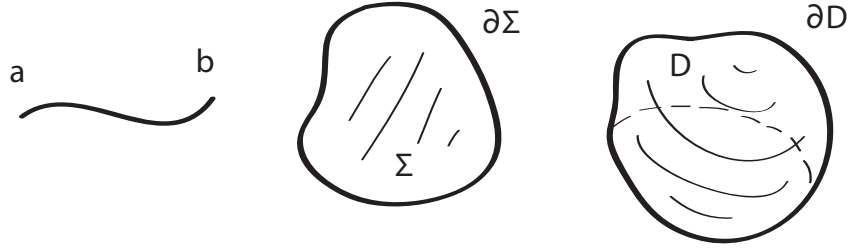


Figure 11: Integration domains: an interval $[a, b]$; a planar domain Σ ; a solid domain D .

that, as a result of Schwarz's identity on mixed derivatives, we have the following relations,

$$\begin{aligned}
 \text{curl} \cdot \text{grad} &= 0 \\
 \text{div} \cdot \text{curl} &= 0
 \end{aligned} \tag{10.19}$$

when applied to arbitrary functions of vector fields. This result goes hand in hand with the fact that the boundary of a boundary of any domain is empty. This is visually manifest from figure 11. Indeed, the boundary $\partial\Sigma$ of the 2-dimensional domain Σ is a *closed curve*, and as such this closed curve has no boundary. Similarly, the boundary ∂D of the 3-dimensional domain D is a *closed 2-dimensional surface*, which again itself has no boundary. Thus, we can write, somewhat formally,

$$\partial\partial = 0 \tag{10.20}$$

a formula which applies to any topological space. This result is reflected in Stokes's integration formulas, and is closely related with the differential relations of (10.19). For example, if we take $\mathbf{A} = \nabla f$ in the second formula of (10.18), the left hand side vanishes because its integrand vanishes identically in view of the differential relations (10.19); the right hand side vanishes not because its integrand vanishes (it does not in general) but rather because $\partial\Sigma$ has no boundary, and then using the first formula of (10.18). Similarly, setting $\mathbf{j} = \nabla \times \mathbf{A}$ in the last formula of (10.18), the lhs again vanishes because its integrand does, and the rhs vanishes because ∂D has no boundary and using the second line. It is this set of observations that justifies writing the integrals on the right hand side with the \oint symbol.

10.8 Differential forms: informal definition

The exterior differential algebra, introduced in Mathematics by Elie Cartan, provides an extremely systematic and compact formalism for unifying and generalizing the grad-curl-div formulas and associated Stokes's theorem. The basic ideas are as follows,

- In the integration formulas, the differential volume element is always *linear* in each independent coordinate differential. For example, in dimensions 2 and 3, we have in Cartesian coordinates, $dx\,dy$ and $dx\,dy\,dz$. We never encounter an object like $(dx)^2$. (Recall that such objects do enter in other setting, like the infinitesimal distance.)
- As a result, we are going to define an algebra in which we formally set $dx\,dx = 0$.
- Since the differential is linear in its argument, consistency requires that we also set $(dx + dy)(dx + dy) = 0$, and this forces us then to have $dx\,dy + dy\,dx = 0$, so that the algebraic product of two differentials must be *anti-symmetric*. To make this property explicit, Cartan introduced the notation $dx \wedge dy = -dy \wedge dx$, an operation referred to as the *wedge product*.

- We introduce the *exterior differential* d (or total derivative). We may do this directly on an arbitrary manifold M of dimension n , and we may express d in terms of a (local) coordinate system x^i , $i = 1, \dots, n$ on M by,

$$d = \sum_{i=1}^n dx^i \frac{\partial}{\partial x^i} \quad (10.21)$$

Denoting the space of differentiable function on M by $\Omega^0(M)$, we may apply d to any function $f \in \Omega^0(M)$, to obtain a 1-form df ,

$$df = \sum_{i=1}^n \frac{\partial f}{\partial x^i} dx^i \quad (10.22)$$

This is not the most general 1-form, as the coefficients of the coordinate differentials dx^i , in general, do not need to form a gradient. The most general real-valued 1-form on M may be expressed in local coordinates by,

$$\omega^{(1)} = \sum_{i=1}^n \omega_i^{(1)}(x) dx^i \quad (10.23)$$

where $\omega_i(x)$ are the components of the 1-form ω with respect to the local coordinate system x^i . The space of all differentiable 1-forms is denoted $\Omega^1(M)$.

- It is immediate to generalize this construction and to introduce p -forms as follows,

$$\omega^{(p)} = \frac{1}{p!} \sum_{i_1, \dots, i_p} \omega_{i_1, \dots, i_p}^{(p)}(x) dx^{i_1} \wedge dx^{i_2} \wedge \dots \wedge dx^{i_p} \quad (10.24)$$

In view of the total anti-symmetry of the wedge product $dx^{i_1} \wedge dx^{i_2} \wedge \dots \wedge dx^{i_p}$ under the interchange of any pair of its indices, the coefficient functions $\omega_{i_1, \dots, i_p}^{(p)}(x)$ must be totally anti-symmetric in its indices. The space of all differentiable p -forms is denoted $\Omega^{(p)}(M)$. Note that for $p > n$, all differential forms vanish.

10.9 Structure relations of the exterior differential algebra

The informal definition of the wedge product, and its anti-symmetry on coordinate differentials $dx^i \wedge dx^j = -dx^j \wedge dx^i$, together with its bilinear property, may be formalized in the following simple manner. For $\omega_1 \in \Omega^{(p)}$, $\omega_2, \omega_2' \in \Omega^{(q)}$, and $\omega_3 \in \Omega^{(r)}$, with $p, q, r \leq n$, we

have

$$\begin{aligned}
\omega_1 \wedge \omega_2 &\in \Omega^{(p+q)} \\
\omega_1 \wedge \omega_2 &= (-)^{pq} \omega_2 \wedge \omega_1 \\
\omega_1 \wedge (\alpha \omega_2 + \beta \omega'_2) &= \alpha \omega_1 \wedge \omega_2 + \beta \omega_1 \wedge \omega'_2 \\
\omega_1 \wedge \omega_2 \wedge \omega_3 &= (\omega_1 \wedge \omega_2) \wedge \omega_3 = \omega_1 \wedge (\omega_2 \wedge \omega_3)
\end{aligned} \tag{10.25}$$

where $\alpha, \beta \in \Omega^{(0)}$. The next-to-last property is linearity, and the last property is associativity of the algebra,

$$\Omega(M) = \bigoplus_{p=0}^n \Omega^{(p)}(M) \tag{10.26}$$

Finally, the properties of the differential operator d are as follows. At an informal level, they may be derived from the elementary relation $d(dx^i) = 0$ on the basic coordinate differentials, and the use of Schwarz's identity on mixed derivatives. More formally, we may define the exterior differential d as follows,

$$\begin{aligned}
d : \Omega^{(p)} &\rightarrow \Omega^{(p+1)} \\
d : \Omega^{(n)} &\rightarrow 0 \\
d(d\Omega^{(p)}) &= 0
\end{aligned} \tag{10.27}$$

such that the following rules hold,

$$\begin{aligned}
d(\omega + \omega') &= d\omega + d\omega' \\
df &= \sum_{i=1}^n (\partial_i f) dx^i \quad f \in \Omega^{(0)}(M)
\end{aligned} \tag{10.28}$$

The one thing one needs to be careful about is Leibnitz's rule for differential forms $\omega_1 \in \Omega^{(p)}$, $\omega_2 \in \Omega^{(q)}$, which becomes,

$$d(\omega_1 \wedge \omega_2) = (d\omega_1) \wedge \omega_2 + (-)^p \omega_1 \wedge (d\omega_2) \tag{10.29}$$

With the help of these rules, we can now work out the exterior derivative of an arbitrary differential form $\omega^{(p)} \in \Omega^{(p)}(M)$, as was given in (10.24),

$$d\omega^{(p)} = \frac{1}{p!} \sum_{i_1, \dots, i_p} \sum_{k=1}^n \partial_k \omega_{i_1, \dots, i_p}^{(p)}(x) dx^k \wedge dx^{i_1} \wedge dx^{i_2} \wedge \dots \wedge dx^{i_p} \tag{10.30}$$

One verifies that, applying a second exterior differential, and using $d(dx^i) = 0$, we have

$$d(d\omega^{(p)}) = \frac{1}{p!} \sum_{i_1, \dots, i_p} \sum_{k, \ell=1}^n \partial_\ell \partial_k \omega_{i_1, \dots, i_p}^{(p)}(x) dx^\ell \wedge dx^k \wedge dx^{i_1} \wedge dx^{i_2} \wedge \dots \wedge dx^{i_p} \quad (10.31)$$

The double derivative $d(d\omega^{(p)})$ vanishes in view of $dx^k \wedge dx^\ell + dx^\ell \wedge dx^k = 0$ and the symmetry under interchange of k, ℓ in $\partial_\ell \partial_k \omega_{i_1, \dots, i_p}^{(p)}(x)$ thanks to Schwarz's identity. The differential relations of (10.19) are found to be just special cases of these formulas when the dimension of M is 3.

10.10 Integration and Stokes's Theorem on forms

There is a natural pairing of manifolds and differential forms, given by integration. Let M be a manifold of dimension n , and let $\omega \in \Omega^{(n)}(M)$, then we have a map,

$$\int_M \omega^{(n)} \rightarrow \mathbf{R} \quad (10.32)$$

But we may also consider the integration of forms $\omega^{(p)}$ of lower rank p on sub-manifolds $\Sigma^{(p)}$ of dimension p of M by

$$\int_{\Sigma^{(p)}} \omega^{(p)} \rightarrow \mathbf{R} \quad (10.33)$$

In fact, one can generalize further, and integrate a form of rank p over a sub-manifold $\Sigma^{(q)}$ of dimension q , which gives a map,

$$\int_{\Sigma^{(q)}} \omega^{(p)} \rightarrow \begin{cases} 0 & \text{if } q > p \\ \mathbf{R} & \text{if } q = p \\ \Omega^{(p-q)}(M) & \text{if } q < p \end{cases} \quad (10.34)$$

Stokes's theorem is now very simply stated, in all generality,

$$\int_D d\omega^{(p)} = \oint_{\partial D} \omega^{(p)} \quad (10.35)$$

where D is any (differentiable) domain of dimension p in M .

10.11 Frobenius Theorem for Pfaffian systems

A very useful application of exterior differential algebra is to the integrability conditions for systems of equations. A simple general case is the Pfaffian system, given on a manifold M of

dimension m in terms of the vanishing of $0 < k < m$ linearly independent differential 1-forms ω_i where $i = 1, \dots, k$. If it exists, the solution will be a $m - k$ dimensional sub-manifold of M . But in general, such a system need not be integrable. The condition for integrability is given by Frobenius's theorem. The necessary and sufficient condition for integrability is that we have

$$d\omega_i \wedge \omega_1 \wedge \dots \wedge \omega_k = 0 \quad (10.36)$$

while the linear independence of the 1-forms guarantees that $\omega_1 \wedge \dots \wedge \omega_k \neq 0$. For example, take a 2-dimensional manifold with a single form ω . Since $d\omega \wedge \omega = 0$ automatically in view of it being rank 3 on a manifold of dimension 2, the system $\omega = 0$ is always integrable. On a 3-dimensional manifold, with a single form ω , we now have a non-trivial condition, $d\omega \wedge \omega = 0$. In components,

$$\begin{aligned} \omega &= \alpha dx + \beta dy + \gamma dz \\ d\omega \wedge \omega &= \mathcal{C} dx \wedge dy \wedge dz = 0 \\ \mathcal{C} &= \alpha(\partial_y \gamma - \partial_z \beta) + \beta(\partial_z \alpha - \partial_x \gamma) + \gamma(\partial_x \beta - \partial_y \alpha) \end{aligned} \quad (10.37)$$

The theorem will be of use in our subsequent discussion of integrable systems.