Contents

I Newtonian Mechanics 5

1 Newton’s equations of motion 7
   1.1 Empirical assumptions . . . . . . . . . . . . . . . . . . . . . . . . 7
   1.2 Energy . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9
   1.3 Conservative systems . . . . . . . . . . . . . . . . . . . . . . . 12
   1.4 Nonconservative systems . . . . . . . . . . . . . . . . . . . . . . 14
   1.5 Reversible systems . . . . . . . . . . . . . . . . . . . . . . . . . 16
   1.6 Linear momentum . . . . . . . . . . . . . . . . . . . . . . . . . 17
   1.7 Angular momentum . . . . . . . . . . . . . . . . . . . . . . . . . 18
   1.8 Exercises . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 20

2 Examples 27
   2.1 One-dimensional systems . . . . . . . . . . . . . . . . . . . . . . 27
   2.2 Central fields . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30
   2.3 Closed bounded orbits . . . . . . . . . . . . . . . . . . . . . . . . 32
   2.4 Kepler’s problem . . . . . . . . . . . . . . . . . . . . . . . . . . . 33
   2.5 Virial theorem . . . . . . . . . . . . . . . . . . . . . . . . . . . . 35
   2.6 Central field scattering . . . . . . . . . . . . . . . . . . . . . . . . 36
   2.7 Exercises . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 40

II Lagrangian Mechanics 45

3 Euler–Lagrange equations of motion 47
   3.1 Principle of least action . . . . . . . . . . . . . . . . . . . . . . . 47
   3.2 Conservative systems . . . . . . . . . . . . . . . . . . . . . . . . . 51
   3.3 Equivalence to Newton’s equations . . . . . . . . . . . . . . . . . 54
   3.4 Nonconservative systems . . . . . . . . . . . . . . . . . . . . . . . 56
   3.5 Momentum and conservation . . . . . . . . . . . . . . . . . . . . . 57
   3.6 Noether’s theorem . . . . . . . . . . . . . . . . . . . . . . . . . . . 59
   3.7 Exercises . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 61
## CONTENTS

### 4 Constraints

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 The d’Alembert–Lagrange principle</td>
<td>69</td>
</tr>
<tr>
<td>4.2 Gauss’ principle of least constraint</td>
<td>71</td>
</tr>
<tr>
<td>4.3 Integrable constraints</td>
<td>74</td>
</tr>
<tr>
<td>4.4 Non-holonomic constraints</td>
<td>74</td>
</tr>
<tr>
<td>4.5 Exercises</td>
<td>76</td>
</tr>
</tbody>
</table>

### 5 Hamilton–Jacobi equation of motion

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Hamilton–Jacobi equation</td>
<td>81</td>
</tr>
<tr>
<td>5.2 Separation of variables</td>
<td>84</td>
</tr>
<tr>
<td>5.3 Conditionally periodic motion</td>
<td>85</td>
</tr>
<tr>
<td>5.4 Geometric optics analogy</td>
<td>87</td>
</tr>
<tr>
<td>5.5 Exercises</td>
<td>89</td>
</tr>
</tbody>
</table>

### III Hamiltonian Mechanics

### 6 Hamilton’s equations of motion

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Hamilton’s equations</td>
<td>95</td>
</tr>
<tr>
<td>6.2 Legendre transformation</td>
<td>97</td>
</tr>
<tr>
<td>6.3 Liouville’s theorem</td>
<td>100</td>
</tr>
<tr>
<td>6.4 Poisson bracket</td>
<td>102</td>
</tr>
<tr>
<td>6.5 Canonical transformations</td>
<td>105</td>
</tr>
<tr>
<td>6.6 Infinitesimal canonical transformations</td>
<td>108</td>
</tr>
<tr>
<td>6.7 Canonical variables</td>
<td>110</td>
</tr>
<tr>
<td>6.8 Exercises</td>
<td>111</td>
</tr>
</tbody>
</table>

### 7 Symplectic geometry

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1 Symplectic structure</td>
<td>115</td>
</tr>
<tr>
<td>7.2 Hamiltonian vector fields</td>
<td>116</td>
</tr>
<tr>
<td>7.3 Integral invariants</td>
<td>118</td>
</tr>
<tr>
<td>7.4 Poisson bracket</td>
<td>120</td>
</tr>
<tr>
<td>7.5 Time-dependent systems</td>
<td>121</td>
</tr>
<tr>
<td>7.6 Locally Hamiltonian vector fields</td>
<td>123</td>
</tr>
<tr>
<td>7.7 Exercises</td>
<td>124</td>
</tr>
</tbody>
</table>

### 8 Contact geometry

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1 Contact structure</td>
<td>127</td>
</tr>
<tr>
<td>8.2 Hamiltonian vector fields</td>
<td>129</td>
</tr>
<tr>
<td>8.3 Dynamics</td>
<td>130</td>
</tr>
<tr>
<td>8.4 Contact transformations</td>
<td>131</td>
</tr>
<tr>
<td>8.5 Time-dependent systems</td>
<td>134</td>
</tr>
<tr>
<td>8.6 Exercises</td>
<td>136</td>
</tr>
</tbody>
</table>

### Bibliography

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bibliography</td>
<td>139</td>
</tr>
</tbody>
</table>
Part I

Newtonian Mechanics

The Newtonian perspective is the most fundamental interpretation of the motion of a mechanical system. Newton’s principle of determinacy is an experimental observation, the mathematical phrasing of which yields Newton’s equations. Together with Galileo’s principle of relativity, these form the axioms of classical mechanics. We will also see how additional empirical assumptions (conservative system, closed system, and the law of action and reaction) yield both physically observable effects (conservation laws and symmetries) and mathematical consequences (special properties for Newton’s equations and its solutions). Although Newtonian mechanics are naturally phrased on Euclidean space, we will observe the fundamental roles of configuration and phase space in order to later generalize this perspective to include a broader class of systems and coordinates.
Chapter 1
Newton’s equations of motion

We build the theory starting from essential experimental observations and then develop some of its immediate consequences. This presentation is based on [Arn89 Ch. 1], [AKN06 Ch. 1], [Gol51 Ch. 1], and [Str15 Ch. 5–7].

1.1 Empirical assumptions

Classical mechanics studies the motion of particles in Euclidean space \( \mathbb{R}^d \). A particle or point mass is a model for a physical object consisting of two pieces of information, the body’s mass (a positive real number) and its position in Euclidean space \( \mathbb{R}^d \), and so the object’s spatial dimensions are ignored. We will consider particle systems, which are collections of \( N \) such particles with positions often denoted \( x_i \in \mathbb{R}^d \) and masses \( m_i \). Altogether, the collection of all positions \( x = (x_1, \ldots, x_N) \) constitute the configuration space \( \mathbb{R}^d \times \cdots \times \mathbb{R}^d = \mathbb{R}^{Nd} \), whose dimension \( Nd \) is referred to as the degrees of freedom of the system.

The motion of a system of \( N \) particles is described by \( N \) maps \( x_i : I \to \mathbb{R}^d \) for \( I \subset \mathbb{R} \) an interval, which are called the particle’s trajectories or the motion of the system. In order to describe how the particles trace their trajectories we need not only the particles’ velocities \( \dot{x}_i \) and accelerations \( \ddot{x}_i \), but also the momentum of the \( i \)th particle,

\[
p_i = m_i \dot{x}_i, \quad i = 1, \ldots, N.
\]

(We will denote time derivatives by \( \dot{f} = df/dt \).) The positions and momenta of the system span the Euclidean space \( \mathbb{R}^{Nd} \times \mathbb{R}^{Nd} \), which is called phase space, and its dimension is twice the number of degrees of freedom. We will often refer to the trajectories of the system plotted in phase space as the system’s phase portrait.

Newton’s principle of determinacy is the experimental observation that the initial state of a mechanical system—the information of all the positions \( x(t_0) \)
and velocities \( \dot{x}(t_0) \) at some moment in time \( t_0 \)—uniquely determines any system’s motion. Mathematically, this means for a particle system there exists a function \( F : (\mathbb{R}^{Nd} \setminus \Delta) \times \mathbb{R}^{Nd} \times \mathbb{R} \to \mathbb{R}^{Nd} \) such that

\[
\ddot{x} = F(x, \dot{x}, t).
\] (1.2)

Here, \( \Delta = \bigcup_{i<j} \{ x_i = x_j \} \) is the union of diagonals, so that we do not consider two particles occupying the same position. Eq. (1.2) is called Newton’s equation, or sometimes Newton’s second law. By default we will assume the particle masses \( m_i \) are all constant, and so eq. (1.2) takes the form

\[
m_i \ddot{x}_i = F_i(x, \dot{x}, t), \quad i = 1, \ldots, N.
\] (1.3)

Historically, we would experimentally observe that a particle’s acceleration is inversely proportional to its mass in order to justify the definition eq. (1.1) of momentum. Treating the velocity independently of position is a common practice in math to reduce a system of ODEs to first-order, so that instead of a system of \( N \) second-order differential equations on configuration space we consider a system of \( 2N \) first-order differential equations on phase space, equations (1.1) and (1.2).

In accordance with the uniqueness part of Newton’s principle and in order to focus on physical techniques, we will always implicitly assume that \( F \) is smooth (infinitely differentiable) on \( (\mathbb{R}^{Nd} \setminus \Delta) \times \mathbb{R}^{Nd} \times \mathbb{R} \); the theorem of existence and uniqueness for ODEs then tells us there always exists a unique solutions to the system of differential equations eq. (1.2) for any initial state and that this solution is also smooth. Although the theorem only guarantees a solution for a finite interval of time, from experimental observation we expect that for any naturally occurring system that the the interval can be extended to all of \( \mathbb{R} \), and so we will occasionally assume that the solutions to our mathematical model eq. (1.3) we have the motion for all time.

Lastly, we will formulate one more mathematical assumption: that the physical laws governing the system’s motion is independent of the choice of coordinates and origin we impose on \( \mathbb{R}^d \). A transformation \( \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d \times \mathbb{R} \) is Galilean provided that it is an affine transformation (a linear transformation and a translation) if it preserves time intervals and for any fixed \( t \in \mathbb{R} \) is an isometry of configuration space \( \mathbb{R}^d \). So if \( g(x, t) = (x', t') \) is . The set of Galilean transformations form a group under function composition.

Example 1.1. It is straightforward to verify that the following are all Galilean transformations:

1. Translations: \( g_1(x, t) = (x + x_0, t + t_0) \) for some fixed \( x_0 \in \mathbb{R}^d, \ t_0 \in \mathbb{R} \)
2. Rotations and reflections: \( g_2(x, t) = (Ax, t) \) for some fixed orthogonal transformation \( A \in O(d) \)
3. Uniform motion with constant velocity: \( g_3(x, t) = (x + vt, t) \) for some fixed velocity \( v \in \mathbb{R}^d \).
1.2. ENERGY

In fact, these examples generate the entire Galilean group (see Exercise 1.3). 

*Galileo’s principle of relativity* is the experimental observation that for the special case of an isolated system invariant under any Galilean transformation, and that a reference frame—a choice of origin and coordinate axes for \( \mathbb{R}^d \)—can always be chosen such that this condition is satisfied. Any such frame is called inertial, and the principle also asserts that all coordinate systems in uniform rectilinear motion with respect to an inertial frame is also inertial. (This is observed, for example, in a car traveling at a constant velocity and noting that motion inside the car is as if the car were at rest.) Physically, this is requiring that space be homogeneous and isotropic and that time be homogeneous. Geometrically, this principle requires that if we graph the trajectories of a particle system in configuration space and apply a Galilean transformation to the entire graph, then the resulting graph will still be trajectories.

If Newton’s equations eq. (1.3) with respect to such an inertial coordinate system, then they must be invariant with respect to the Galilean group. Let \( \mathbf{x}(t) \) denote a solution in an inertial coordinate system. Applying the Galilean group generators of Example 1.1, we find the following conditions on \( \mathbf{F}_i \):

1. Translations: \( \mathbf{F}_i(\mathbf{x}, \dot{\mathbf{x}}, t) \equiv \mathbf{F}_i(\mathbf{x}_j - \mathbf{x}_k, \dot{\mathbf{x}}) \).

2. Rotations and reflections: \( \mathbf{F}_i(A\mathbf{x}, A\dot{\mathbf{x}}) = A\mathbf{F}_i(\mathbf{x}, \dot{\mathbf{x}}) \) for \( A \in O(n) \).

3. Uniform motion: \( \mathbf{F}_i(\mathbf{x}_j - \mathbf{x}_k, \dot{\mathbf{x}}) \equiv \mathbf{F}_i(\dot{\mathbf{x}}_j - \dot{\mathbf{x}}_k) \).

Note that the third type of transformations in Example 1.1 change neither \( \ddot{\mathbf{x}} \) nor \( \mathbf{x}_i - \mathbf{x}_j \). In the case \( N = 1 \) of one particle, this requires that \( \mathbf{F} \) is independent of \( \mathbf{x}, \dot{\mathbf{x}}, t \) and is rotationally invariant, and so \( \mathbf{F} \equiv 0 \). Consequently, the particle’s acceleration in an inertial coordinate system is equal to zero; this is *Newton’s first law.*

### 1.2 Energy

The kinetic energy of the \( i \)th particle is

\[
T_i = \frac{1}{2} m_i |\dot{\mathbf{x}}_i|^2 = \frac{1}{2m_i} |\mathbf{p}_i|^2 = \frac{1}{2} \dot{\mathbf{x}}_i \cdot \mathbf{p}_i,
\]

and summing over \( i = 1, \ldots, N \) gives us the total system kinetic energy \( T \). From observation, we know that the magnitude of the velocity and hence the kinetic energy can be increased and decreased by the force \( \mathbf{F}_i \) acting on the \( i \)th particle, depending on the force’s magnitude and direction. This is measured through the work done by the force \( \mathbf{F}_i \) on the \( i \)th particle from time \( t_0 \) to \( t \), defined by the line integral

\[
W_i = \int_{\mathbf{x}(t_0)}^{\mathbf{x}(t)} \mathbf{F}_i \cdot d\mathbf{s}_i = \int_{t_0}^t \mathbf{F}_i(\mathbf{x}(\tau)) \cdot \dot{\mathbf{x}}(\tau) \, d\tau.
\]

(We use \( d\mathbf{s}_i \) to denote the line element of the trajectory \( \mathbf{x}_i(t) \), and so the second equality is just the definition of path integration.)
Proposition 1.2. The increase in total kinetic energy is equal to the total work on the system.

Proof. Summing over \( i = 1, \ldots, N \) and differentiating the kinetic energy eq. (1.4) we obtain
\[
\dot{T} = \sum_{i=1}^{N} m_i \dot{x}_i \cdot \ddot{x}_i = \sum_{i=1}^{N} \dot{x}_i \cdot \dot{F}_i = \dot{x} \cdot F
\]
by using Newton’s equations eq. (1.3). Integrating, we arrive at
\[
T(t) - T(t_0) = \int_{t_0}^{t} \dot{T} \, dt = \int_{t_0}^{t} \dot{x} \cdot F \, d\tau = \int_{x(t_0)}^{x(t)} F \cdot ds = W
\]
where \( W \) is the total work. \( \square \)

Although work is measured in the physical space \( \mathbb{R}^d \), the total work is naturally defined on configuration space \( \mathbb{R}^{Nd} \). So the change in kinetic energy can also be thought of as the work done by the tuple of forces \( F \) on the lifted path \( x(t) \) in configuration space.

For some systems there is also a potential energy. Physically, a system of particles is called conservative if the force \( F \) depends only on the positions \( x \) (not on time or velocity \( \dot{x} \)) and if the total work along any path connecting two points \( y, z \) in configuration space,

\[
W = \int_{y}^{z} F(s) \cdot ds,
\]  

is independent of the path. The path here is arbitrary, and is not limited to trajectories. This is equivalent to the work around any simple closed path vanishing, since two paths with the same endpoints can be concatenated to form a closed path.

Example 1.3. If the interaction forces depend only on particle distances,
\[
F_i = \sum_{j=1}^{N} F_{ij}, \quad F_{ij} = f_{ij}(r_{ij}) e_{ij}, \quad r_{ij} = |x_i - x_j|, \quad e_{ij} = \frac{x_i - x_j}{r_{ij}},
\]
then the system is conservative, with potential energy \( U = \sum_{i<j} U_{ij}, \ U_{ij} = \int f_{ij} \, dr \).

Let’s pause, fix \( N - 1 \) of the particles, and consider only wiggling the \( i \)th particle. Let \( S \) be a smooth 2-dimensional submanifold of \( \mathbb{R}^d \) with boundary. Using Stokes’ theorem, we can rewrite
\[
0 = \int_{\partial S} F_i \cdot ds_i = \int_S dF_i
\]
1.2. ENERGY

where \( dF_i \) is the exterior derivative of \( F_i \) when viewed as a 1-form in the \( i \)th coordinate on \( \mathbb{R}^d \) (for \( n = 3 \), \( dF_i \) is the curl \( \nabla_i \times F_i \) dotted with the unit normal vector field). For this to hold for all such surfaces \( S \), then we must have \( dF_i = 0 \) (\( \nabla_i \times F_i = 0 \) for \( n = 3 \)). So the physical definition of a conservative force is an integral formulation of \( F_i \) being closed when viewed as a 1-form on \( \mathbb{R}^d \), which avoids the issue of \( dF_i \) not being defined at \( \bigcup \{ x_j \} \). From differential geometry, we expect that \( F_i \) being closed should imply that \( F_i \) is exact (\( F_i = -\nabla U(x_i) \)) on \( \mathbb{R}^d \), although the 0-form antiderivative may not be defined on \( \bigcup \{ x_j \} \).

**Theorem 1.4.** A system is conservative if and only if there exists a potential energy, a smooth function \( U : \mathbb{R}^{Nd} \setminus \Delta \rightarrow \mathbb{R} \) such that \( F_i = -\nabla U \).

**Proof.** First suppose that the work for all of the forces \( F_i \) does not depend on the path. Then the line integral

\[
U(x) = -\int_{x_0}^{x} F(s) \cdot ds
\]

is well defined as a function for a fixed choice of \( x_0 \in \mathbb{R}^{Nd} \setminus \Delta \), and we have \( F = -\nabla \sum_i U_i(x_i) \).

Conversely, if \( F \) is a conservative force with potential energy \( U \), then the fundamental theorem of calculus tells us that

\[
\int_{x_0}^{x} F \cdot ds = -U(x) + U(x_0).
\]

That is, the work is independent of the path. \( \square \)

**Remark.** In Theorem 1.4 it is assumed that \( F \) and \( U \) are defined on all of \( \mathbb{R}^{Nd} \setminus \Delta \). The statement does not hold true in general on just an open subset of configuration space \( \mathbb{R}^{Nd} \).

**Example 1.5.** The vector field

\[
F(x, y) = \left(\frac{y}{x^2 + y^2}, -\frac{x}{x^2 + y^2}\right)
\]

is defined on \( \mathbb{R}^2 \setminus \{0\} \), and on this set \( F \) may be written as the negative gradient of the planar polar coordinate angle. However, the work done on a particle traveling once clockwise about the unit circle is \( 2\pi \), which corresponds to the fact that we cannot define an angular coordinate continuously on all of \( \mathbb{R}^2 \setminus \{0\} \). Consequently, if we consider a particle subject to this force we have an example of a nonconservative force.

It is often helpful for physical intuition to picture a small ball rolling down the graph of \( U(x) \). Suppose we have a solution \( x(t) \) to eq. (1.8) with \( E(x(t)) = E_0 \). Since kinetic energy is nonnegative, then a ball at position \( U(x(t)) \) is confined to the region where \( U(x) \leq E_0 \), which is often called a potential well. The smaller
the potential energy the greater the kinetic energy and hence the greater the velocity; this tells us that as the pebble gains velocity as it rolls into the well. This picture makes some facts very intuitive, like that local minima and maxima of $U(x)$ are stable and unstable equilibria for the system, respectively. For a bounded potential well in a conservative system, the ball rolls right through any minima and up towards the boundary $U^{-1}(E_0)$.

Now that we have two notions of mechanical energy—kinetic and potential—we may consider their sum $E = T + U$, called the total energy.

**Proposition 1.6** (Conservation of energy). The total energy $E = T + U$ of a conservative system is preserved under the motion: $E(t) = E(t_0)$.

**Proof.** As in Proposition 1.2 we differentiate:

$$\frac{d}{dt} E = \dot{T} + \nabla U \cdot \dot{x} = \dot{x} \cdot F - F \cdot \dot{x} = 0.$$

\[\square\]

### 1.3 Conservative systems

In section 1.2 we saw that for physically conservative systems the total mechanical energy is constant along trajectories. We will now give this phenomenon a name, and examine the mathematical consequences of this seemingly simple observation. Exercise 1.6 provides an example where the conservative quantity is not an energy, for which the results of this section are more surprising.

Recall from section 1.1 that we can always reduce a system of ODEs to a first-order system. Suppose we have a system of the form

$$\dot{x} = f(x) \quad (1.8)$$

which is conservative:

**Definition 1.7.** The system of ODEs eq. (1.8) is called conservative if there is a smooth function $E(x)$ (that is nonconstant on open sets) such that $\frac{d}{dt}(E(x(t))) = 0$ for any solution $x(t)$ to eq. (1.8).

Note that geometrically this requires that trajectories $x(t)$ lie in level sets of $E(x, \dot{x})$ in phase space.

Any point $x_*$ where $f(x_*)$ is called a fixed point or equilibrium of eq. (1.8), since this implies that the constant solution $x(t) \equiv x_*$ is a solution. Around any point that is not an equilibrium, the implicit function theorem guarantees a neighborhood on which the level set through that point is the graph of a smooth function, and so non-equilibrium trajectories trace out smooth curves in phase space. Moreover, if $U(x) \neq E$ for all $x$ then the level set cannot contain an equilibrium and must be smooth everywhere.

A fixed point is attracting if $x(t) \to x_*$ as $t \to \infty$ for all initial conditions sufficiently close to $x_*$. 

**Proposition 1.8.** A conservative system cannot have any attracting (or repulsive) fixed points.
1.3. CONSERVATIVE SYSTEMS

Proof. Suppose $x^*$ were an attracting fixed point. Then all points in an open ball around $x^*$ would have to be at the same energy $E(x^*)$ since energy is constant on trajectories and all trajectories in the ball flow to $x^*$. But then $E(x)$ would be constant on an open set, which contradicts our definition. Changing $t \mapsto -t$ yields the analogous argument for repulsive fixed points.

To examine the stability of a fixed point, we would like to Taylor expand about $x^*$ to replace (1.8) with the linear system

$$\frac{d}{dt}(x - x^*) = \dot{x} \approx f(x^*) + f'(x^*) \cdot (x - x^*) = f'(x^*) \cdot (x - x^*), \quad (1.9)$$

where $f'(x^*)$ is the Jacobian matrix of $f$ evaluated at $x^*$. From ODE theory, it turns out that this first term is sufficiently dominant to determine the stability of $x^*$ if $f'(x^*)$ has at least one eigenvalue with nonzero real part. This fact is sharp, as the following example illustrates:

**Example 1.9.** Consider the system

$$\begin{align*}
\dot{x} &= -y + ax(x^2 + y^2) \\
\dot{y} &= x + ay(x^2 + y^2)
\end{align*} \quad \iff \begin{align*}
\dot{r} &= ar^3 \\
\dot{\theta} &= 1
\end{align*}$$

on $\mathbb{R}^2$, where $a \in \mathbb{R}$ is a constant. When $a = 0$ we obtain the linearized system at $(x^*, y^*) = (0, 0)$, which has eigenvalues $\pm i$ and predicts that the origin is surrounded by periodic solutions or closed orbits—such a fixed point is called a center. However, when $a < 0$ ($a > 0$) we see that $r(t)$ is decreasing (increasing) monotonically and so the origin becomes a stable (unstable) spiral.

Centers are delicate, in the sense that trajectories nearby need to perfectly match up after one revolution, and the neglected terms in the Taylor expansion eq. (1.9) can push them inwards or outwards. This example illustrates that the linearized system predicting a center is in general unilluminating as to the stability of the fixed point. However, a conserved quantity is enough to recover the prediction:

**Theorem 1.10 (Nonlinear centers for conservative systems).** Suppose $x_*$ is an isolated fixed point of eq. (1.8) and that $E(x)$ is a conservative quantity. If $x_*$ is a local minimum (or maximum) of $E$, then all trajectories sufficiently close to $x_*$ are closed.

Sketch: Since $E$ is constant on trajectories, each trajectory is contained in some contour of $E$. Near a local maximum or minimum the contours are closed. The only remaining question is whether the trajectory actually goes all the way around the contour or whether it stops at a fixed point on the contour. But because $x_*$ is an isolated fixed point, there cannot be any fixed points on contours sufficiently close to $x_*$. Hence all trajectories in a sufficiently small neighborhood of $x_*$ are closed orbits, and therefore $x_*$ is a center.

**Remark.** We need to assume that $x_*$ is isolated, otherwise there could actually be fixed points on the energy contour—see Exercise [1.7].
Example 1.11 (Harmonic oscillator). For the one-dimensional system \( m\ddot{x} = -kx \), we have \( U(x) = \frac{1}{2}kx^2 \) and \( E(x,p) = \frac{1}{2}(\frac{1}{m}p^2 + kx^2) \). The trajectories are periodic and trace out the elliptic level sets of \( E \) in phase space \( \mathbb{R}_x \times \mathbb{R}_p \). This system has the explicit solution

\[
x(t) = x_0 \cos \left( \sqrt{\frac{k}{m}} t \right) + \frac{p_0}{\sqrt{km}} \sin \left( \sqrt{\frac{k}{m}} t \right),
\]

\[
p(t) = -\sqrt{km}x_0 \sin \left( \sqrt{\frac{k}{m}} t \right) + p_0 \cos \left( \sqrt{\frac{k}{m}} t \right),
\]

which are parametric equations for an axes-parallel ellipse centered at the origin.

1.4 Nonconservative systems

From experience we know that in practice systems are rarely conservative—if we were to test an ideally conservative system, dissipative forces like kinetic friction would dampen mechanical motion and prevent perpetual motion. In this case we would now have \( \frac{d}{dt}E \leq 0 \), which has new consequences for solutions.

Definition 1.12. Consider the two-dimensional ODE system (1.8) with a fixed point \( x_\ast \). Suppose there exists a continuous function \( E(x) \) on a connected neighborhood \( U \) of the fixed point that is smooth on \( U \setminus \{x_\ast\} \). \( x_\ast \) is a strict global minimum with value zero, and for all \( c \) sufficiently small the level sets \( E(x) = c \) within \( U \) form simple closed curves. If also:

(a) \( \frac{d}{dt}E(x) \leq 0 \) for all \( x \) in \( U \setminus \{x_\ast\} \) then \( E \) is a weak Liapunov function;

(b) \( \frac{d}{dt}E(x) < 0 \) for all \( x \) in \( U \setminus \{x_\ast\} \) then \( E \) is a strong Liapunov function.

For our image of the ball rolling down the graph of the potential energy, the surface of the graph is now slightly sticky. The ball may still roll through a minimum, but does not have enough energy to approach the boundary \( U^{-1}(E_0) \) again and so the permitted region for the ball continually shrinks.

Example 1.13 (Damped harmonic oscillator). Consider the one dimensional system \( m\ddot{x} = -b\dot{x} - kx \) with \( b > 0 \) a positive damping constant. The total energy is still \( E = \frac{1}{2}(m\dot{x}^2 + kx^2) \), but now

\[
\frac{d}{dt}E = m\ddot{x} + kx\dot{x} = -b\dot{x}^2 \leq 0
\]

and so total energy \( E \) is a weak (but not strong) Liapunov function. The origin is globally attracting with three qualitatively different phase portraits:

- \( 0 < b < 2\sqrt{km} \) (under damped): the origin is a stable spiral and the system oscillates infinitely many times with exponentially decaying amplitude;
1.4. NONCONSERVATIVE SYSTEMS

- \( b = 2\sqrt{km} \) (critically damped): the origin is a stable degenerate node and the system moves toward the origin as quickly as possible; and

- \( b > 2\sqrt{km} \) (over damped): the origin is a stable node and the system returns to the origin without oscillating.

**Theorem 1.14.** Consider the smooth two-dimensional ODE system (1.8) with a fixed point \( \mathbf{x}_* \).

(a) If there exists a weak Liapunov function near the fixed point \( \mathbf{x}_* \) then the fixed point is Liapunov stable: for all \( \epsilon > 0 \) there exists \( \delta > 0 \) such that \( \| \mathbf{x}(t_0) - \mathbf{x}_* \| < \delta \) implies \( \| \mathbf{x}(t) - \mathbf{x}_* \| < \epsilon \) for all \( t \geq t_0 \).

(b) If there exists a strong Liapunov function near the fixed point \( \mathbf{x}_* \) then \( \mathbf{x}_* \) is also globally asymptotically stable: for all initial conditions the corresponding solutions \( \mathbf{x}(t) \rightarrow \mathbf{x}_* \) as \( t \rightarrow \infty \).

In particular, there can be no periodic solutions, unlike conservative systems. For a proof and more details, see [JS07, Ch. 10].

Now let us examine the behavior of solutions when the friction term dominates. For the damped harmonic oscillator in Example 1.13, this is the limit \( b \gg mk \) (i.e. \( mk/b^2 \rightarrow 0 \))—see Exercise 1.8 for details. In this limit, the \( \ddot{x} \) term is negligible and we may be left with something of the form:

**Definition 1.15.** The ODE system (1.8) is a gradient system if it is of the form \( \dot{x} = -\nabla U(x) \), for some smooth function \( U(x) \) (which for mechanical systems may not be the same as the potential energy).

**Remark.** Although conservative systems \( \ddot{x} = -\nabla U(x) \) and gradient systems \( \dot{x} = -\nabla U(x) \) look similar, they display entirely opposite behavior.

From the definition, we see that the vector field \( -\nabla U(x) \) in configuration space which is tangent to trajectories always points in the direction of steepest descent for \( U(x) \), and hence is orthogonal to level sets (rather than parallel as with conservative systems). For our image of the ball rolling down the graph of the potential energy, the ball now slows and never reaches the first minimum it encounters, as if the potential energy graph was a water tank. Closed orbits are of course impossible again (see Exercise 1.9).

**Example 1.16.** In the over damping limit for the harmonic oscillator we are left with \( \dot{x} = -kx/b \), and so \( U(x) = kx^2/2b \). This is now a first-order system with solution \( x(t) = x_0e^{-kt/b} \), which is the limiting (slow time scale) behavior for the over damped oscillator after the transient (fast time scale) behavior is negligible. For the phase portrait we can take \( p = m\dot{x} = -mkx/b \), but we can no longer take a second arbitrary initial condition \( p(0) \).
1.5 Reversible systems

Other than conserved quantities, many mechanical systems obey another symmetry: a symmetry in time. A mechanical system of the form

\[ m\ddot{x} = F(x) \]  

(1.10)

that is independent of time and velocity (or if \( F \) is even in velocity in time) is invariant under the change of variables \( t \mapsto -t \), since \( \ddot{x} \) picks up two factors of \(-1\). Because \( p \mapsto -p \) under this change of variables, phase space is reflected about the position axes \( \{(x, p) : p = 0\} \) and the arrows on trajectories are reversed. The resulting trajectories are still trajectories for the original system since the equation is unchanged, and so the phase portrait must be symmetric about the position axes with arrows reverse.

We will now give this phase space symmetry a name and study its mathematical consequences. Exercise\[1.10\] provides a two-dimensional example where the symmetry is not reflection across the \( x \)-axis.

**Definition 1.17.** The two-dimensional ODE system (1.8) is **reversible** if it is invariant under the change of variables \( t \mapsto -t, \ x \mapsto R(x) \) for some smooth involution \( R \) (i.e. \( R(R(x)) = x \)).

Involutions include reflections across a hyperplane and rotations about an axis by \( \pi \).

Recall from section\[1.2\] the difficulty of classifying an equilibrium as a center.

**Theorem 1.18** (Nonlinear centers for reversible systems). Suppose that the smooth two-dimensional ODE system (1.8) is reversible, and that \( x_* \) is a fixed point both for the system \( (f(x_*) = 0) \) and for \( R(R(x_*) = x_*) \). If the linearized system about \( x_* \) has \( x_* \) as a center, then all trajectories sufficiently close to \( x_* \) are indeed closed.

**Sketch:** After a change of variables, we may assume that \( x_* = 0 \) and that \( R \) is a linear reflection. Consider a trajectory that starts on the hyperplane fixed by \( R \). Sufficiently near the origin, the flow swirls around the origin thanks to the dominant influence of the linear center, and so the trajectory eventually intersects the hyperplane again. By reversibility, we can reflect this trajectory to obtain a twin trajectory with the same endpoints but with its arrow reversed. Together the two trajectories form a closed orbit, as desired. Hence all trajectories sufficiently close to the origin are closed.

This argument can also be used to show the existence of lone closed or homoclinic orbits, once we know that the trajectory starts and returns to the hyperplane of symmetry.
1.6 Linear momentum

Suppose the force on the $i$th particle can be decomposed into

$$ F_i = \sum_{j=1, j \neq i}^{n} F_{ij} + F'_i, $$

(1.11)

where $F_{ij}$ is the interaction force between the $i$th and $j$th particle, and $F'_i$ is the external force on the $i$th particle. A system is called closed if there are no external forces. Further assume that the interaction forces obey the law of action and reaction or Newton’s third law—another experimental observation that the force two particles exert on each other are equal and opposite: $F_{ij} = -F_{ji}$. Not all interaction forces obey this law, but we will always implicitly assume that they do. If we were also working in an inertial frame then the interaction forces would also be collinear: $F_{ij} = f_{ij}e_{ij}$, where $e_{ij}$ is the unit vector $(x_j - x_i)/|x_j - x_i|$ from the $i$th to the $j$th particle.

**Example 1.19.** Any system of the form in Example 1.3, which includes the motion of gravitational bodies, obeys the law of action and reaction since $e_{ij} = -e_{ji}$. Interaction forces for non-mechanical systems may not obey this law. For example, a particle with electric charge $q$ placed in an electromagnetic field is acted upon by the Lorentz force

$$ F = q \left[ E + \frac{1}{c} (v \times H) \right], $$

(1.12)

where $E$, $H$ are the strengths of the electric and magnetic fields (they satisfy the Maxwell system of equations) and $c$ is the speed of light.

The effect of all external forces together can be observed through the total (linear) momentum,

$$ \mathbf{P} = \sum_{i=1}^{N} \mathbf{p}_i. $$

(1.13)

**Proposition 1.20 (Conservation of linear momentum).** The change in the total momentum is equal to the total force $\sum_i F_i$, which for the decomposition (1.11) is the total external force. In particular, for a closed system the total linear momentum is conserved.

**Proof.**

$$ \mathbf{P} = \sum_{i=1}^{n} \dot{\mathbf{p}}_i = \sum_{i=1}^{n} \mathbf{F}_i = \sum_{i,j=1, i \neq j}^{n} F_{ij} + \sum_{i=1}^{n} F'_i = \sum_{i=1}^{n} F'_i $$

In the last equality, we note that $F_{ij} = -F_{ji}$ causes the double sum over interaction forces to cancel pairwise. $\square$
By taking a dot product in the previous proposition, we see that sometimes the individual components of the total momentum can still be conserved:

**Corollary 1.21.** If the total external force is perpendicular to an axis, then the projection of the total momentum onto that axis is conserved.

The system center of mass (sometimes barycenter) is

$$X = \frac{\sum_{i=1}^{N} m_i x_i}{\sum_{i=1}^{N} m_i}. \quad (1.14)$$

This does not depend on the choice or origin, and it is special because the total momentum relative to this point vanishes (exercise Exercise 1.11). Moreover, the total momentum is mathematically the same as that of a particle of mass $\sum_{i=1}^{N} m_i$ lying at the center of mass:

$$P = \sum_{i=1}^{N} p_i = M \dot{X}, \quad M = \sum_{i=1}^{N} m_i. \quad (1.15)$$

**Proposition 1.22** (Newton's first law). The center of mass moves as if all masses were concentrated at it and all forces were applied to it. In particular, the center of mass of a closed system moves uniformly and rectilinearly.

**Proof.** Differentiating eq. (1.15) yields

$$M \ddot{X} = \dot{P} = \sum_{i=1}^{n} F_i.$$

The right-hand side vanishes for a closed system. \qed

1.7 Angular momentum

Now that we have discussed linear momentum, we would like to consider angular momentum, for which we will have to specialize to $n = 3$ for this section so that we may use the cross product. The angular momentum (about the origin) of a particle is

$$L_i = x_i \times p_i = m_i x_i \times \dot{x}_i, \quad (1.16)$$

and the torque (sometimes moment of force) is

$$N_i = x_i \times F_i. \quad (1.17)$$

Since the force can be divided into interaction and external forces, we also have the external torque $N'_i = x_i \times F'_i$. The relationship between angular momentum and external torque is analogous to before:
Proposition 1.23 (Conservation of angular momentum). The change in total angular momentum is equal to the total torque, which for the decomposition (1.11) is the total external torque. In particular, for a closed system the total angular momentum is conserved.

Proof. Using Newton’s equations (1.2) we have

$$\dot{L}_i = \frac{d}{dt}(x_i \times p_i) = x_i \times \dot{p}_i + \dot{x}_i \times p_i = N_i + 0.$$ 

Note that the product $\dot{x}_i \times p_i = 0$ vanishes because $\dot{x}_i$ is parallel to $p_i$. Summing over $i = 1, \ldots, N$ yields $\dot{L} = N$. As in the proof of Proposition 1.20 the total torque is equal to the total external torque,

$$\dot{L} = \sum_{i=1}^{n} x_i \times F_i = \sum_{i,j=1, i \neq j}^{n} x_i \times F_{ij} + \sum_{i=1}^{n} x_i \times F'_i = \sum_{i=1}^{n} N'_i.$$ 

This is because the pairwise vanishing guaranteed by the law of action and reaction,

$$x_i \times F_{ij} + x_j \times F_{ji} = (x_i - x_j) \times F_{ij} = 0.$$ 

In particular, for a closed system the external torque vanishes and we have $\dot{L} = 0$. \[\square\]

Taking a dot product we also obtain component-wise conservation:

Corollary 1.24. If the total external torque is perpendicular to an axis, then the projection of the total angular momentum onto that axis is conserved.

As with linear momentum, from outside the system the momentum evolves as if it were all concentrated about the center of mass and all the external torques were applied to it.

Proposition 1.25. The total angular momentum and total torque about the center of mass is the sum of that of and about the center of mass.

Proof. We expand about the center of mass:

$$L = \sum_{i=1}^{N} [X + (x_i - X)] \times m_i [\dot{X} + (\dot{x}_i - \dot{X})]$$

$$= X \times M\dot{X} + \left[ \sum_{i=1}^{N} m_i (x_i - X) \right] \times \dot{X}$$

$$+ X \times \left[ \sum_{i=1}^{N} m_i (\dot{x}_i - \dot{X}) \right] + \sum_{i=1}^{N} (x_i - X) \times m_i (\dot{x}_i - \dot{X})$$

$$= X \times P + 0 + 0 + \sum_{i=1}^{N} (x_i - X) \times m_i (\dot{x}_i - \dot{X})$$

where we used eq. (1.15) for the first term and Exercise 1.11 for the vanishing of the square-bracketed terms. Since $L = N$, the second statement follows. \[\square\]
There is one last dynamical quantity of interest, the moment of inertia of the $i$th particle about the origin,

$$I_i = m_i |\mathbf{x}_i|^2.$$  \hfill (1.18)

The moment of inertia plays the role for angular velocity, \( \omega_i = \frac{\mathbf{x}_i \times \mathbf{\dot{x}}_i}{|\mathbf{x}_i|^2}, \) analogous to mass for linear velocity. This is because

$$L_i = \mathbf{x}_i \times p_i = \mathbf{x}_i \times (m_i \omega_i \times \mathbf{x}_i) = m_i \left[ (\mathbf{x}_i \cdot \mathbf{x}_i) \omega_i - (\mathbf{x}_i \cdot \omega_i) \mathbf{x}_i \right] = I_i \omega_i + 0,$$ \hfill (1.20)

since $\mathbf{\dot{x}}_i = \omega_i \times \mathbf{x}_i$. As in the previous proposition, the formula $L = I \omega$ can also be broken into the sum of that of and about the center of mass.

**Proposition 1.26.** The total moment of inertia evolves according to \( \ddot{I} = 4T + 2F \cdot \mathbf{x} \). For a conservative system with a homogeneous potential of degree \( k \) (i.e. $U(\mathbf{x}) = c|\mathbf{x}|^k$) we have $\ddot{I} = 4E - 2(k + 2)U$.

**Proof.** The first statement is a straightforward calculation. For the second, we note that for a homogeneous potential we have

$$F \cdot \mathbf{x} = -\nabla U \cdot \mathbf{x} = -kU, \quad 4T = 4E - 4U.$$ \hfill \( \square \)

### 1.8 Exercises

**1.1 (Pendulum).** Consider a mass $m$ attached to the end of a rigid massless rod of length $\ell$ with the other end suspended at a fixed point. We allow the rod to rotate in a vertical plane, subject to a constant downward gravitational acceleration $g$.

(a) Let $x$ denote the angle from the vertical directly below the pivot, and show that

$$\ddot{x} = -\frac{g}{\ell} \sin x.$$  

(We have not allowed for angular coordinates yet, so at the moment we are cheating and taking $x \in \mathbb{R}$.)

(b) Sketch the potential energy and the phase portrait, and convince yourself how the trajectories correspond to a small ball rolling down the graph of the potential. Note that near the origin in the phase plane the diagram looks like that of the harmonic oscillator (cf. Example 1.11), which is a consequence of the small angle approximation $\sin x \approx x$. Identify the equilibria (constant solutions) and the separatrix (the eye-shaped boundary separating different modes of behavior). How many trajectories make up the separatrix from $-\pi \leq x \leq \pi$, and to what motion do they correspond?
1.8. EXERCISES

(c) Now add a damping term:

\[ \ddot{x} = -b\dot{x} - \frac{g}{\ell} \sin x \]

where \( b > 0 \). Sketch the new phase portrait, and show that \( \dot{E} \) decreases monotonically along all trajectories that are not equilibria.

1.2 (A way to compute \( \pi \) [Gal03]). In this example we will see a geometric aspect of phase space appear as a physically measurable quantity, which reinforces that phase space is inherent to a mechanical system and not merely abstract. Consider a frictionless horizontal ray with a vertical wall at the origin. One small block of mass \( m \) is initially at rest on the surface, and a big block of mass \( M \gg m \) is pushed towards the small block so that the small block is sandwiched between the large block and the wall. We will count the number \( N \) of collisions the small block makes with the big block and the wall.

(a) Let \( v_1 \) and \( v_2 \) denote the velocities of the large and small blocks respectively, and consider the convenient rescaling \( y_1 = \sqrt{M}v_1, \ y_2 = \sqrt{m}v_2 \).

Plot the initial energy level set in the \((y_1,y_2)\)-plane, to which the motion is confined since the surface is frictionless.

(b) Initially we have \( v_1 < 0, \ v_2 = 0 \)—plot this point in the same \((y_1,y_2)\)-plane. Assume that the collision is purely elastic, so that when the blocks collide the total momentum is conserved. Plot the total momentum level set which contains the initial point—the outcome velocities is determined by the other intersection of the level sets. After the first collision the small block will hit the wall, and we will assume that this collision is elastic so that \( v_2 < 0 \) is replaced by \( -v_2 \)—plot this new point as well. Plot a few more iterates of this two-collision pattern in the \((y_1,y_2)\)-plane.

(c) The pattern repeats until \( v_1 > 0 \) and \( 0 < v_2 < v_1 \), so that the large block is moving away, the small block will not collide with the wall again, and the small block cannot catch up. Sketch the end zone region in the \((y_1,y_2)\)-plane as well.

(d) Connect consecutive points occupied by the system in the \((y_1,y_2)\)-plane. Since the lines are either vertical or parallel, the angle \( \theta \) any point between any two consecutive lines is the same. Show that \( \theta = \arctan \sqrt{m/M} \).

(e) For any point, the angle \( \theta \) at that point subtends an arc on the circle opposite that point. Show that the angle at the origin which subtends the same arc is \( 2\theta \), and that the total number \( N \) of collisions that occur is the maximal integer such that \( N\theta < \pi \).

(f) Take \( M = 100^n m \) for \( n \) a positive integer. Show that \( 10^{-n} \cdot N \to \pi \) as \( n \to \infty \), and so the number of collisions \( N \) begins to spell out the digits of \( \pi \) for \( n \) large. In fact, \( N \) will be a number with \( n + 1 \) digits whose first \( n \) digits coincide with the first \( n \) decimal digits of the number \( \pi \) (starting with 3).
1.3 (Galilean group generators). Show that every Galilean transformation $g$ of the space $\mathbb{R}^d \times \mathbb{R}$ can be written uniquely as the composition $g_1 \circ g_2 \circ g_3$ of the three types of Galilean transformations in Example 1.1. Start by writing $g$ as a general affine transformation on $\mathbb{R}^d \times \mathbb{R}$,

$$g(x, t) = (Ax + vt + x_0, b \cdot x + kt + t_0),$$

and show that $b = 0$, $k = 1$, and $A \in O(d)$. What is the dimension of the Galilean group for $d = 3$?

1.4. Suppose a mechanical system of $N$ particles in $\mathbb{R}^d$ is in some inertial frame and all of the initial velocities are zero. Show that the particles always remain in the $(d-1)$-dimensional linear subspace of $\mathbb{R}^d$ in which they were initially contained.

1.5 (Rotating reference frame). Suppose we have a system in an inertial coordinate system $z \in \mathbb{R}^3$ (like coordinates relative to a stationary sun), so that Newton’s equations (1.3) obey the conditions (1)–(3) of section 1.1. Consider a non-inertial coordinates $x$ (like coordinates relative to a fixed point on Earth’s surface) defined by

$$t \mapsto t, \quad x = B(t)z + b(t),$$

where $b(t) \in \mathbb{R}^3$ is the new origin and $B(t) \in O(3)$ for all $t$.

(a) Show that the equations of motion in the new frame are

$$m_i \ddot{z}_i = F_i \left( z_k - z_j, B^{-1} \dot{B}(z_k - z_j) + (\dot{z}_k - \dot{z}_j) \right) + \Phi_i + \Psi_i,$$

where

$$\Phi_i = -m_i \left( B^{-1} \dot{B}z_i + B^{-1} \dot{b} \right), \quad \Psi_i = -2m_i B^{-1} \dot{B} \dot{z}_i.$$

The new forces $\Phi$ and $\Psi$ that appear on the right-hand side of the equations of motion for $z$ are called inertial or fictitious forces.

(b) Differentiate the definition of an orthogonal matrix to show that $B^{-1} \dot{B}$ is skew-symmetric, and write $B^{-1} \dot{B}z = \omega \times z$ for some vector $\omega$ called the angular velocity of the moving frame. Now

$$\Psi_i = -2m_i \omega \times \dot{z}_i.$$

$\Psi_i$ is called the Coriolis force, and depends on the velocity. In the northern hemisphere of the earth it deflects every body moving along the earth to the right and every falling body eastward.

(c) Use the product rule for $(B^{-1} \dot{B})^\top$ and the skew-symmetry of $B^{-1} \dot{B}$ to show that

$$B^{-1} \dot{B} = \left( B^{-1} \dot{B} \right)^\top + B^{-1} \dot{B} B^{-1} \dot{B}.$$
1.8. EXERCISES

Writing $\mathbf{w} = B^{-1} \mathbf{b}$ for the \textit{acceleration} and $\alpha = \dot{\omega}$ for the \textit{angular acceleration} of the moving frame, we obtain

$$\Phi_i = -m_i [\mathbf{w} + \omega \times (\omega \times \mathbf{z}_i) + \alpha \times \mathbf{z}_i].$$

The second term is called the \textit{centrifugal force} and the third term is the \textit{inertial rotation force} or the \textit{Euler force}, both of which only depend on position. The former is always directed outward from the instantaneous axis of rotation and acts even on a body at rest in the coordinate system \(z\). The latter is only present for nonuniform rotation.

1.6 (A non-mechanical conservative system). In dimensionless form, the Lotka–Volterra predator-prey model takes the form

$$\dot{x} = x(1 - y), \quad \dot{y} = \mu y(x - 1),$$

where $\mu > 0$ is a constant.

(a) Use the chain rule to find a differential equation for $\frac{dy}{dx}$.

(b) Find a conserved quantity $E(x, y)$ for this system by integrating this separable differential equation.

(c) Show that all trajectories are periodic for initial conditions $x_0, y_0 > 0$.

1.7 (Minimum of a conservative system that is not a center). Consider the system

$$\dot{x} = xy, \quad \dot{y} = -x^2.$$  

Show that $E(x, y) = x^2 + y^2$ is a conserved quantity and plot the phase portrait for this system. Although the origin is a minimum for $E$, it is not an isolated fixed point nor a center.

1.8 (Harmonic oscillator in the over damped limit). This is an example of the “over damped” limit yielding a gradient system, as advertised in section 1.4. The objective is to find what limit we need to take for the damped harmonic oscillator:

$$m\ddot{x} = -b\dot{x} - kx, \quad b, k > 0,$$

that justifies neglecting the $\ddot{x}$ term.

(a) By equating the units of all terms, determine the dimensions of $b$ and $k$.

(b) Define a new dimensionless variable $\tau$ via $t = T\tau$ where $T$ is a constant with units time to be chosen. Find the new differential equation in terms of $\tau$. Divide the equation by the coefficient of the $x$ term, pick $T$ to cancel the coefficient of the $\dot{x}$ term, and check that $T$ has units time.

(c) The coefficient of the $\ddot{x}$ term should be $mk/b^2$, and so the limit in which this term is negligible is $\epsilon := mk/b^2 \to 0$. Find the general solution $x(t) = c_1 e^{k_1 t} + c_2 e^{k_2 t}$ for the linear equation $\epsilon \ddot{x} + \dot{x} + x = 0$. 


(d) Recall that $1/|k|$ is called the characteristic time of $e^{kt}$, since after a time $1/|k|$ the function has decreased (since $k < 0$) by a factor of $1/e$. Find the leading term in the Taylor expansion of $1/k_1$ and $1/k_2$ about $\epsilon = 0$—these are called the fast and slow time scales for the solution.

1.9. Show that nonconstant periodic solutions are impossible in a gradient system, by considering the change in $U(x)$ around such an orbit. In particular, since we can always find an antiderivative in one dimension, this shows that any one-dimensional first order ODE has no periodic solutions.

1.10 (A non-mechanical reversible system). Show that the system

$$
\begin{align*}
\dot{x} &= -2 \cos x - \cos y, \\
\dot{y} &= -2 \cos y - \cos x
\end{align*}
$$

is reversible, by sketching the phase portrait and finding a symmetry. Note that the presence of stable and unstable nodes guarantees that this system is not conservative.

1.11. (a) Show that the center of mass (1.14) does not depend on the choice of origin.

(b) Show that both the “total position” and total momentum relative to the center of mass vanishes:

$$
\sum_{i=1}^{N} m_i (x_i - X) = 0, \quad \sum_{i=1}^{N} m_i (\dot{x}_i - \dot{X}) = 0.
$$

1.12 (Book stacking [Nah16]). We will stack books of mass 1 and length 1 on a table in an effort to produce the maximum amount of overhang.

(a) Place the first book with its left edge at $x = 0$ and its right edge lined up with the end of the table at $x = 1$. By considering the center of mass of the book, determine the distance $S(1)$ we can slide the book over the edge of the table before it falls.

(b) Starting with a stack of two books instead of one, we can reason as in part (a) and slide the top book forward a distance of $S(1)$ while keeping the bottom book stationary. By considering the center of mass of the two books, determine the distance $S(2)$ we can slide this two-book configuration before it falls.

(c) Now start with three books, slide the top one a distance of $S(1)$ and then the top two books as in part (b) in order to produce an overhang $S(2)$ from the edge of the bottom book. Determine the distance $S(3)$ we can slide the three-book configuration before it falls. Note that when multiplied by 2, the sequence $S(1), S(2), S(3), \ldots$ begins to spell out a familiar series.

(d) Postulate a formula for $S(n)$ and prove it by induction. Note that the overhang $S(n)$ tends to infinity as $n \to \infty$. 
1.13. Consider a conservative system in $\mathbb{R}^3$ with coordinates so that the center of mass $\mathbf{X} = 0$ is the origin and is at rest, $\dot{\mathbf{X}} = 0$.

(a) Show that $K^2 \leq 2IT$.

(b) Show that the trajectories in configuration space must be contained in the intersection of the hyperplane $\{ \mathbf{x} \in \mathbb{R}^{3N} : \sum_{i=1}^{N} m_i \mathbf{x}_i = 0 \}$ with the set $\{ \mathbf{x} \in \mathbb{R}^{3N} : U(\mathbf{x}) + |\mathbf{L}|^2/(2I(\mathbf{x})) \leq E \}$, where $\mathbf{L}$ and $E$ are constants of the motion.
Chapter 2

Examples

In this section we will examine some examples to see how these tools are applied and to show their purpose and power. These represent only a small portion of solved problems and are meant to provide a baseline intuition; a thorough study of classical mechanics should include many more, including a study of rigid bodies and spinning tops. The examples presented here are adapted from [Arn89, Ch. 2], [LL76, Ch. 3], and [Gol51, Ch. 3].

2.1 One-dimensional systems

Suppose we have a system with one degree of freedom of the form

\[ m \ddot{x} = F(x). \]  \hspace{1cm} (2.1)

Any such system is conservative, since we can always find an antiderivative \( U(x) \) (unique up to an additive constant) such that \( F = -\frac{dU}{dx} = -U'(x) \). From section 1.2 we know that the total energy

\[ E = \frac{1}{2} m \dot{x}^2 + U(x) \]  \hspace{1cm} (2.2)

is conserved for motions of the system. Since there is only one degree of freedom, this provides an implicit first-order equation (rather than second-order) for the motion of the system. In this section we will reap the consequences of this simple observation.

Since the kinetic energy must be nonnegative then we have \( E \geq U \) at every point in phase space. A point \((x_0, y_0)\) in phase space will then be an equilibrium if \( y_0 = 0 \) (so that \( \dot{x} = 0 \)) and if \( U'(x_0) = 0 \) (so that \( \ddot{x} = 0 \)). Solving the energy equation (2.2) for \( \dot{x} \) we obtain

\[ \frac{dx}{dt} = \sqrt{\frac{2[E - U(x)]}{m}}. \]  \hspace{1cm} (2.3)
The velocity is zero for values of $x$ with $U(x) = E$; such values are called turning points for the motion. Since the motion is confined to the region $U(x) \leq E$ (and must be connected), these turning points provide limits for the motion of the system.

For naturally occurring systems we often have $U(x) \to +\infty$ as $x \to \pm\infty$, and so there are often two turning points and the motion of the system is bounded; we will assume that this is the case. Such motion must be periodic since it is confined to one component of the energy level set and the trajectory cannot double back. Treating eq. (2.3) as a separable differential equation, we arrive at

$$t(E) = \sqrt{\frac{m}{2}} \int \frac{dx}{\sqrt{E - U(x)}} + t_0$$

for $t_0$ an integration constant. The period $\tau$ (not to be confused with the kinetic energy $T$) of bounded motion between turning points $x_1(E) < x_2(E)$ is thus the integral of $t(E)$ from $x_1$ to $x_2$ to $x_1$. This second portion would be negative, and since time should be increasing we need to insert a factor of $-1$. Therefore we arrive at two times the integral from $x_1$ to $x_2$:

$$\tau(E) = \sqrt{2m} \int_{x_1(E)}^{x_2(E)} \frac{dx}{\sqrt{E - U(x)}}. \quad (2.4)$$

This is an explicit formula for the period in terms of the energy. The period of motion also has a geometric interpretation. Let $S(E)$ denote the area enclosed by a closed trajectory in the phase plane corresponding to the energy $E$ and bounded by the turning points $(x_1,0)$ and $(x_2,0)$. As in section 1.5, the time symmetry of eq. (2.1) tells us that the trajectory in phase must be symmetric about the $x$-axis. Therefore, expressing the top half of the curve as a function $y(x) = \dot{x}(x)$ we may write

$$S(E) = 2 \int_{x_1(E)}^{x_2(E)} y(x) \, dx = 2 \int_{x_1(E)}^{x_2(E)} \sqrt{\frac{2[E - U(x)]}{m}} \, dx$$

by eq. (2.3). We can differentiate this using the Leibniz integral rule to get

$$\frac{dS}{dE} = 2 \int_{x_1(E)}^{x_2(E)} \frac{d}{dE} \sqrt{\frac{2[E - U(x)]}{m}} \, dx$$

$$+ 2 \sqrt{\frac{2[E - U(x_2)]}{m}} \frac{dx_2}{dE} - 2 \sqrt{\frac{2[E - U(x_1)]}{m}} \frac{dx_1}{dE}.$$ 

Since $x_1, x_2$ are turning points we know $U(x_1) = U(x_2) = E$, and so the second and third terms vanish:

$$\frac{dS}{dE} = 2 \int_{x_1(E)}^{x_2(E)} \frac{2}{m[E - U(x)]} \, dx = \frac{2}{m} \int_{x_1(E)}^{x_2(E)} \frac{dx}{\sqrt{E - U(x)}}.$$
2.1. ONE-DIMENSIONAL SYSTEMS

Therefore, comparing this expression to what we found for the period in eq. (2.4), we conclude

\[ \tau = m \frac{dS}{dE}. \]

(2.5)

That is, the period of motion for closed trajectories is proportional to the rate of change of the enclosed area with respect to energy.

Not only does the potential determine the period, but the converse is also partially true. Suppose now that the period of motion is experimentally known in a region bounded by \( x_1, x_2 \) where \( U(x) \) has only one local minimum, and we wish to find the potential energy \( U \) in terms of \( \tau \). For simplicity, redefine the potential \( U \) and the coordinate \( x \) such that \( U \) obtains its minimum \( U = 0 \) at \( x = 0 \). Away from this minimum, each value of \( U \) corresponds to exactly two values of \( x \); this allows us to write \( x_+ - x_- \) in the two intervals \( (x_1, 0), (0, x_2) \) respectively and manipulate the period formula (2.4) by changing variables:

\[
\tau(E) = \sqrt{2m} \left[ \int_{x_1(E)}^{x_2(E)} \frac{dx}{\sqrt{E - U(x)}} \right] + \sqrt{2m} \int_0^E \frac{dx}{\sqrt{E - U(x)}}
\]

\[
= \sqrt{2m} \left[ \int_{E}^{0} \frac{dx_-}{dU} \frac{dU}{\sqrt{E - U}} \right] + \sqrt{2m} \int_0^E \frac{dx_+}{dU} \frac{dU}{\sqrt{E - U}}
\]

\[
= \sqrt{2m} \int_0^E \left( \frac{dx_+}{dU} - \frac{dx_-}{dU} \right) \frac{dU}{\sqrt{E - U}}
\]

Dividing both sides of the equality by \( \sqrt{W - E} \) where \( W = U \) and integrating with respect to \( E \) from 0 to \( W \) yields

\[
\int_0^W \frac{\tau(E) \, dE}{\sqrt{W - E}} = \sqrt{2m} \int_0^W \left( \frac{dx_+}{dU} - \frac{dx_-}{dU} \right) \frac{dU \, dE}{\sqrt{(W - E)(E - U)}}
\]

\[
= \sqrt{2m} \int_0^W \left( \frac{dx_+}{dU} - \frac{dx_-}{dU} \right) \left[ \int_W^U \frac{dE}{\sqrt{(W - E)(E - U)}} \right] dU
\]

\[
= \sqrt{2m} \int_0^W \left( \frac{dx_+}{dU} - \frac{dx_-}{dU} \right) \left[ \tan^{-1} \left( \frac{W + U - 2E}{2\sqrt{(W - E)(E - U)}} \right) \right]_{E=W}^{E=U} dU
\]

\[
= \sqrt{2m} \int_0^W \left( \frac{dx_+}{dU} - \frac{dx_-}{dU} \right) \pi \, dU = \pi \sqrt{2m} (x_+(W) - x_-(W))
\]

where in the last step we noted \( x_+(0) = 0 = x_-(0) \), and in the second equality we switched order of integration (we assumed \( U < W \) without loss of generality, since the integral over \( E \) is symmetric). Rearranging reveals the result:

\[
x_+(U) - x_-(U) = \frac{1}{\pi \sqrt{2m}} \int_0^U \tau(E) \, dE
\]

(2.6)

That is, the period determines the potential energy up to an additive constant.
2.2 Central fields

In this section we will consider the motion of a single particle of mass \( m \) in a central field \( \mathbf{F} \). A vector field \( \mathbf{F} \) in \( \mathbb{R}^3 \) is called central (about the origin) if all of the vectors are radial and that the magnitude of the force is only a function of the radial coordinate \( r = |\mathbf{x}| \); in other words, \( \mathbf{F} = F(r)\hat{r} \). (This definition of course extends to \( \mathbb{R}^d \), but we will shortly need to restrict our attention to \( \mathbb{R}^3 \) in order to discuss angular momentum.)

A central field must be conservative, and the corresponding potential energy \( U = U(r) \) depends only on the distance from the origin. This is because if we write \( \mathbf{F} = F(r)\hat{r} \), then the work
\[
\int_{x_1}^{x_2} \mathbf{F} \cdot d\mathbf{s} = \int_{|x_1|}^{|x_2|} F(r) \, dr
\]
is path independent, and so from the conservative force Theorem [1.4] we know the force is conservative. Consequently, we can write \( \mathbf{F} = -(dU/dr)\hat{r} \) for a radial potential energy \( U(r) \).

The torque of the particle is
\[
\mathbf{N} = \mathbf{x} \times \mathbf{F} = F(r)(\hat{r} \times \hat{r}) = 0,
\]
and so by Proposition [1.23] we know the particle’s angular momentum \( \mathbf{L} \) is conserved. Therefore \( \mathbf{L} = \mathbf{x} \times \mathbf{p} \) is constant as the system evolves, and since \( \mathbf{L} \) must be perpendicular to \( \mathbf{x} \), then we know that (for \( \mathbf{L} \) nonzero) the particle’s motion must be coplanar in the plane orthogonal to \( \mathbf{L} \). When \( \mathbf{L} = 0 \), then \( \mathbf{x} \) is parallel to the velocity \( \dot{\mathbf{x}} \), and thus the particle’s motion must be collinear. In both cases, the motion is coplanar.

Using polar coordinates \((r, \phi)\) on this plane, the velocity is
\[
\dot{\mathbf{x}} = \frac{d}{dt} \begin{pmatrix} r \cos \phi \\ r \sin \phi \end{pmatrix} = \dot{r} \begin{pmatrix} \cos \phi \\ \sin \phi \end{pmatrix} + r \dot{\phi} \begin{pmatrix} -\sin \phi \\ \cos \phi \end{pmatrix} = \dot{r} \hat{r} + r \dot{\phi} \hat{\phi},
\]
and so the angular momentum is
\[
\mathbf{L} = |\mathbf{x} \times \mathbf{p}| = |(r\hat{r}) \times (m\dot{\mathbf{x}})| = mr^2 \dot{\phi}.
\]

We can also observe this conservation geometrically. Write the area sector \( \Delta S \) swept by the radius vector \( \mathbf{x} \) over an angle \( \Delta \phi \) to first order as
\[
\Delta S = \frac{1}{2} \mathbf{x} \cdot \mathbf{x} \Delta \phi + O(\Delta t^2) = \frac{1}{2} r^2 \dot{\phi} \Delta t + O(\Delta t^2).
\]

Therefore
\[
\frac{dS}{dt} = \frac{1}{2} r^2 \dot{\phi} = \frac{1}{2m} L \tag{2.9}
\]
This is known as Kepler’s second law: in equal times the particle’s radial vector sweeps out equal areas.
2.2. CENTRAL FIELDS

We can now use the constant \( L \) to help us write the total energy as

\[
E = T + U = \frac{1}{2} m (\dot{r}^2 + r^2 \dot{\phi}^2) + U(r) = \frac{1}{2} m \dot{r}^2 + \frac{L^2}{2mr^2} + U(r). \tag{2.10}
\]

If we define the effective potential to be

\[
U_{\text{eff}} = U(r) + \frac{M^2}{2mr^2}, \tag{2.11}
\]

then (2.10) looks like a one-dimensional system for the coordinate \( r \) with the new potential. The added term in the effective potential is called the centrifugal energy. When the effective potential is equal to the total energy, then we have \( \dot{r} = 0 \) which is a turning point for the particle. The particle is not at rest at such a point as was the case in the previous section however, unless the angular momentum is zero. Rearranging the total energy (2.10) we get

\[
\frac{dr}{dt} = \sqrt{\frac{2}{m} [E - U_{\text{eff}}(r)]} \tag{2.12}
\]

This is a separable differential equation, with solution

\[
t = \int \frac{dr}{\sqrt{\frac{2}{m} [E - U_{\text{eff}}(r)]}} + \text{constant}. \tag{2.13}
\]

We can also solve the separable differential equation (2.7) for \( \phi \) to find

\[
\phi = \int \frac{Md r}{r^2 \sqrt{2m [E - U_{\text{eff}}(r)]}} + \text{constant}, \tag{2.14}
\]

where we plugged in \( dr/dt \) from eq. (2.12) and integrated. This yields an equation for \( \phi \) as a function of \( r \). Note that the constancy of \( L \) requires that \( \phi \) cannot change sign.

As in the previous section, the motion is confined to the region \( U_{\text{eff}} \leq E \). If the particle is unbounded for some finite energy \( E \), then the potential energy at infinity \( U_{\infty} = \lim_{r \to \infty} U(r) = \lim_{r \to \infty} U_{\text{eff}} \) must be finite—assuming it exists (as it often does for physical systems). If \( E > U_{\infty} \), then we can define the velocity at infinity \( v_{\infty} \) via \( E = \frac{1}{2} mv_{\infty}^2 + U_{\infty} \). Conversely, in the region \( r \to 0 \), in order for the particle to reach the center \( U(r) \) must not outgrow the centrifugal energy:

\[
\lim_{r \to 0^+} r^2 U(r) < \frac{L^2}{2m}. \tag{2.15}
\]

Suppose now that for some given \( U_{\text{eff}} \) we have motion bounded by two turning points \( r_{\text{min}} \) and \( r_{\text{max}} \), which confines the particle within an annulus defined by these two radii. Points where \( r = r_{\text{min}} \) are called pericenters and where \( r = r_{\text{max}} \) are called apocenters. As before, the time symmetry for the one-dimensional system in \( r \) allow us to conclude that the particle’s motion is reversible, and so the trajectory will be symmetric about any ray from the origin.
through a pericenter or apocenter. According to our solution (2.14), the angle between successive pericenters (or apocenters) is

$$\Phi = 2 \int_{r_{\min}}^{r_{\max}} \frac{L \, dr}{r^2 \sqrt{2m[E - U_{\text{eff}}(r)]}}$$

(2.16)

by symmetry. In general, \( \Phi \neq 2\pi m/n \) for some integers \( m, n \), in which case the particle’s orbit is not closed (and is in fact dense) in the annulus. There are only two central fields in which all bounded orbits are closed: the first is the harmonic oscillator, \( U(r) = k r^2 \) for \( k > 0 \) (cf. section 1.3), and the second is when \( U(r) = -k/r \) for \( k > 0 \) which will be covered in section 2.4.

### 2.3 Closed bounded orbits

There are two standard examples of central fields: the harmonic oscillator potential

$$U(r) = k r^2, \quad k > 0,$$

(2.17)

and the gravitational potential

$$U(r) = -\frac{k}{r}, \quad k > 0.$$  

(2.18)

In this section, we will show that these are the only two central fields in which all bounded orbits are closed.

Suppose \( U(r) \) is a central field in which all bounded orbits are closed. Then for any given orbit bounded by \( r_{\min} \leq r \leq r_{\max} \), the angle between the pericenter and apocenter is

$$\Phi = \int_{r_{\min}}^{r_{\max}} \frac{M}{r^2 \sqrt{2m[E - U(r)]} - L^2/r^2} \, dr.$$  

(2.19)

If we make the substitution \( x = L/r \), the angle

$$\Phi = \int_{x_{\min}}^{x_{\max}} \frac{1}{\sqrt{2m[E - U(L/x)] - x^2/2m}} \, dx$$

(2.20)

looks like \((2m)^{-1}\) times the period for oscillations in the one-dimensional system \( W(x) = U(L/x) + x^2/2m \). Orbits near circular orbits have \( \Phi \approx \Phi_{\text{cir}} \) which we can calculate (cf. Exercise 2.2) to be

$$\Phi_{\text{cir}} = \frac{\pi L}{r^2 \sqrt{U''(r)}} = \pi \sqrt{\frac{U'}{3U'' + rU''}}.$$  

(2.21)

For closed orbits we must have \( \Phi_{\text{cir}} = 2\pi m/n \) for some integers \( m, n \), and in particular must be independent of \( r \). This yields a differential equation for
2.4. Kepler’s Problem

Consider a particle whose motion is governed by the central field

\[ U(r) = -\frac{k}{r}, \quad k > 0, \]  

which we might recognize as the gravitational field due to a fixed mass at the origin, or the attractive electric force of a fixed charge at the origin. As discussed in section 2.2, this corresponds to the one-dimensional effective potential

\[ U_{\text{eff}}(r) = -\frac{k}{r} + \frac{L^2}{2mr^2}. \]  

Note that \( \lim_{r \to 0^+} U_{\text{eff}}(r) = +\infty \) and \( \lim_{r \to \infty} U_{\text{eff}}(r) = +\infty. \) Since

\[ \frac{dU_{\text{eff}}}{dr}(r) = \frac{k}{r^2} - \frac{L^2}{mr^3} \]  

\( \Phi_{\text{circ}} = \frac{\pi}{\sqrt{\alpha + 2}}, \) \hspace{1cm} (2.22)

with the \( \alpha = 0 \) case corresponding to \( U(r) = b \log r. \) Now we split into cases.

If \( \alpha > 0, \) then \( U(r) \to \infty \) as \( r \to \infty. \) Substituting \( x = x_{\text{max}} y \) in the integral for \( \Phi \) yields

\[ \Phi = \int_{y_{\text{min}}}^{1} \frac{1}{\sqrt{2[W^*(1) - W^*(y)]}} \, dy, \quad W^*(y) = \frac{y^2}{2} + \frac{1}{x_{\text{max}}^2} U \left( \frac{L}{x_{\text{max}} y} \right). \]  

As \( E \to \infty, \) note that \( x_{\text{max}} \to \infty \) and the second term in \( W^* \) can be neglected. Moreover, in this limit we also have \( y_{\text{min}} \to 0 \) and so we can evaluate the integral to be \( \lim_{E \to \infty} \Phi(E, L) = \pi/2. \) But \( \Phi_{\text{circ}} \) is a constant, and so this requires that \( \alpha = 2. \)

If \( \alpha < 0, \) then \( U(r) = ar^\alpha \) for \( -2 < \alpha < 0. \) In the limit \( E \to 0^-, \) we calculate

\[ \lim_{E \to 0^-} \Phi(E, L) = \int_{0}^{1} \frac{1}{\sqrt{x^{\alpha} - x^2}} \, dx = \frac{\pi}{2 + \alpha}. \]  

Comparing this to eq. (2.22), we see that we must have \( \alpha = -1. \)

Therefore, the only two allowed cases are \( \alpha = -1, 2 \) which correspond to the gravitational and harmonic oscillator potentials. We have already seen that all orbits for the harmonic oscillator are closed, and in the next section we will solve the equations of motion for the gravitational potential and find that all bounded orbits are closed.

2.4 Kepler’s problem

Consider a particle whose motion is governed by the central field

\[ U(r) = ar^\alpha \]  

for \( \alpha \geq -2, \alpha \neq 0, \) or \( U(r) = b \log r. \) Plugging this back into \( \Phi_{\text{circ}} \) yields

\[ \Phi_{\text{circ}} = \frac{\pi}{\sqrt{\alpha + 2}}, \]  

for \( \alpha \geq -2, \alpha \neq 0, \) or \( U(r) = b \log r. \) Now we split into cases.

Right away, we see that if \( \alpha = 0 \) then we have \( \Phi_{\text{circ}} = \pi/\sqrt{2}; \) this is not commensurable with \( 2\pi \) and so small oscillations will not be closed.

If \( \alpha > 0, \) then \( U(r) \to \infty \) as \( r \to \infty. \) Substituting \( y = y_{\text{max}} x \) in the integral for \( \Phi \) yields

\[ \Phi = \int_{y_{\text{min}}}^{1} \frac{1}{\sqrt{2[W^*(1) - W^*(y)]}} \, dy, \quad W^*(y) = \frac{y^2}{2} + \frac{1}{x_{\text{max}}^2} U \left( \frac{L}{x_{\text{max}} y} \right). \]  

As \( E \to \infty, \) note that \( x_{\text{max}} \to \infty \) and the second term in \( W^* \) can be neglected. Moreover, in this limit we also have \( y_{\text{min}} \to 0 \) and so we can evaluate the integral to be \( \lim_{E \to \infty} \Phi(E, L) = \pi/2. \) But \( \Phi_{\text{circ}} \) is a constant, and so this requires that \( \alpha = 2. \)

If \( \alpha < 0, \) then \( U(r) = ar^\alpha \) for \( -2 < \alpha < 0. \) In the limit \( E \to 0^-, \) we calculate

\[ \lim_{E \to 0^-} \Phi(E, L) = \int_{0}^{1} \frac{1}{\sqrt{x^{\alpha} - x^2}} \, dx = \frac{\pi}{2 + \alpha}. \]  

Comparing this to eq. (2.22), we see that we must have \( \alpha = -1. \)
has only one root for \( r \in (0, \infty) \) at \( r = L^2/mk \), then \( U_{\text{eff}} \) obtains a global minimum with value
\[
U_{\text{eff}, \min} = U_{\text{eff}} \left( \frac{L^2}{k} \right) = -\frac{mk^2}{2L^2}.
\] (2.28)

Since this must be the only extrema, we conclude that \( E \geq 0 \) yields unbound motion, and \( E < 0 \) requires that the particle be bound with \( E \geq -mk^2/2L^2 \) if \( L \neq 0 \).

Using the general solutions found in section 2.2, we find the equation of motion:
\[
\phi(r) = \cos^{-1} \left( \frac{\frac{L}{r} - \frac{mk}{r}}{\sqrt{2mE + \frac{m^2k^2}{L^2} \cos \phi}} \right)
\] (2.29)
where we have picked the origin such that the integration constant is zero.

Rearranging, we get
\[
\sqrt{2mE + \frac{m^2k^2}{L^2} \cos \phi} = \frac{L}{r} - \frac{mk}{L}
\]
\[
r \left( \frac{mk}{L} + \sqrt{2mE + \frac{m^2k^2}{L^2} \cos \phi} \right) = L
\]
\[
r \left( 1 + \frac{L}{mk} \sqrt{2mE + \frac{m^2k^2}{L^2} \cos \phi} \right) = \frac{L^2}{mk}
\]
\[
r = \frac{L^2/mk}{1 + \sqrt{1 + (2EL^2/mk^2)} \cos \phi}.
\]
Define the quantities \( p = L^2/mk \) and \( \epsilon = \sqrt{1 + (2EL^2/mk^2)} \), so that we may write
\[
r(\phi) = \frac{p}{1 + \epsilon \cos \phi}.
\] (2.30)
This is the parametric equation for a conic section with one focus at the origin, where \( \epsilon \in [0, \infty) \) is the eccentricity and \( 2p \) is the latus rectum. Note that \( E \geq 0 \) yields unbound orbits (\( \epsilon = 1 \Rightarrow \) parabolas and \( \epsilon > 1 \Rightarrow \) hyperbolas) and \( -mk^2/2L^2 < E < 0 \) yields bound orbits (\( \epsilon = 0 \Rightarrow \) circles and \( 0 < \epsilon < 1 \Rightarrow \) ellipses). From geometry, the semi-major and -minor axes respectively are given by
\[
a = \frac{p}{1 - \epsilon^2} = \frac{k}{2|E|} \, , \quad b = \frac{p}{\sqrt{1 - \epsilon^2}} = \frac{M}{\sqrt{2m|E|}}.
\] (2.31)

For the planets in our solar system, the eccentricities are very small and the trajectories are nearly circular. Kepler observed this, and in his first law stated that the planetary orbits are circles with the Sun not at the center.

Now we determine the period \( \tau \) of a bounded elliptic orbit. Integrating Kepler’s second law (2.9) over one orbit, we get
\[
\pi ab = \int dS = \int \frac{1}{2m} L \, dt = \frac{1}{2} L \tau \quad \Rightarrow \quad \tau = \frac{2\pi mab}{L} = \pi k \sqrt{\frac{m}{2|E|^3}}.
\] (2.32)
2.5. VIRIAL THEOREM

This is an explicit formula for the particle’s period in terms of its energy. Using the formula [2.31] for the semi-major axis \( a \) in terms of the energy \( E \), we see that

\[
\tau = \pi k \sqrt{\frac{8ma^3}{2k^3}} = 2\pi \sqrt{\frac{ma^3}{k}}.
\]  

(2.33)

This demonstrates Kepler’s third law: the square of the orbital period of a planet is directly proportional to the cube of the semi-major axis of its orbit.

In practice, we know that a mass (or charge) sitting at the origin in order to create such a field is not unaffected by the particle’s presence—but this is easily remedied. The two-body problem asks for the equations of motion for a closed system consisting of two interacting particles with position \( x_i \) and mass \( m_i \), for \( i = 1, 2 \). The system is conservative with potential

\[
U(x_1, x_2) \equiv U(|x_1 - x_2|) = -\frac{Gm_1m_2}{|x_1 - x_2|}
\]  

(2.34)

where \( G \) is a constant called the gravitational constant. Define \( x = x_1 - x_2 \) to be the distance between the particles and redefine the origin to be the center of mass: \( m_1x_1 + m_2x_2 = 0 \). We can recover the positions form \( x \) via

\[
x_1 = \frac{m_2x}{m_1 + m_2}, \quad x_2 = -\frac{m_1x}{m_1 + m_2}.
\]  

(2.35)

The total energy can be written solely in terms of \( x \) as

\[
E = \frac{1}{2}m_1|x_1|^2 + \frac{1}{2}m_2|x_2|^2 + U(x_1, x_2) = \frac{1}{2}\mu|x|^2 + U(|x|),
\]  

(2.36)

with the reduced mass \( \mu \equiv m_1m_2/(m_1 + m_2) \). That is, the two-body system is equivalent to a single-particle of mass \( \mu \) moving in the external central field \( U(|x|) \). Solving for the motion of \( x \) in the reduced problem (2.36) yields the solutions to the original problem (2.34) using the relations (2.35). In this way, any force that satisfies the law of action and reaction gives rise to a central field. In particular, for two gravitational bodies (or two attractive charges) this means the equations of motion for each will be conic sections with a shared focus at the origin.

2.5 Virial theorem

The virial theorem is a formula for the time average of a system’s kinetic energy, which in the case of central field yields a remarkably simple relation. Consider a mechanical system of \( N \) particles in \( \mathbb{R}^d \). It turns out the key quantity to examine is the evolution of \( \sum_i x_i \cdot p_i \), which is not unreasonable since it is equal to \( 2I \) where \( I \) is the moment of inertia. We calculate

\[
\frac{d}{dt} \left( \sum_i x_i \cdot p_i \right) = \sum_i \dot{x}_i \cdot p_i + \sum_i x_i \cdot \dot{p}_i = 2T + \sum_i F_i \cdot x_i.
\]
CHAPTER 2. EXAMPLES

Taking a time average of both sides, i.e.

\[
\overline{f} = \lim_{T \to \infty} \frac{1}{T - t_0} \int_{t_0}^{T} f(\tau) \, d\tau
\]

(which is independent of choice of \(t_0\)), we get

\[
\frac{d}{dt} \left( \sum_i x_i \cdot p_i \right) = 2 \overline{T} + \sum_i F_i \cdot x_i. \tag{2.37}
\]

Note that for any bounded function \(f\) we have

\[
\frac{df}{dt} = \lim_{T \to \infty} \frac{1}{T} \int_0^T \frac{df}{d\tau}(\tau) \, d\tau = \lim_{T \to \infty} \frac{f(T) - f(0)}{T} = 0.
\]

Therefore, since all particle motion is bounded, then the quantity \(\sum_i x_i \cdot p_i\) must also be bounded, and so the left-hand side of eq. (2.37) vanishes:

\[
\overline{T} = -\frac{1}{2} \sum_i F_i \cdot x_i. \tag{2.38}
\]

This consequence is known as the *virial theorem*.

If we return to a single particle moving in a central field as in the previous section, then the force \(F = -dU/dr\) is conservative and is in the direction of \(\hat{r}\), and so

\[
\overline{T} = \frac{1}{2} U'(r)r. \tag{2.39}
\]

Moreover, if we have a power-law force \(F(r) \propto r^n\), then \(U(r) = ar^{n+1}\) and so

\[
U'(r)r = (n + 1)ar^{n+1} = (n + 1)U \implies \overline{T} = \frac{n + 1}{2} U. \tag{2.40}
\]

In particular, for gravitational force we have \(n = -2\) and

\[
\overline{T} = -\frac{1}{2} U. \tag{2.41}
\]

### 2.6 Central field scattering

One-body scattering is a uniform beam of particles in \(\mathbb{R}^3\), all of the same mass and energy, approaching a central force which is assumed to decay far away from the origin, so that the particle motion begins and ends colinearly at infinity. Examples of such systems include a beam of asteroids passing by a star, or a (classical treatment of a) beam of charged particles passing near an attractive or repulsive central charge. One of the beam characteristics is its *intensity* \(I\): the number of particles crossing unit area normal to the initial direction of travel per unit time. For the solid angle \(\Omega \subset S^2\) (where \(S^2\) has area \(4\pi\)), the scattering
2.6. CENTRAL FIELD SCATTERING

cross section \( \sigma(\Omega) \) is given by the number of particles scattered per unit solid angle per unit time divided by the incident intensity.

As established in section 2.2, we know the angular momentum \( L = |L| \) and energy \( E \) for each particle is conserved. Define the impact parameter \( s \) for a particle of mass \( m \) and initial velocity \( v_0 \) via

\[
L = mv_0s = s\sqrt{2mE},
\]

(2.42)
because initially (and hence always since it is conserved) we have \( E = \frac{1}{2}mv_0^2 \). Because the field is spherically symmetric, the scattering is rotationally symmetric about the incident beam’s axis, and so the quantities \( s \) and \( E \) are sufficient to determine the angle \( \Theta \) between the incident and scattered beams. This symmetry also tells us that the element of solid angle \( d\Omega \) is related to the element of angle \( d\Theta \) by \( d\Omega = 2\pi \sin \Theta d\Theta \). The particles between \( s, s + ds \) scattered between \( \Theta, \Theta + d\Theta \) is therefore

\[
2\pi Is \, ds = -2\pi \sigma(\Theta) I \sin \Theta d\Theta.
\]

(2.43)
The minus sign arises for repulsive fields, which is what we will focus on due to physical importance. Rearrangement yields

\[
\sigma(\Theta) = -\frac{s \sin \Theta}{\sin \Theta \, d\Theta}.
\]

(2.44)

An important example of central field scattering is for a beam of charged particles of charge \( -q \) incident towards a fixed charge \( -Q \), for which the field is given by the Coulomb force

\[
F(r) = \frac{qQ}{r^2} \hat{r} = \nabla U, \quad U(r) = \frac{qQ}{r}.
\]

(2.45)

Although in the previous section we assumed that \( U = -k/r \) with \( k > 0 \) in order to model celestial motion, nowhere did our calculations require that \( k = -qQ \) be positive, and so our solutions are still conic sections:

\[
r(\phi) = \frac{p}{1 + \epsilon \cos \phi} = -\frac{L^2}{mqQ(1 + \epsilon \cos \phi)}.
\]

(2.46)
The eccentricity is now given by

\[
\epsilon = \sqrt{1 + \frac{2EL^2}{m(qQ)^2}} = \sqrt{1 + \left(\frac{2Es}{qQ}\right)^2} \geq 1.
\]

(2.47)
So the motion is necessarily hyperbolic for \( E, s > 0 \). Moreover, the radius \( r(\phi) \) in eq. (2.46) must be positive, and so we also have the restriction

\[
\cos \phi < -\frac{1}{\epsilon}.
\]

(2.48)
Unlike in section 2.4 where the field center was the interior focus for hyperbolic motion, the origin is now the hyperbola’s exterior focus. Since \( r \to \infty \) as \( \phi \) approaches the bounds given by eq. (2.48), then the range \( \Phi \) of \( \phi \) and the total angle of deflection \( \Theta \) sum to \( \pi \), and so

\[
\sin \left( \frac{\Theta}{2} \right) = \sin \left( \frac{\pi}{2} - \frac{\Phi}{2} \right) = \cos \left( \frac{\Phi}{2} \right) = \frac{1}{\epsilon}
\]

\[
\cot \left( \frac{\Theta}{2} \right) = \sqrt{\csc^2 \left( \frac{\Theta}{2} \right) - 1} = \sqrt{\epsilon^2 - 1} = \frac{2Es}{qQ}
\]

This gives us the relation between \( s \) and \( \Theta \), and so using the cross-section (2.44) we find the Rutherford scattering cross section

\[
\sigma(\Theta) = -\frac{qQ}{2E} \cot \left( \frac{\Theta}{2} \right) \frac{d}{d\Theta} \left( \frac{qQ}{2E} \cot \left( \frac{\Theta}{2} \right) \right)
\]

\[
= - \left( \frac{qQ}{2E} \right)^2 \frac{\sin(\Theta/2) \cos(\Theta/2)}{\sin(\Theta/2)^2} \left( - \frac{1}{2\sin^2(\Theta/2)} \right) \quad (2.49)
\]

In context, this formula is a classical approximation of a quantum system, and thus it has limitations. For example, if we were to integrate this expression over the sphere we would obtain an infinite total scattering cross section, which is of course physically impossible.

As with Kepler’s problem, in practice the central charge is not fixed but rather recoils as a result of the scattering. Here, we cannot get away with a simple substitution of the reduced mass since \( \Theta \) measures the angle between the initial and final vectors between the charges, and no longer the deflection angle \( \vartheta \) between the initial and final directions of the scattered particle. Assume the “central” particle is initially stationary, and let the subscript 1 refer to the scattered particle, primed coordinates denote coordinates relative to the center of mass (as opposed to the laboratory coordinates), and \( \mathbf{X} \) represent the position of the center of mass in laboratory coordinates. As before, we have \( \mathbf{x}_1 = \mathbf{X} + \mathbf{x}'_1 \), \( \mathbf{v}_1 = \mathbf{x}_1 = \mathbf{X} + \mathbf{v}'_1 \). In the final configuration, the center of mass velocity \( \dot{\mathbf{X}} \) still lies along the initial direction by conservation of momentum (the two-particle system is conservative), and so geometrically we have

\[
\tan \vartheta = \frac{|\mathbf{v}'_1| \sin \Theta}{|\mathbf{v}'_1| \cos \Theta + |\mathbf{X}|}.
\]  

(2.50)

Using the coordinate relations (2.35) and the reduced mass \( \mu \) of the two particles, we know

\[
\mathbf{v}'_1 = \frac{\mu}{m_1} \mathbf{v},
\]  

(2.51)
2.6. CENTRAL FIELD SCATTERING

where \( x \equiv x_1 - x_2 \) as before. Moreover, energy must be conserved and so letting \( v_0 = X_{\infty} \) denote the initial relative velocity, we get

\[
|v'_1| = \frac{\mu}{m_1}|v_0|
\] (2.52)

This, in tandem with the conservation of total linear momentum:

\[
(m_1 + m_2)\dot{X} = m_1v_0 \quad \Rightarrow \quad |\dot{X}| = \frac{\mu}{m_2}|v_0|,
\] (2.53)

allows us to rewrite eq. (2.50) as

\[
\tan \vartheta = \frac{\frac{\mu}{m_1}|v_0|\sin \Theta}{\frac{\mu}{m_1}|v_0|\cos \Theta + \frac{\mu}{m_2}|v_0|} = \frac{\sin \Theta}{\cos \Theta + \frac{m_2}{m_1}}.
\] (2.54)

As we would expect experimentally, when \( m_2 \gg m_1 \) then the initially stationary particle does not experience significant recoil and \( \vartheta \approx \Theta \). We would now like to determine the cross section \( \sigma'(\vartheta) \) in the laboratory system. The number of particles scattered into a fixed element of solid angle must be the same in both the laboratory and relative to the center of mass. Consequently,

\[
2\pi I\sigma(\Theta)\sin(\Theta)\,d\Theta = -2\pi I s\,ds = 2\pi I \sigma'(\vartheta)\sin(\vartheta)\,d\vartheta,
\]

and so

\[
\sigma'(\vartheta) = \sigma(\Theta)\frac{\sin \Theta}{\sin \vartheta} \frac{d\Theta}{d\vartheta} = \sigma(\Theta)\frac{d(\cos \Theta)}{d(\cos \vartheta)}.
\] (2.55)

Our model illustrates elastic scattering, in that the total kinetic energy is conserved. Since the second particle is initially at rest and is then set into motion, the velocity \( |v_1| \) after scattering is less than \( |v_0| \). Starting with the law of cosines we have

\[
|v'_1|^2 = |v_1|^2 + |\dot{X}|^2 - 2|v_1||\dot{X}|\cos \vartheta,
\]

and so

\[
\left|\frac{v_1}{v_0} - \frac{2\mu}{m_2}\frac{|v_1|}{|v_0|}\cos \vartheta - \frac{m_2 - m_1}{m_2 + m_1}\right| = 0,
\] (2.56)

using the conserved quantities (2.52) and (2.53).

Rutherford was interested in \( \alpha \)-particle scattering, for which corrections due to \( m_1/m_2 \ll 1 \) are small. When \( m_1 = m_2 \), the angle relation (2.54) becomes

\[
\tan \vartheta = \frac{\sin \Theta}{\cos \Theta + 1} = \tan \left( \frac{\Theta}{2} \right) \quad \Rightarrow \quad \vartheta = \frac{\Theta}{2}.
\] (2.57)

So \( 0 \leq \vartheta \leq \pi/2 \) and there can be no back scattering. The results (2.55) and (2.56) simplify to

\[
\sigma'(\vartheta) = 4\cos \vartheta \sigma(2\vartheta), \quad \left|\frac{v_1}{v_0}\right| = \cos \vartheta.
\] (2.58)
2.7 Exercises

2.1 (Pendulum period). Show that for the pendulum of Exercise 1.1,
\[ \ddot{x} = -\frac{g}{\ell} \sin x, \]
the motion with turning points \( \pm \theta_0 \) has period
\[ \tau = 4 \sqrt{\frac{l}{g}} K \left( \sin \left( \frac{\theta_0}{2} \right) \right), \]
where
\[ K(k) = \int_0^{\pi/2} \frac{d\xi}{\sqrt{1 - k^2 \sin^2 \xi}} \]
is the complete elliptic integral of the first kind. Taylor expanding about \( \theta_0 = 0 \), find the one-term expansion:
\[ \tau \approx 2\pi \sqrt{\frac{l}{g}} \left( 1 + \frac{\theta_0^2}{16} + \ldots \right). \]
Note that the zeroth order term is the constant period approximation obtained when we take \( \sin x \approx x \) and replace the pendulum with a harmonic oscillator.

2.2 (Small one-dimensional oscillations). Let \( E_0 \) be the value of the potential energy of a one-dimensional system at a minimum point \( \xi \). Show that the period \( \tau_0 = \lim_{E \to E_0} \tau(E) \) of small oscillations in a neighborhood of the point \( \xi \) is \( 2\pi/\sqrt{U''(\xi)} \).

2.3. Show that for the central field with potential
\[ U(r) = -\frac{k}{r^3}, \]
the motion falls into the origin in finite time.

2.4 (Method of similarity). Suppose that the potential energy of a central field is a homogeneous function of degree \( \nu \):
\[ U(\alpha r) = \alpha^\nu U(r) \quad \text{for all } \alpha > 0. \]
Show that if a curve \( \gamma \) is the orbit of a motion, then the homothetic curve \( \alpha \gamma \) is also an orbit (under the appropriate initial conditions). Determine the ratio of the circulation times along these orbits. Conclude from this the isochronicity of the harmonic oscillator (\( \nu = 2 \)) and Kepler’s third law (\( \nu = -1 \)).

2.5 (Escape velocity). Let \( r_0 \) denote the radius of the Earth, and \( g \approx 9.8 \text{ m/sec}^2 \) the gravitational acceleration at the Earth’s surface. The gravitational potential energy of the Earth is
\[ U(r) = -\frac{gr_0^2}{r}. \]
Determine with what velocity a particle must be given on the surface of the Earth in order for it to travel infinitely far away.
2.6 (Cosmic velocities). Consider the gravitational potential energy of the Earth as in the previous problem. The escape velocity \( v_2 \) is sometimes called the second cosmic velocity. The first cosmic velocity is the velocity of motion on a circular orbit of radius close to the radius of the earth. Find the magnitude of the first cosmic velocity \( v_1 \) and show that \( v_2 = \sqrt{2}v_1 \).

2.7 (Geosynchronous orbit [Nah16]). It is useful for communication satellites to be in geosynchronous orbit, so that their orbital period is that of the Earth’s and the satellite appears to hover in the sky. We will calculate the height of this orbit for Earth in two different ways.

(a) Let \( m \) be the satellite’s mass, \( M = 5.98 \times 10^{24} \) kilograms be the Earth’s mass, \( v \) be the satellite’s velocity, and \( R_s \) the radius of the satellite’s circular orbit. First equate the gravitational and centripetal accelerations of a circular orbit, then substitute \( v = \frac{2\pi R_s}{T} \) where \( T = 86,400 \) seconds is the period of the orbit, and solve for \( R_s \).

(b) Use Kepler’s third law to calculate the same value for \( R_s \).

2.8 (Satellite paradox [Nah16]). Satellites in low Earth orbit experience significant atmospheric drag, which counterintuitively increases the speed of the satellite.

(a) For a circularly orbiting satellite in Earth’s gravitational potential, conclude from the virial theorem that the satellite’s total energy is

\[
E = -\frac{1}{2}mv^2.
\]

Alternatively, this relation can be obtained by equating the gravitational and centripetal accelerations, solving for \( v \), and substituting \( v \) into the kinetic energy.

(b) Differentiate the result from part (a) to determine \( \frac{dv}{dt} \) in terms of \( \frac{dE}{dt} \). Assuming the energy loss rate (dissipated power) of the satellite is

\[
\frac{dE}{dt} = -vf_d, \quad f_d > 0,
\]

show that \( \frac{dv}{dt} > 0 \).

2.9 (Solar and lunar tides [Nah16]). If we differentiate the gravitational potential (2.34), we find the gravitational force between two bodies of masses \( m_1 \) and \( m_2 \) to have magnitude

\[
F(r) = G\frac{m_1 m_2}{r^2}.
\]
where $r$ is the distance between the bodies’ centers and $G$ is the universal gravitational constant. For the Earth, Sun, and Moon, we have

\[
M_s = \text{mass of the Sun} = 2 \times 10^{30} \text{ kilograms},
M_m = \text{mass of the Moon} = 7.35 \times 10^{22} \text{ kilograms},
R_s = \text{Earth–Sun separation} = 93 \times 10^6 \text{ miles},
R_m = \text{Earth–Moon separation} = 2.39 \times 10^5 \text{ miles}.
\]

(a) Show that, even though the Sun is much farther from the Earth, that the Sun’s gravitational force on the Earth is greater than the Moon’s, and find their ratio.

(b) Since the Earth is not a point mass, then the Sun’s gravitational force is stronger (weaker) on the side of the Earth closest (farthest) from the Sun. This causes water to bulge at the points closest an furthest from the Sun, which is called the solar high tides. Calculate the maximum difference in gravitational force in terms of Earth’s radius $R$ for both the Sun and the Moon.

(c) Extract the leading term in the limit $R/R_s \ll 1$ and $R/R_m \ll 1$ for each expression in part (b). Calculate their ratio and conclude that, although the Sun’s gravitational force is stronger, the lunar tides are more than twice as large as the solar tides.

2.10 (Energy of the ocean tides [Nah16]). The high tides are not directly in line with the centers of the Earth and Moon, but are rather carried ahead slightly by the Earth’s rotation and frictional forces. This means that the Moon’s gravitational pull on both tides produce torque: the Moon’s pull on the farther tide increases the Earth’s rotational speed, but the stronger pull on the nearer tide is counter-rotational, and so the overall effect decreases the Earth’s rotational speed. Atomic clocks are measured that the length of a day is increasing at the rate of about 2 milliseconds per century.

(a) Let $\Omega$ denote the angular rotation rate of the Earth and let $T$ denote the length of a day in seconds, so that $\Omega T = 2\pi$. By integrating the kinetic energy of an infinitesimal mass $dm$ over the volume of the Earth, show that the rotational energy $E$ is given by

\[
E = \frac{1}{2} \Omega^2 I,
\]

where

\[
I = \iiint_{r \leq R} (x^2 + y^2) \, dm
\]

is the moment of inertia.
(b) Show that for a solid sphere of radius $R$ and constant mass density $\rho$ the moment of inertia is

$$I = \frac{8\pi}{15} R^4 \rho,$$

or in terms of the total mass $M$,

$$I = \frac{2}{5} MR^2.$$

The Earth is not a constant-density sphere, and so rather than 0.4 the coefficient is approximately 0.3444.

(c) Write the rotational energy $E$ as a function of the period $T$, and show that

$$\frac{dE}{dT} = -4 \cdot 0.3444 M \pi^2 R^2 T^{-3}.$$

Taking $T = 86,400$ seconds, the length of a day, and $\Delta T = 2 \times 10^{-3}$ seconds, $M = 5.98 \times 10^{24}$ kilograms, and $R = 6.38 \times 10^6$ meters, find the change in the Earth’s rotational energy $\Delta E$ over a century. Dividing $\Delta E$ by the number of seconds in a century, conclude that the power of the ocean tides is 3,260 gigawatts or 4.37 billion horsepower.

2.11 (Moon recession rate [Nah16]). As in Exercise 2.10, tidal friction decreases the Earth’s rotational angular momentum. Consequently, the Moon’s orbital angular momentum increases to conserve total angular momentum, which results in the Moon drifting away from the Earth. We will estimate this recession rate, assuming that all of the momentum is transferred to Moon’s orbit (and not rotation).

(a) Consider the Moon as a point mass $m$ orbiting circularly about the Earth at a radius $r$, with speed $v$ and angular speed $\omega$ radians per second. Show that the Moon’s orbital angular momentum is $L_m = mrv$.

(b) The gravitational force on the Moon by the Earth has magnitude

$$F = \frac{GMm}{r^2},$$

where $M$ is the mass of the Earth and $G$ is the universal gravitational constant. Equating the gravitational and centripetal accelerations of the Moon, find $v$ as a function of $r$ and use this to find the angular momentum $L_m$ as a function of $r$.

(c) Using part (b) of Exercise 2.10, the spin angular momentum of the Earth is $L_e = 0.3444 MR^2 \Omega$ where $\Omega$ is Earth’s rotation rate. Expressing $\Omega$ in terms of the day length $T = 86,400$ seconds, find $L_e$ as a function of $T$ and calculate $dL_e/dT$. 
(d) Using the daily change $\Delta T = 2 \times 10^{-5}/365$ seconds in the length of a day, approximate the daily change and then the yearly change in $\Delta L_e$.

(e) Equating change in the Moon’s orbital momentum $\Delta L_m$ with the change in Earth’s rotational momentum $|\Delta L_e|$, find the yearly change in the Moon’s orbital radius. Using the values

\[
M = \text{mass of the Earth} = 5.98 \times 10^{24} \text{ kilograms}, \\
m = \text{mass of the Moon} = 7.35 \times 10^{22} \text{ kilograms}, \\
r = \text{radius of Moon’s orbit} = 3.84 \times 10^8 \text{ meters}, \\
R = \text{Earth’s radius} = 6.37 \times 10^6 \text{ meters}, \\
G = \text{gravitational constant} = 6.67 \times 10^{-11} \frac{\text{meters}^3}{\text{kilograms} \times \text{seconds}^2},
\]

find that the Moon is receding from the Earth at a rate of $3.75 \text{ centimeters of 1.48 inches per year}$. This value is in outstanding agreement with measurements made by a laser on Earth and corner cube reflectors placed on the Moon during the Apollo 11 mission.
Newton’s equations are rephrased as the abstract and coordinate-free principle of least action, so that the configuration space can be endowed with any coordinates or replaced by any manifold. The subsequent Euler–Lagrange equations of motion are often the preferred perspective in physics, and consist of a second-order ordinary differential equation for each degree of freedom. These equations are independent of coordinate choice, which is a powerful tool in applications as we are free to choose the most convenient coordinates to describe our system. When the chosen coordinates align with symmetries of the system, the Euler–Lagrange equations are effective in identifying the corresponding conserved quantities, reducing the number of equations, and simplifying the system. Further study of the action—an initially abstract quantity introduced to formulate the principle of least action—leads to the Hamilton–Jacobi equation of motion: a single partial differential equation with two times the independent variables as the degrees of freedom.
Chapter 3

Euler–Lagrange equations of motion

We will reformulate the equations of motion as the coordinate-free principle of least action, the unpacking of which will demonstrate some of the fundamental techniques of variational calculus. This presentation of the calculus of variations follows [AKN06, Ch. 1], and in the rest of the chapter is based on [LL76, Ch. 1–2], [Arn89, Ch. 3], and [Gol51, Ch. 1–2].

3.1 Principle of least action

For a mechanical system with \( N \) particles, we will replace the configuration space \((\mathbb{R}^d)^N\) with a smooth manifold \( M \). We will often denote by \( q \) any (local) coordinates on \( M \) (sometimes called generalized coordinates in physics), and the number of components \( n = \dim M \) is called the number of degrees of freedom.

**Definition 3.1.** A Lagrangian system \((M, \mathcal{L})\) is a smooth \( n \)-dimensional manifold \( M \) called the configuration space together with a smooth time-dependent function \( \mathcal{L}(q, v, t) : TM \times I \to \mathbb{R} \) (where \( I \subset \mathbb{R} \) is an interval) on the tangent bundle called the Lagrangian.

As we will see in section 3.2, the following abstract condition in fact encodes Newton’s equations.

**Definition 3.2** (Hamilton’s principle of least action). A motion \( q(t) \) of the Lagrangian system \((M, \mathcal{L})\) between any two times \( t_0 < t_1 \) is a smooth path on \( M \) that is a critical point of the action functional

\[
S(q(t)) = \int_{t_0}^{t_1} \mathcal{L}(q(t), \dot{q}(t), t) \, dt.
\]  

(3.1)

The principle is called “least” action because it turns out that the motion is often a minimum, but we do not need this additional assumption. In this section we will see how to extract equations of motion from this assumption.
The goal of variational calculus is to find the extrema of functionals. A path from \( a_1 \in M \) to \( a_2 \in M \) (not necessarily distinct) starting at \( t_1 \) and ending at \( t_2 > t_1 \) is a smooth map \( \gamma : [t_1, t_2] \to M \) with \( \gamma(t_1) = a_1 \) and \( \gamma(t_2) = a_2 \). Let \( \Omega \) denote the collection of all such paths, which can be thought of as an “infinite-dimensional manifold”. A functional \( \Phi \) is a function from \( \Omega \) into \( \mathbb{R} \).

**Example 3.3.** The arc length of the graph of a smooth function \( x(t) \) on some \( (t_0, t_1) \) into \( M = \mathbb{R}^d \) is a functional. It takes as input a path \( \gamma \) on \( \mathbb{R}^d \) and returns
\[
\Phi(\gamma) = \int_{t_0}^{t_1} \sqrt{1 + \dot{x}^2} \, dt.
\]
Intuitively, we expect the path of minimum length between two points to be the line segment connecting those two points, and mathematically we see that \( \Phi(\gamma) \geq \int_{t_0}^{t_1} \dot{t} \, dt = t_1 - t_0 \) with equality if and only if \( \dot{x} = 0 \). The more systematic machinery we develop in this section should also return this answer.

We want to obtain a differential equation which a functional’s extrema must satisfy, arising from the vanishing of some sort of first derivative. For more concrete examples we do not need to assume arbitrary smoothness (cf. Exercise 3.2) nor even assume the existence of a minimizer (cf. Exercise 3.6), but here we will assume existence (as before) and remain within the context of smooth manifolds. A (fixed-endpoint) variation of a path \( \gamma \in \Xi \) is a smooth map \( H(s, t) = H_s(t) : (-\epsilon, \epsilon) \times [t_0, t_1] \to M \) for some \( \epsilon > 0 \) such that \( H(0, \cdot) = \gamma \), \( H_s = H(s, \cdot) \in \Xi \) for all \( s \in (-\epsilon, \epsilon) \), and \( H(s, t_i) = a_i \) for \( i = 1, 2 \) for all \( s \in (-\epsilon, \epsilon) \). A variation can be thought of a path on \( \Xi \) through \( \gamma \) at \( s = 0 \).

Just like with functions on Euclidean space, we would like to say that a functional \( \Phi \) is differentiable if
\[
\Phi(H_s) = \Phi(\gamma) + s \delta \Phi \left( \frac{\partial H_s}{\partial s} \right) + o(s) \quad \iff \quad \lim_{s \to 0} \frac{\Phi(H_s) - \Phi(\gamma)}{s} = \delta \Phi \left( \frac{\partial H_s}{\partial s} \right)
\]
for all variations \( H \), where \( \delta \Phi \) should be linear in \( \partial H_s / \partial s \). When this is the case, \( \delta \Phi \) is of course the derivative or first variation of \( \Phi \), and \( \gamma \) is a critical point for \( \Phi \) if \( \delta \Phi \equiv 0 \). However, we first need to figure out what \( \delta \Phi \) and \( \partial H_s / \partial s \) are.

The derivative \( \partial H_s / \partial s \) should be tangent to \( \Xi \) at the path \( H_s \). A tangent vector \( W \) to \( \Xi \) at a path \( \gamma \in \Xi \) is a function which associates to each \( t \in [t_0, t_1] \) a tangent vector \( W_t \in T_{\gamma(t)} M \), and is smooth in the sense that for every smooth function \( f : M \to \mathbb{R} \) the function \( t \mapsto df(W_t) \in \mathbb{R} \) is smooth. The tangent space \( T_\gamma \Xi \) is the space of all tangent vectors \( W \) at \( \gamma \) such that \( W_{t_0} = 0 = W_{t_1} \).

Now we can define the tangent vector \( \partial H_s / \partial s(0) \) to the variation \( H_s \) at \( s = 0 \) to be
\[
\frac{\partial H}{\partial s}(0, t) \in T_{\gamma(t)} M \text{ for } t \in [t_0, t_1].
\]
Since
\[
\frac{\partial H}{\partial s}(0, t_0) = 0, \quad \frac{\partial H}{\partial s}(0, t_1) = 0
\]
by the fixed endpoint requirement, then we indeed have \( \partial H_s / \partial s(0) \in T_{s=0} \Omega \). Now we can set
\[
\delta \Phi \left( \frac{\partial H_s}{\partial s} \right) = \frac{d}{ds} \Phi \circ H_s \bigg|_{s=0}.
\]
As we would hope, \( \delta \Phi \) is well-defined and linear.

For the rest of this section, we will restrict our attention to the action functional \( S \) defined at the beginning of the section. If we take \( L(x, y, t) = \sqrt{1 + y^2} \), we recover the functional of Example 3.3. The reason we insisted that \( L \) be differentiable is so that the action is differentiable.

**Proposition 3.4 (Euler–Lagrange equations).** The action functional is differentiable with derivative
\[
\delta S \left( \frac{\partial H_s}{\partial s} \right) = \int_{t_0}^{t_1} \left( \frac{\partial \mathcal{L}}{\partial q} - \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{q}} \right) \frac{\partial H_s}{\partial s} dt.
\]
Consequently, a path \( q(t) \) is a critical point for the action (i.e. \( \delta S \equiv 0 \)) if and only if \( q(t) \) solves the differential equations
\[
\frac{d}{dt} \left[ \frac{\partial \mathcal{L}}{\partial \dot{q}} (q(t), \dot{q}(t), t) \right] - \frac{\partial \mathcal{L}}{\partial q} (q(t), \dot{q}(t), t) = 0.
\]

The \( n \)-many second-order differential equations (3.3) are called the Euler–Lagrange equations for the functional \( S \), or simply the Lagrange equations when applied to a mechanical system. Note that eq. (3.3) must hold for any choice of coordinates \( q \), where \( \partial \mathcal{L} / \partial q \) is the gradient of \( \mathcal{L}(q, v, t) \) in the \( q \) variables. The derivative \( \partial \mathcal{L} / \partial \dot{q} \) represents the derivative of the Lagrangian \( \mathcal{L}(q, v, t) \) in the velocity variables \( v \), and should really be noted as \( (\partial \mathcal{L} / \partial v)(q, \dot{q}, t) \). Although eq. (3.3) is often shortened to
\[
\frac{d}{dt} \left( \frac{\partial \mathcal{L}}{\partial \dot{q}} (q(t), \dot{q}(t), t) \right) - \frac{\partial \mathcal{L}}{\partial q} (q(t), \dot{q}(t), t) = 0,
\]
we are meant to plug in \( q \) and \( \dot{q} \) before taking the total time derivative \( d / dt \), which will for example turn \( \dot{q} \) into \( \ddot{q} \).

**Proof.** Fix a set of coordinates \( q \) on \( M \), let \( q(t) \) be a path, and let \( H \) be a variation of \( q(t) \). For the rest of this proof we will pull everything back to the domain of \( q \), the coordinate patch lying within \( \mathbb{R}^n \). Let \( x(t) \) denote the path in \( \mathbb{R}^n \), \( \mathcal{L}(x, x, t) \) denote the Lagrangian in \( \mathbb{R}^n \), and \( x(t) + h(t) \) the variation \( H_s(t) \) in \( \mathbb{R}^n \) (we are suppressing the \( s \)-dependence of \( h \)). Note that the fixed endpoint condition requires that \( h(t_0) = 0 = h(t_1) \).

Now that we are in Euclidean space, we can proceed with the usual variational argument. Since \( \mathcal{L} \) is differentiable, we may Taylor expand to get
\[
\mathcal{L}(x + h, \dot{x} + \dot{h}, t) = \mathcal{L}(x, \dot{x}, t) + \frac{\partial \mathcal{L}}{\partial x} \cdot h + \frac{\partial \mathcal{L}}{\partial \dot{x}} \cdot \dot{h} + O(h^2).
\]
Therefore,

\[ S(\gamma + h) - S(\gamma) = \int_{t_0}^{t_1} \left[ L(x + h, \dot{x} + \dot{h}, t) - L(x, \dot{x}, t) \right] \, dt \]

\[ = \int_{t_0}^{t_1} \left( \frac{\partial L}{\partial x} \cdot h + \frac{\partial L}{\partial \dot{x}} \cdot \dot{h} \right) \, dt + O(h^2). \]

We are able to pull the term \( O(h^2) \) outside of the integral since \( L \) is continuously differentiable. The integral in the rightmost expression will be the derivative \( \delta S \). Integrating by parts yields

\[ \int_{t_0}^{t_1} \frac{\partial L}{\partial \dot{x}} \cdot \dot{h} \, dt = -\int_{t_0}^{t_1} \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{x}} \right) \dot{h} \, dt + \left( \frac{\partial L}{\partial \dot{x}} \cdot h \right) \bigg|_{t_0}^{t_1}. \]

Because \( h(t_0) = 0 = h(t_1) \), we know that the boundary term must vanish:

\[ \left( \frac{\partial L}{\partial \dot{x}} \cdot h \right) \bigg|_{t_0}^{t_1} = 0, \]

and plugging this back into \( \delta S \) produces eq. (3.2) as desired.

It is clear from the formula that if eq. (3.3) is satisfied then the functional derivative (3.2) \( \delta S \) must be identically zero. Conversely, assuming that \( \delta S = 0 \) for any variation, we know that

\[ \int_{t_0}^{t_1} \left( \frac{\partial L}{\partial x} - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} \right) \cdot h \, dt = 0 \quad (3.4) \]

for all smooth \( h : [t_0, t_1] \to \mathbb{R}^n \) with \( h(t_0) = 0 = h(t_1) \). We can conclude that the integrand without \( h \) identically vanishes (and so eq. (3.3) holds) using the following basic analysis lemma. \( \square \)

**Lemma 3.5.** If a continuous function \( f : [t_0, t_1] \to \mathbb{R} \) satisfies \( \int_{t_0}^{t_1} f(t)h(t) \, dt = 0 \) for all smooth functions \( h : [t_0, t_1] \to \mathbb{R} \) with \( h(t_0) = 0 = h(t_1) \), then \( f(t) = 0 \) for all \( t_0 \leq t \leq t_1 \).

**Remark.** In the proof of Proposition 3.4, we only need that the coordinates are spanning in order to have enough directions \( h(t) \) to conclude that the Euler–Lagrange equation holds. Therefore, there is no issue in introducing more than \( n \) coordinates as long as they span the same space.

**Corollary 3.6.** Transforming the Lagrangian by a total time derivative:

\[ L(q, \dot{q}, t) \mapsto L' = L(q, \dot{q}, t) + \frac{d}{dt} [f(q, t)] \]

leaves the Euler–Lagrange equations (3.3) unchanged.
Proof. The new action is given by

$$S' = \int_{t_0}^{t_1} L(q, \dot{q}, t) + \int_{t_0}^{t_1} \frac{df}{dt} dt = S + f(q^{(2)}, t_1) - f(q^{(1)}, t_0)$$

and the addition of a constant to $S$ does not affect the principle of least action.

Example 3.7. We will apply the Euler–Lagrange equations (3.3) to determine the extrema of the arc length functional of Example 3.3. Using Cartesian coordinates $x$ on $\mathbb{R}^d$, the Euler–Lagrange equations are

$$0 = \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{x}} - \frac{\partial \mathcal{L}}{\partial x} = \frac{d}{dt} \left( \frac{\dot{x}}{\sqrt{1 + \dot{x}^2}} \right) - 0.$$ 

This means

$$\frac{\dot{x}}{\sqrt{1 + \dot{x}^2}} = a$$

for some constant $a \in \mathbb{R}^n$. Squaring and rearranging yields

$$\dot{x} = b$$

for a new constant $b \in \mathbb{R}^n$. Finally, integrating yields

$$x(t) = bt + c$$

for $b, c \in \mathbb{R}^d$. That is, the extrema of the arc length functional are straight lines. If we had instead chosen, say, polar coordinates, the Euler–Lagrange equations would be different but the solutions would still describe straight lines.

### 3.2 Conservative systems

In the previous sections we saw how the principle of least action yields the $n = \dim M$ second-order Euler–Lagrange equations

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

which when applied to mechanical systems are simply referred to as Lagrange’s equations. In this section we will see how eq. (3.5) is a formulation of Newton’s equations, and so Newton’s principle of determinacy is contained in Hamilton’s more concise principle of least action. For this reason, many graduate physics texts opt to begin with the principle of least action and view Newton’s equations as a consequence.

First we will see how to apply the Euler–Lagrange equations in order to obtain the equations of motion. Let $x_i \in \mathbb{R}^d$ be Cartesian coordinates on
Euclidean space. For a conservative system of \( N \) particles the Lagrangian is given by

\[ L(x, \dot{x}, t) = T(\dot{x}) - U(x), \tag{3.6} \]

where \( T \) and \( U \) are the kinetic and potential energies. The principle of least action implies that the motion \( x(t) = (x_1(t), \ldots, x_N(t)) \) of the system satisfies the Euler–Lagrange equations \((3.5)\). For this Lagrangian we have

\[
\begin{align*}
\frac{\partial L}{\partial \dot{x}_i} &= \frac{\partial T}{\partial \dot{x}_i} = \frac{\partial}{\partial x} \left( \sum_{i=1}^{N} \frac{1}{2} m_i \dot{x}_i^2 \right) = \mathbf{p}, \\
\frac{\partial L}{\partial x} &= -\frac{\partial U}{\partial x} = -\nabla U = \mathbf{F},
\end{align*}
\]

Therefore the Euler–Lagrange equations are Newton’s equations \((1.2)\). The advantage now is that we are now longer restricted to Euclidean space.

**Example 3.8.** Consider a pendulum consisting of a mass \( m \) attached to the end of a rigid massless rod of length \( \ell \) with the other end fixed, allowed to rotate in a vertical plane subject to a constant downward gravitational acceleration \( g \). Let \( \theta \) denote the angle from the vertical directly below the pivot, which entirely describes the system. The configuration space is the circle \( S^1 \), and the Lagrangian is a (time-independent) function defined on \((\theta, \dot{\theta}) \in TS^1 = S^1 \times \mathbb{R}\).

The kinetic energy is \( T = \frac{1}{2} mv^2 = \frac{1}{2} m \ell^2 \dot{\theta}^2 \), and since the force acting on the mass is \( F = ma = -mg \sin \theta \) \( \text{(cf. Exercise [1.1])} \) then the potential energy is

\[ U = -mg \ell \cos \theta. \]

(We picked our integration constant so that \( U = 0 \) when the mass is at the height of the pivot \( \theta = \pi/2 \).) The Lagrangian is thus

\[ \mathcal{L} = T - U = \frac{1}{2} m \ell^2 \dot{\theta}^2 + mg \ell \cos \theta, \]

and so the Euler–Lagrange equation is

\[ 0 = \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{\theta}} - \frac{\partial \mathcal{L}}{\partial \theta} = m \ell^2 \ddot{\theta} + mg \ell \sin \theta \implies \ddot{\theta} + \frac{g}{\ell} \sin \theta = 0, \]

which agrees with what we found with Newton’s equation in Exercise \([1.1]\).

In fact, the reason for the form of the Lagrangian \((3.6)\) stems from Galileo’s principle of relativity, presented in section \([1.1] \), now applied to the Lagrangian \( \mathcal{L} \) rather than the force. Consider a single particle in space whose motion is denoted by the Cartesian coordinates \( x \in \mathbb{R}^d \). In an inertial frame, the Lagrangian \( \mathcal{L}(x, \dot{x}, t) \) of this particle cannot be explicitly dependent on position or time by homogeneity: \( \mathcal{L} \equiv \mathcal{L}(\dot{x}) \). Also \( \mathcal{L} \) cannot be dependent on the direction of
v = \dot{x} either since space is isotropic: \mathcal{L} \equiv \mathcal{L}(|v|^2) (since \mathcal{L} must be smooth). Since \partial \mathcal{L}/\partial x = 0, then the Euler–Lagrange equation for this particle is
\[
\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{x}} = 0. \tag{3.7}
\]
The quantity \partial \mathcal{L}/\partial \dot{x} is therefore constant in time, and since \mathcal{L} \equiv \mathcal{L}(v) then v must be constant. This is the Lagrangian mechanics demonstration of Newton’s first law: in an inertial frame, any free motion takes place with constant velocity (both magnitude and direction).

Now we know that \mathcal{L} \equiv \mathcal{L}(|v|^2), but we would like to conclude that for our free particle we must have \mathcal{L} = T. Consider an inertial frame K moving with infinitesimal velocity \epsilon relative to an inertial frame K’. Then \dot{v}’ = v + \epsilon and
\[
|\dot{v}’|^2 = |v|^2 + 2v \cdot \epsilon + |\epsilon|^2,
\]
and so
\[
\mathcal{L}(|\dot{v}’|^2) = \mathcal{L}(|v|^2) + 2 \frac{\partial \mathcal{L}}{\partial |v|^2} v \cdot \epsilon + \mathcal{O}(|\epsilon|^2) \tag{3.8}
\]
to first order in \epsilon. However, since both frames are inertial the two Lagrangians should be equivalent for all \epsilon, which means the rightmost term of eq. (3.8) must be a total time derivative (cf. Corollary 3.6). Since the Lagrangian can only be a function of |v|^2, then this term could only be a total time derivative if it is linear in v. We therefore have that \partial \mathcal{L}/\partial |v|^2 is independent of v, and hence \mathcal{L} is proportional to |v|^2. This allows us to write
\[
\mathcal{L} = \frac{1}{2}m|v|^2 \tag{3.9}
\]
where m is the particle’s mass; we would experimentally observe that a particle’s acceleration is inversely proportional to its mass as in section 1.1. Note that m cannot be negative since looking at the action (3.1) we see that Hamilton’s principle would yield maxima instead of minima. We did not use the third type of Galilean transformations, but we can check that the expression (3.9) is invariant with respect to rectilinear motion \dot{v}’ = v + v_0:
\[
\mathcal{L}’(\dot{v}’) = \frac{1}{2}m|\dot{v}’|^2 = \frac{1}{2}m|\dot{v} + v_0|^2 = \frac{1}{2}m|v|^2 + m\dot{v} \cdot v_0 + \frac{1}{2}mv_0^2
\]
\[
= \mathcal{L}(v) + \frac{d}{dt} \left( mx \cdot v_0 + \frac{1}{2}mv_0^2 \right),
\]
and so the Euler–Lagrange equations are unaffected by Corollary 3.6.

For a system of free noninteracting particles, the Lagrangian for each individual particle cannot be dependent on the coordinates of any other, and so \mathcal{L} must be additive:
\[
\mathcal{L} = \sum_{i=1}^{N} \frac{1}{2}m_i|v_i|^2 = T, \tag{3.10}
\]
which is the kinetic energy as claimed. Note that (at least symbolically)
\[
|\dot{x}|^2 = \left( \frac{ds}{dt} \right)^2 = \frac{(ds)^2}{(dt)^2},
\]
and so we only need to know the line element $ds$ or metric $ds^2$ in order to transform the kinetic energy into other coordinate systems. If we wish to express the Cartesian coordinates $x_i = f_i(q_1, \ldots, q_n)$ in terms of generalized coordinates $q = (q_1, \ldots, q_n)$, then

$$T = \frac{1}{2} \sum_{i,j=1}^{n} a_{ij}(q) \dot{q}_i \dot{q}_j$$

(3.11)

where $a_{ij}$ are functions of the coordinates only. That is, the kinetic energy $T$ in generalized coordinates is still a quadratic function of velocities, but may also depend on the other coordinates. Mathematically, a conservative Lagrangian system is determined by a Riemann manifold, with a metric that determines the kinetic energy, and a potential function.

For a general system of particles which may interact we have to insert a function to the Lagrangian, which for a conservative system is the potential energy:

$$L(q, \dot{q}) = T(q, \dot{q}) - U(q).$$

(3.12)

Now that we are no longer in Euclidean space, we require that for a conservative system on a manifold $M$ the force (when viewed as a 1-form) be exact $F = -\nabla U$ (and not merely closed). Now the time reversibility (cf. section 1.5) is easily seen as the time independence of the Lagrangian—time reversal $t \mapsto -t$ in the quadratic kinetic energy (3.11) preserves each product $\dot{q}_i \dot{q}_j$. In other words, when the Lagrangian is time-independent the total energy is conserved (cf. section 1.5).

### 3.3 Equivalence to Newton’s equations

Now we will see how to obtain the Euler–Lagrange equations eq. (3.5) from Newtonian mechanics, so that the principle of least action is equivalent to Newton’s principle of determinacy and not stronger.

A mechanical system with a configuration manifold $M$ can always be embedded —and in experiment is naturally—embedded in some Euclidean space $x \in \mathbb{R}^N$. Within $M$, the motion of the system is dictated by some known force $F$. The effect of constraining the motion to the manifold $M$ can be thought of as a force $N$ orthogonal to $M$, called the constraint force. Newton’s equations for this system is

$$m_i \ddot{x}_i = F_i + N_i.$$  

(3.13)

Rearranging, we see that $m_i \ddot{x}_i - F_i = N_i$ is orthogonal to $M$, and so

$$(m_i \ddot{x}_i - F_i) \cdot \xi_i = 0$$

(3.14)

for all vectors $\xi = (\xi_1, \ldots, \xi_N)$ tangent to $M$. This is Newton’s equation in the tangent plane to the surface $M$. Summing over all particles, we get the d’Alembert-Lagrange principle,

$$\sum_{i=1}^{N} (m_i \ddot{x}_i - F_i) \cdot \xi_i = 0,$$

(3.15)
which dictates the motion of a system with constraints. Note that for a free system $M = \mathbb{R}^n$ we may take any vector $\xi \in \mathbb{R}^n$, and so we recover Newton’s equations.

Let $q = (q_1, \ldots, q_n)$ be local coordinates on $M$. Then by the chain rule

$$\dot{x}_i = \sum_{j=1}^{n} \frac{\partial x_i}{\partial q_j} \dot{q}_j,$$

and so we may write the kinetic energy

$$T(q, \dot{q}) = \sum_{i=1}^{N} \frac{1}{2} m_i |\dot{x}_i|^2 = \sum_{i,j=1}^{n} a_{ij}(q) \dot{q}_i \dot{q}_j$$

as a positive definite quadratic form on $M$.

Expressing the force in terms of the coordinates $q_j$, the covectors $Q_j$ defined by the 1-form equation

$$\sum_{i=1}^{N} F_i \, dx_i = \sum_{j=1}^{n} Q_j \, dq_j,$$  \hspace{1cm} (3.16)

or equivalently

$$Q_j = \sum_{i=1}^{N} F_i \frac{\partial x_i}{\partial q_j},$$  \hspace{1cm} (3.17)

are called the generalized forces. They dictate the evolution of the kinetic energy via the following expression.

**Proposition 3.9.** The Newtonian motion $q(t)$ of the system satisfies the equation

$$\frac{d}{dt} \frac{\partial T}{\partial \dot{q}} - \frac{\partial T}{\partial q} = Q.$$  \hspace{1cm} (3.18)

We will prove this in the next section.

Now that we are no longer in Euclidean space, we require that for a conservative system the 1-form $Q \, dq$ is not merely closed but also exact and may be written as $-dU$ for a potential energy $U(q)$. For such a system we are naturally led to define $\mathcal{L} = T - U$ so that

$$\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{q}} - \frac{\partial \mathcal{L}}{\partial q} = 0,$$  \hspace{1cm} (3.19)

and in section 3.1 we saw that this implies $q(t)$ is a critical point for the action functional. In this way, the principle of least action can be seen as a consequence of Newton’s equations.
Thus far we have considered Lagrangians of the form \( \mathcal{L} = T - U \), and the resulting system is automatically conservative since it is derived from the potential. In this section we will extend Lagrangian mechanics to include nonconservative systems. In the spirit of Proposition 3.9, we will see that the correct choice is

\[
\mathcal{L} = T + W, \tag{3.20}
\]

where \( W \) is the total work of the (generalized) forces \(3.17\): \( W = \sum_{i=1}^{n} Q_i q_i \).

As in the set up for the proof of Proposition 3.4, we can work in Euclidean space. Consider a system whose motion is described by the \( n \) coordinates \( q_j(t) \in \mathbb{R}^d \) from time \( t_1 \) to \( t_2 \), and write a fixed-endpoint variation of the motion as \( q_j(t) + \delta q_j(t) \) with \( \delta q_j(t_1) = \delta q_j(t_2) = 0 \). Repeating the integration by parts procedure from Proposition 3.4, we get

\[
\delta T = \int_{t_1}^{t_2} \sum_j \left( \frac{\partial T}{\partial q_j} \delta q_j + \frac{\partial T}{\partial \dot{q}_j} \delta \dot{q}_j \right) \, dt
\]

\[
= \left[ \sum_j \frac{\partial T}{\partial q_j} \delta q_j \right]_{t_1}^{t_2} + \int_{t_1}^{t_2} \sum_j \left( \frac{\partial T}{\partial q_j} - \frac{d}{dt} \frac{\partial T}{\partial \dot{q}_j} \right) \delta q_j \, dt
\]

\[
= \int_{t_1}^{t_2} \sum_j \left( \frac{\partial T}{\partial q_j} - \frac{d}{dt} \frac{\partial T}{\partial \dot{q}_j} \right) \delta q_j \, dt,
\]

where in the last equality we noted that \( \delta q_j(t_1) = \delta q_j(t_2) = 0 \) for an endpoint-fixed variation.

The quantity \( \delta W \) represents the work done by the forces on the system in taking the actual to the varied path \( q_j(t) + \delta q_j(t) \). Adding it in, the principle of least action for the new Lagrangian \(3.20\) yields

\[
0 = \delta \int_{t_1}^{t_2} L(q, \dot{q}, t) \, dt = \int_{t_1}^{t_2} (\delta T + \delta W) \, dt
\]

\[
= \int_{t_1}^{t_2} \left( \sum_j \left[ \frac{\partial T}{\partial q_j} - \frac{d}{dt} \frac{\partial T}{\partial \dot{q}_j} \right] \delta q_j + \sum_j Q_j \delta q_j \right) \, dt
\]

\[
= \int_{t_1}^{t_2} \sum_j \left( \frac{\partial T}{\partial q_j} - \frac{d}{dt} \frac{\partial T}{\partial \dot{q}_j} + Q_j \right) \delta q_j \, dt
\]

Since this must be true for all \( \delta q_j \) and each generalized coordinate \( q_j \) can be varied independently from the others, then we conclude

\[
\frac{d}{dt} \frac{\partial T}{\partial \dot{q}_j} - \frac{\partial T}{\partial q_j} = Q_j, \quad j = 1, \ldots, d. \tag{3.22}
\]
This is the extension of the Lagrange’s equations for nonconservative forces.

Nowhere did we assume that the generalized forces are not derivable from potentials, and thus as a sanity check that equations (3.22) reduce to the familiar formulation of Lagrange’s equations when the forces are conservative. In this case, we can use the same integration by parts procedure in reverse to obtain

\[
\int_{t_1}^{t_2} \delta W \, dt = \int_{t_1}^{t_2} \sum_j Q_j \delta q_j \, dt = -\int_{t_1}^{t_2} \sum_j \left[ \frac{\partial U}{\partial q_j} \delta q_j \right] \, dt
\]

\[
= \left[ \sum_j \frac{\partial U}{\partial q_j} \delta q_j \right]_{t_1}^{t_2} - \int_{t_1}^{t_2} \sum_j \left( \frac{\partial U}{\partial q_j} \delta q_j + \frac{\partial U}{\partial \dot{q}_j} \delta \dot{q}_j \right) \, dt
\]

\[
= -\int_{t_1}^{t_2} \sum_j \left( \frac{\partial U}{\partial q_j} \delta q_j + \frac{\partial U}{\partial \dot{q}_j} \delta \dot{q}_j \right) \, dt = -\int_{t_1}^{t_2} \delta U \, dt.
\]

In other words, \( \delta W = -\delta U \), and so the new Lagrangian (3.20) generates the same motion as the conservative Lagrangian \( \mathcal{L} = T - U \).

Even when the forces are nonconservative, all we needed for the previous paragraph is the ability to write the \( j \)th component of the generalized force as

\[
Q_j = -\frac{\partial U}{\partial q_j} + \frac{d}{dt} \frac{\partial U}{\partial \dot{q}_j}.
\] (3.23)

In this case, we recover the familiar form of Lagrange’s equations [3.5] with no right-hand side, but now with a velocity-dependent potential.

So \( U = T - \mathcal{L} \) is sometimes the potential energy even for nonconservative systems, and the potential energy is generally time-dependent. A common case is that the system \( A \) with coordinates \( q_A \) is not closed, but it moves in an external field due to a system \( B \) with coordinates \( q_B(t) \) independent of \( q_A \) and the entire system \( A + B \) is closed. This system has a Lagrangian of the form

\[
L = T_A(q_A, \dot{q}_A) - U(q_A, q_B(t)).
\] (3.24)

We may ignore \( T_B \) since it depends only on time and is thus a full time derivative. Equation (1.14) is a Lagrangian of the usual type, but with \( U \) being possibly time-dependent. If system \( A \) is a single particle, then the Euler–Lagrange equations yield

\[
m\ddot{q} = -\frac{\partial U}{\partial q}(q, t) = F(q, t).
\] (3.25)

As an example, if \( F \equiv F(t) \) is uniform (i.e. independent of position) then \( U = F \cdot x \) in Euclidean space.

### 3.5 Momentum and conservation

Mathematically there are two special cases when we can turn the second order Euler–Lagrange equations into first order equations, which physically correspond to conservation laws.
CHAPTER 3. EULER–LAGRANGE EQUATIONS OF MOTION

For a Lagrangian system, the (generalized) momentum of a particle is defined to be
\[ p_i = \frac{\partial L}{\partial \dot{q}_i}. \] (3.26)

If \( q_i = x_i \) is a Cartesian coordinate, then the Lagrangian has a term \( \frac{1}{2} m_i x_i^2 \) and so \( p_i = m_i x_i \) is the linear momentum along the \( x_i \)-axis. If \( q_i = \phi_i \) is the azimuthal angular coordinate in \( \mathbb{R}^3 \), then the Lagrangian has a term \( \frac{1}{2} m_i r_i^2 \dot{\phi}_i^2 \) and so \( p_i = m_i r_i^2 \dot{\phi}_i \) is the angular momentum about the axis perpendicular to the \( \phi \)-plane.

We can similarly define the (generalized) force as
\[ F_i = \frac{\partial L}{\partial q_i} = \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} = \frac{dp_i}{dt}, \] (3.27)
so that Newton’s equation is still satisfied symbolically. This definition agrees with the 1-form definition \((3.16)\) for \( Q_i \), and so both forces \( F_i \) and momenta \( p_i \) should be interpreted as covectors; conversely, the velocities \( \dot{q}_i \) are by definition tangent vectors. This fact is built into the kinetic energy:
\[ T_i = \frac{1}{2} v_i \cdot p_i = \frac{1}{2} \langle v_i, p_i \rangle, \] (3.28)
and so it is indeed natural and not merely a convention.

The Euler–Lagrange equations immediately imply that when the Lagrangian is independent of a generalized coordinate \( q_i \) then the force \( F_i \equiv 0 \) vanishes and the momentum \( p_i \) is conserved:
\[ \frac{dp_i}{dt} = \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} = \frac{\partial L}{\partial q_i} = 0. \] (3.29)
Mathematically, we can record this fact as saying:

**Corollary 3.10.** (Conservation of momentum) If the Lagrangian is independent of a variable \( q_i \), we can replace the second order Euler–Lagrange equations by the first order equation
\[ \frac{\partial L}{\partial \dot{q}_i} = \text{constant}. \]

Any such coordinate \( q_i \) is called cyclic. This includes the conservation laws we saw in sections 1.6 and 1.7. One of the strengths of Lagrangian mechanics is that when there is a cyclic coordinate then there is one less Euler–Lagrange equation to solve, which we already saw when conservation of angular momentum reduced a three-dimensional system to a two-dimensional system in section 2.2.

Geometrically, this requires the trajectories \( q(t) \) in configuration manifold \( M \) to lie in the level sets of \( p_i \).

In section 3.2 we mentioned that when the cyclic coordinate is time then the conserved quantity is the total energy. We can extend our definition of the *total*
3.6. NOETHER’S THEOREM

energy (or Hamiltonian $H$) to any system as

$$E = \sum_{i=1}^{d} \dot{q}_i p_i - \mathcal{L},$$

(3.30)

since for a conservative system on Euclidean space we have

$$E = \sum_i v_i \frac{\partial}{\partial v_i} \left[ \sum_j \frac{1}{2} m_j v_j^2 - U(x) \right] - \sum_i \left[ \frac{1}{2} m_i \dot{v}_i^2 - U(x) \right]$$

$$= \sum_i m_i \dot{v}_i^2 - \sum_i \frac{1}{2} m_i \dot{v}_i^2 + U = T + U.$$

More generally, we can show:

**Proposition 3.11** (Conservation of energy). The solution $q(t)$ to the Euler–Lagrange equation must satisfy the first order differential equation

$$H(q) := \dot{q} \cdot \frac{\partial \mathcal{L}}{\partial \dot{q}}(q, \dot{q}, t) - \mathcal{L}(q, \dot{q}, t) = \int_{t_0}^{t} \frac{\partial \mathcal{L}}{\partial \dot{q}}(q(s), \dot{q}(s), s) \, ds + \text{constant.}$$

In particular, if the Lagrangian $\mathcal{L}$ is independent of $t$, then the antiderivative on the right-hand side vanishes and the Hamiltonian $H(q)$ is constant.

In the context of variational calculus this is sometimes called the second Euler–Lagrange equation. Notice again the natural paring of the velocity $\dot{q}$ with the momentum $\frac{\partial \mathcal{L}}{\partial \dot{q}}$.

**Proof.** Using the chain rule and the Euler–Lagrange equations, we have

$$\frac{d \mathcal{L}}{dt} = \frac{\partial \mathcal{L}}{\partial t} + \frac{\partial \mathcal{L}}{\partial q_i} \dot{q}_i + \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \ddot{q}_i$$

$$= \frac{\partial \mathcal{L}}{\partial t} + \left( \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial q_i} \right) \dot{q}_i + \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \ddot{q}_i$$

$$= \frac{\partial \mathcal{L}}{\partial t} + \frac{d}{dt} \left( \dot{q}_i \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \right).$$

Integrating in time gives the desired equation for the Hamiltonian. \qed

3.6 Noether’s theorem

The conservation of momentum (Corollary 3.10) demonstrates that a system that is continuously symmetric along the coordinate $q_i$ possesses the corresponding conserved quantity $\partial \mathcal{L}/\partial \dot{q}_i$. We will now demonstrate that every one-parameter group of symmetries which preserves a Lagrangian system has a conserved quantity.
Consider a Lagrangian system consisting of a smooth \( n \)-dimensional manifold \( M \) with a time-independent Lagrangian \( L : TM \to \mathbb{R} \). We say that a diffeomorphism \( h : M \to M \) is a symmetry of the system if
\[
L(h_s(q, v)) = L(q, v) \quad \text{for all } (q, v) \in TM.
\] (3.31)

Here, \( h_s(q, v) = dh(q, v) \) is the pushforward or differential of the point \( q \in M \) and tangent vector \( v \in T_qM \), and is equal to the point \( h(q) \in M \) with the tangent vector \( dh_q(v) \in T_{h(q)}M \).

**Proposition 3.12** (Noether’s theorem). *If the time-independent Lagrangian system \((M, L)\) has a differentiable one-parameter group of diffeomorphisms \( h_s : M \to M, s \in \mathbb{R} \), then the Lagrangian motion of the system has a conserved quantity (or “integral”) \( I : TM \to \mathbb{R} \). In local coordinates,
\[
I(q, v) = \frac{\partial L}{\partial \dot{q}}(q, v) \cdot \frac{d}{ds} h^s(q) \bigg|_{s=0}.
\] (3.32)

**Remark.** The quantity \( I \) is independent of the choice of local coordinates \( q \). \( I \) measures the rate of change of \( L(q, v) \) as \( v \) is varied in \( T_qM \) in the direction of the tangent vector \( \frac{d}{ds}|_{s=0} h^s(q) \), which is why the formula (3.32) looks like the chain rule for \( \frac{d}{ds}|_{s=0} L((h^s)_*(q, v)) \), and this does not intrinsically involve a choice of coordinates.

**Proof.** As in the set up for the proof of the Euler–Lagrange equations (Proposition 3.4), we may cover \( M \) with coordinate patches and use a partition of unity to reduce the statement to Euclidean space \( \mathbb{R}^n \). Let \( \phi(t) := q(t) : \mathbb{R} \to M \) denote a solution to Lagrange’s equations. Since all of the symmetries \( h_s \) preserve the Lagrangian \( L \) then
\[
0 = \frac{\partial L}{\partial s}(\Phi(s, \dot{\Phi}(s))) = \frac{\partial L}{\partial q} \cdot \frac{\partial \Phi}{\partial s} + \frac{\partial L}{\partial \dot{q}} \cdot \frac{\partial \Phi}{\partial s},
\] (3.33)

where everything on the right-hand side is evaluated at \((\Phi(s, t), \dot{\Phi}(s, t)) \in T\mathbb{R}^n \). As we have already argued, for fixed \( s \) the map \( \Phi(s, \cdot) : \mathbb{R} \to \mathbb{R}^n \) satisfies Lagrange’s equation:
\[
\frac{\partial}{\partial t} \left[ \frac{\partial L}{\partial \dot{q}}(\Phi(s, t), \dot{\Phi}(s, t)) \right] = \frac{\partial L}{\partial q}(\Phi(s, t), \dot{\Phi}(s, t)).
\]
for each $s \in \mathbb{R}$. Plugging this into the right-hand side of eq. (3.33), we see that

$$
0 = \frac{\partial}{\partial t} \left[ \frac{\partial \mathcal{L}}{\partial \dot{q}} (\Phi, \dot{\Phi}) \right] \cdot \frac{\partial \Phi}{\partial s}(\Phi, \dot{\Phi}) + \frac{\partial \mathcal{L}}{\partial \dot{q}} (\Phi, \dot{\Phi}) \cdot \frac{\partial \dot{\Phi}}{\partial s}(\Phi, \dot{\Phi})
$$

by the chain rule and using that $\frac{\partial \dot{\Phi}}{\partial s} = \frac{\partial^2 \Phi}{\partial s \partial t} = \frac{d}{dt}(\frac{\partial \Phi}{\partial s})$.

**Example 3.13** (Translational symmetry). Consider the conservative $N$-particle Lagrangian

$$
\mathcal{L}(x, v) = \sum_{i=1}^{N} \frac{1}{2} m_i |v_i|^2 - U(x),
$$

where $X = (x_1, \ldots, x_N)$, $x_i \in \mathbb{R}^d$ and similarly for $v$. If the potential energy is invariant under translations along the first coordinate axis $e_1 \in \mathbb{R}^n$, then the system is symmetric with respect to the $N$ translations

$$
h_s^j : (\mathbb{R}^d)^N \rightarrow (\mathbb{R}^d)^N, \quad h_s^j(x_1, \ldots, x_N) = (x_1, \ldots, x_j + se_1, \ldots, x_N)
$$

for $j = 1, \ldots, N$. Noether’s theorem yields $N$ conserved quantities

$$
I_j(x, v) = \frac{\partial \mathcal{L}}{\partial \dot{x}} \cdot (0, \ldots, 0, e_1, 0, \ldots, 0) = m_j v_{j1},
$$

which we recognize as the first component of the $j$th particle’s momentum $p_j = m_j v_j$ (cf. Corollary 1.21).

**Example 3.14** (Rotational symmetry). Now suppose $d = 3$, and that the conservative $N$-particle Lagrangian (3.34) is invariant under rotations about the first coordinate axis $e_1 \in \mathbb{R}^3$. Then the system is symmetric with respect to the $N$ rotations

$$
h_s^j(x_1, \ldots, x_N) = (x_1, \ldots, \cos(s)x_j + \sin(s)e_1 \times x_j + (1 - \cos s)x_{j1}e_1, \ldots, x_N)
$$

for $j = 1, \ldots, N$. Noether’s theorem returns the $N$ conserved quantities

$$
I_j(x, v) = \frac{\partial \mathcal{L}}{\partial \dot{x}} \cdot (0, \ldots, 0, e_1 \times x_j, 0, \ldots, 0) = p_j \cdot (e_1 \times x_j) = (x_j \times p_j) \cdot e_1,
$$

which we recognize as the first component of the $j$th particle’s angular momentum $L_j = x_j \times p_j$ (cf. Corollary 1.24).

### 3.7 Exercises

**3.1** (Functional derivative is well-defined). Show that if $\Phi$ is a differentiable functional, then its differential is linear and is independent of the choice in variation of $\gamma$. 
3.2. Prove the following version of Proposition 3.4 on Euclidean space with weaker assumptions: If \( L \in C^1(\mathbb{R}^n \times \mathbb{R}^n \times [t_0, t_1]; \mathbb{R}) \) and \( q \in C^1([t_0, t_1]; \mathbb{R}^d) \) is a (fixed-endpoint) critical point for the action functional defined in eq. (3.1), then \( \partial L / \partial \dot{q}(q, \dot{q}, t) \) is a \( C^1 \) function on \([t_0, t_1]\) and \( q \) solves the Euler–Lagrange equations (3.3).

3.3. (Geodesics on the sphere [Tro96, Ch. 1]) In Example 3.7, we saw that the geodesics—paths of shortest length between two given points—in \( \mathbb{R}^d \) are straight lines. We will repeat this procedure for the sphere \( S^2 \).

(a) Using coordinates \((\phi, \theta)\), we can parameterize a path \( x(t) \) in \( S^2 \subset \mathbb{R}^3 \) as

\[
x(t) = (\cos \phi(t) \sin \theta(t), \sin \phi(t) \sin \theta(t), \cos \theta(t)), \quad t \in [0, 1]
\]

(\( \phi \) is the azimuth and \( \theta \) is the zenith angle). Find the formula for the arc length functional \( \Phi[x(t)] \).

(b) After rotating the sphere we may assume that \( \theta(0) = 0, \theta(1) = \theta_1, \) and \( \phi(1) = 0. \) Find a simple lower bound for \( \Phi[x(t)] \) using the \( \phi'(t) \) term. By considering when equality occurs, conclude that the geodesic connecting the north pole \( x(0) \) to the point \( x(1) \) is the shorter arc of the great circle (an equator, or circle of maximum circumference on the sphere) connecting the two points. This is another example where we can determine that the critical path for the functional is a minimum.

3.4. (Brachistochrone [Tro96, Ch. 6]) The brachistochrone between two points in the plane is the curve on which a frictionless bead would traverse the quickest subject to a downward gravitational acceleration. In 1696, Johann Bernoulli challenged mathematicians to find the shape of the brachistochrone, and it was his brother Jakob Bernoulli who provided a solution which was later refined into the calculus of variations.

(a) After translating, we may assume that the initial point is the origin \((0, 0)\) and the second point is given by some \((x_1, y_1)\) with \( x_1 > 0 \) and \( y_1 < 0. \) Explain why it is reasonable to assume that the brachistochrone is the graph of a function \( y(x), x \in [0, x_1] \) as opposed to a general parametric curve. Show that the time is takes the bead to traverse this curve is

\[
\Phi[y(x)] = \int_0^{x_1} \sqrt{1 + y'(x)^2} \frac{v(x)}{v(x)} \, dx
\]

where \( v(x) \) is the bead’s speed.

(b) With constant downward acceleration \( g \), show that \( v(x) = \sqrt{2gy(x)} \).

(c) Using the conservation of total energy (3.30), find the first order differential equation

\[
\sqrt{\frac{y}{c^2} - y'} = 1.
\]
3.7. EXERCISES

(d) Introducing a new dependent variable $\theta(x)$ so that

$$y = c^2 \sin^2 \frac{\theta}{2} = \frac{c^2}{2} (1 - \cos \theta), \quad 0 \leq \theta < 2\pi,$$

show that

$$\frac{c^2}{2} (1 - \cos \theta) \theta' = 1.$$

(e) By integrating the two equations of the previous part, obtain the parametric equations

$$x = c^2 (\theta - \sin \theta) + c_1, \quad 0 \leq \theta \leq \theta_1,$$
$$y = c^2 (1 - \cos \theta),$$

In order for $x(0) = 0 = y(0)$, we must have $c_1 = 0$. That is, the brachistochrone is a cycloid; these equations describe the path traced by a fixed point on a circle of radius $c^2$ as it rolls along the $x$-axis in the lower half-plane.

3.5. (Catenary [Tro96 Ch. 3]) The catenary is the shape taken by a cable hung by both ends under its own weight.

(a) Consider a cable of length $L$ hung between two equal height supports separated by a distance $H < L$. Let $y$ denote the vertical coordinate with $y = 0$ at the point where the cable is fastened, and let $y(s)$ denote the shape of the cable where $s$ is the arc length along the cable, so that $y(0) = 0 = y(L)$. If the weight per unit length is a constant $W$, explain why the cable shape $y(s)$ minimizes the mass integral

$$F[y(s)] = W \int_0^L y(s) \, ds.$$

If we had chosen the horizontal coordinate $x$ as the independent variable, the resulting functional is not convex.

(b) The functional with Lagrangian $(y, y', t) \mapsto Wy$ is only convex and may not have a unique minimizer. Show that in order to span the supports, the cable must satisfy the constraint

$$\int_0^L \sqrt{1 - y'(s)^2} \, ds = H.$$

(Note that $|y'(s)| = 1$ requires $x'(s) = 0$ which would produce a cusp.) The new Lagrangian

$$\mathcal{L}(y, y', t) = Wy - \lambda \sqrt{1 - y'^2}$$

is strictly convex for $\lambda > 0$. 
(c) Apply the Euler–Lagrange equations and integrate once to obtain the first order differential equation

\[ \frac{\lambda y'(s)}{\sqrt{1 - y'^2(s)}} = s + c \]

for the catenary, where \( c \) is a constant.

(d) Since we expect \( y(s) \) to be unique by convexity, then we may place additional assumptions. Assume that the cable shape is symmetric about the midpoint \( \ell = L/2 \), and conclude that \( y' (\ell) = 0 \) and \( c = -\ell \). Solve for \( y'(s) \) and integrate on \([0, s]\) to obtain

\[ y(s) = \sqrt{\lambda^2 + (\ell - s)^2} - \sqrt{\lambda^2 + \ell^2} \quad \text{on } [0, \ell]. \]

(e) Show that the constraint of part (b) yields

\[ \int_0^\ell \frac{\lambda}{\sqrt{\lambda^2 + (\ell - s)^2}} \, ds = \frac{H}{2}. \]

Make the substitution \((\ell - s) = \lambda \sinh \theta\), evaluate the integral, and conclude that

\[ \lambda = \frac{\ell}{\sinh(H/L)}. \]

(f) Show that \( x(s) \) is given by

\[ x(s) = \int_0^s \sqrt{1 - y'(t)^2} \, dt = \frac{H}{2} - \lambda \sinh^{-1} \left( \frac{\ell - s}{\lambda} \right). \]

Together with \( y(s) \) from (d) we have a parametric equation for the catenary. Eliminate the variable \( s \) and conclude

\[ y(x) = \lambda \cosh \left( \frac{x - H/2}{\lambda} \right) - \sqrt{\lambda^2 + \ell^2} \]

for \( x \in [0, H] \). That is, the catenary equation is hyperbolic cosine.

3.6. (Variational method for PDE [Evans10 Ch. 8]) In this exercise we will prove the existence of a solution to the elliptic divergence-form PDE

\[-\nabla \cdot (A(x) \nabla u(x)) = 0 \quad \text{for } x \in \Omega, \quad u(x) = 0 \quad \text{for } x \in \partial \Omega,\]

for an open set \( \Omega \subset \mathbb{R}^n \). Here, \( A(x) = (a_{ij}(x)) \) is a symmetric \( n \times n \) matrix with \( a_{ij} \in H^2(\Omega) \) (the Sobolev space), and we also assume that \( A \) is uniformly elliptic:

\[ \lambda I \preceq A(x) \preceq \Lambda I \quad \text{for } x \in \Omega \]

(in the sense of positive definite matrices).
3.7. Exercises

(a) For \( u \in H^1_0(\Omega) \) (the closure of \( C_\infty(\Omega) \) in \( H^1(\Omega) \)), show that the energy functional

\[
E[u] = \frac{1}{2} \int_\Omega \nabla u(x) \cdot A(x) \nabla u(x) \, dx
\]

is finite. Show that for \( \phi \in C_\infty(\Omega) \) the first variation at \( u \) is

\[
\lim_{\epsilon \to 0} \frac{E[u + \epsilon \phi] - E[u]}{\epsilon} = \int_\Omega \nabla \phi \cdot A \nabla u
\]

Consequently, if \( u \) is a minimum of \( E \), then the above expression vanishes for all \( \phi \in C_\infty(\Omega) \); such a function \( u \in H^1_0(\Omega) \) is called a weak solution of the PDE and boundary condition, since one formal integration by parts would produce the PDE but we do not assume \( u \) is twice differentiable.

(b) Let \( u_j \) be a sequence in \( H^1_0(\Omega) \) such that

\[
E[u_j] \to \inf_{H^1_0(\Omega)} E[u] \quad \text{as} \quad j \to \infty.
\]

Using Poincaré’s inequality, show that \( E[u] \) is bounded below and hence our sequence is bounded. Conclude that there exists a weakly convergent subsequence \( u_{jk} \to u \) in \( H^1(\Omega) \) using the Riesz representation theorem.

(c) Show that \( E \) is (sequentially) weakly lower semicontinuous:

\[
E[u] \leq \liminf_{k \to \infty} E[u_{jk}].
\]

Conclude that \( u \in H^1_0(\Omega) \) is a minimum for \( E \), and hence is a weak solution to the PDE by part (a).

3.7 (Two pendulums connected by a spring). Consider two pendulums of unit length and unit mass in a constant gravitational field \( g \). Suppose they are connected by a massless spring with spring constant \( k \) whose resting length is the same as their distance of separation.

(a) Let \( \theta_1 \) and \( \theta_2 \) be the angles the pendulums make with the downward vertical. Find the Lagrangian for the system for small angles, so that \( \sin \theta \approx \theta \).

(b) Define new variables

\[
q_1 = \frac{\theta_1 + \theta_2}{\sqrt{2}}, \quad q_2 = \frac{\theta_1 - \theta_2}{\sqrt{2}},
\]

and show that Lagrangian separates into two harmonic oscillators:

\[
\mathcal{L} = \frac{1}{2} (q_1^2 + q_2^2) + \frac{1}{2} (\omega_1^2 q_1^2 + \omega_2^2 q_2^2).
\]

Find \( \omega_1 \) and \( \omega_2 \). When \( q_2 = 0 \), we have \( \theta_1 = \theta_2 \) and so both pendulums swing in phase with each other with frequency \( \omega_1 \). When \( q_1 = 0 \), we have \( \theta_2 = -\theta_1 \) and so the pendulums swing in exact opposite phase with frequency \( \omega_2 \).
(c) For \( k \ll 1 \) we will see that an exchange of energy occurs. Suppose that at time \( t = 0 \) we have \( \theta_1 = 0 = \theta_2, \, \dot{\theta}_2 = 0, \) and \( \dot{\theta}_1 = v_0. \) Using part (b), show that the motion is given by

\[
\theta_1(t) = \frac{v_0}{2} \left( \sin t + \frac{1}{\omega} \sin \omega t \right), \quad \theta_2(t) = \frac{v_0}{2} \left( \sin t - \frac{1}{\omega} \sin \omega t \right)
\]

with \( \omega = \omega_2. \) For \( k \ll 1 \) we have \( 1/\omega \approx 1, \) and so

\[
\theta_1(t) \approx v_0 \cos \epsilon t \sin \omega' t, \quad \theta_2(t) \approx -v_0 \cos \omega' t \sin \epsilon t
\]

for some \( \epsilon \) and \( \omega'. \) Show that to leading order as \( k \to 0 \) we have

\[
\epsilon \approx \frac{k}{2}, \quad \omega' \approx 1,
\]

and so after a time \( T = \pi/2\epsilon \approx \pi/k \) the pendulums have switched roles and now essentially only the second pendulum is oscillating.

3.8 (Charged particle in an electromagnetic field). The motion of a charged particle is an example of a nonconservative Lagrangian system. A particle with charge \( q \) moving through the vector fields \( E, D, B, H \) on \( \mathbb{R}^3 \) with the scalar charge density \( \rho \) and vector current density \( j \) obeys Maxwell’s equations (in Gaussian units):

\[
\nabla \times E + \frac{1}{c} \frac{\partial B}{\partial t} = 0, \quad \nabla \cdot D = 4\pi \rho, \\
\nabla \times H - \frac{1}{c} \frac{\partial D}{\partial t} = \frac{4\pi}{c} j, \quad \nabla \cdot B = 0.
\]

The force on the charge is given by the Lorentz law

\[
F = q \left[ E + \frac{1}{c} (v \times B) \right].
\]

(a) The fourth Maxwell equation requires that \( B \) is divergence-free, and so we introduce a vector potential \( A \) for \( B \) so that \( B = \nabla \times A. \) Using the first equation, introduce a scalar potential \( \phi \) so that the electric field becomes

\[
E = -\nabla \phi - \frac{1}{c} \frac{\partial A}{\partial t}
\]

and the Lorentz force is

\[
F = q \left[ -\nabla \phi - \frac{1}{c} \frac{\partial A}{\partial t} + \frac{1}{c} (v \times (\nabla \times A)) \right].
\]

(b) The Lorentz force is nonconservative, but if we can put it in the form (3.23) then we will have a Lagrangian for this system. The first term \( q\phi \) is already
of the desired form. Show that the $x$-component of the rightmost term $v \times (\nabla \times A)$ may be rewritten as

$$(v \times (\nabla \times A))_x = \frac{\partial}{\partial x}(v \cdot A) - \frac{dA_x}{dt} + \frac{\partial A_x}{\partial t}.$$ 

By symmetry, we get the same relation for the other components with $x$ replaced by the respective variable.

(c) Show that the $x$-component of the Lorentz force can be written as

$$F_x = -\frac{\partial U}{\partial x} + \frac{d}{dt}\frac{\partial U}{\partial v_x},$$

for the potential energy

$$U = q\phi - \frac{q}{c}v \cdot A.$$ 

Symmetrically, the $y$- and $z$-components of the Lorentz force are also of this form if we replace $x$ with the respective coordinates. Consequently, the Lagrangian for this system is

$$\mathcal{L} = T - U = \frac{1}{2}mv^2 - q\phi + \frac{q}{c}v \cdot A$$

where $m$ is the particle’s mass.

3.9 (Noether’s theorem with time dependence). Consider a time-dependent Lagrangian system $\mathcal{L} : TM \times \mathbb{R} \to \mathbb{R}$.

(a) Prove Noether’s theorem for this system by applying Proposition 3.12 to the extended configuration space $M_1 = M \times \mathbb{R}$ with Lagrangian

$$\mathcal{L}_1 \left( q, t, \frac{dq}{d\tau}, \frac{dt}{d\tau} \right) = \mathcal{L} \left( q, \frac{dq}{dt/\tau}, \frac{dt}{d\tau} \right)$$

and the new “time” variable $\tau$, to obtain a conserved quantity $I(q, \dot{q}, t)$ on $M$.

(b) Apply this to a time-independent Lagrangian $\mathcal{L}(q, v, t) \equiv \mathcal{L}(q, v)$ to conclude that the total energy is conserved (cf. Proposition 1.6).
Chapter 4

Constraints

In this chapter we focus on the theoretical formulations of constraints in Lagrangian mechanics, without delving into the practical realizations of constraints; for an introduction to practical methods, see [AKN06, Sec. 1.6]. The treatment of holonomic constraints is based on [AKN06, Ch. 1], and [Gol51, Ch. 1–2] for non-holonomic constraints.

4.1 The d’Alembert–Lagrange principle

A holonomic constraint placed on a Lagrangian system requires that the system’s motion is confined to a submanifold $S$ of the phase space $TM$ that can be locally expressed in the form

$$f_1(q, \dot{q}, t) = \cdots = f_k(q, \dot{q}, t) = 0.$$  \hspace{1cm} (4.1)

A constraint that is not holonomic is called a non-holonomic constraint.

**Example 4.1.** Rigid-body motion in Euclidean space, in which the distances between all particles are fixed, is a holonomic constraint. This constraint may be expressed as

$$(x_i - x_j)^2 - c^2_{ij} = 0$$  \hspace{1cm} (4.2)

where $c_{ij}$ are the inter-particle distances. Conversely, the $j$th particle of a gas inside of a rigid spherical container of radius $a$ centered at the origin must obey the equation

$$r_j^2 - a^2 \leq 0,$$

which is non-holonomic. These two examples happen to be time-independent, but this need not be the case.

Tangent vectors $\xi \in T_q M$ which are tangent to the submanifold $S$ satisfy

$$\frac{\partial f_1}{\partial \dot{q}}(q, \dot{q}, t) \cdot \xi = \cdots = \frac{\partial f_k}{\partial \dot{q}}(q, \dot{q}, t) \cdot \xi = 0$$  \hspace{1cm} (4.3)
are called virtual velocities of the constrained motion at the state \((q, \dot{q}) \in TM\) and time \(t\). Analogous to the constraint forces of section 3.3 we can ensure that the motion \(q(t)\) is constrained to \(S\) by insisting that

\[
\left( \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{q}} - \frac{\partial \mathcal{L}}{\partial q} \right) \cdot \xi = 0 \tag{4.4}
\]

for all virtual velocities \(\xi\) at the state \((q(t), \dot{q}(t))\).

**Definition 4.2 (d’Alembert–Lagrange principle).** A motion of the Lagrangian system \((M, \mathcal{L})\) subject to the holonomic constraints (4.1) is a smooth trajectory solving eq. (4.4) for all virtual velocities \(\xi\).

By the definition (4.3) of the virtual velocities, this requires that the left-hand side of the Euler–Lagrange equations be within the span of the derivatives \(\partial f_i/\partial \dot{q}\):

\[
\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{q}} - \frac{\partial \mathcal{L}}{\partial q} = \sum_{i=1}^{k} \mu_i \frac{\partial f_i}{\partial \dot{q}} \tag{4.5}
\]

for some constants \(\mu_i\). These are called Lagrange’s equations with multipliers. Mathematically, we recognize the \(\mu_i\) as playing the role of Lagrangian multipliers: in order for the action functional to obtain a minimum on the submanifold \(S\), the gradient (the left-hand side in eq. (4.5)) of the action must be orthogonal to the submanifold. Physically, if we imagine that the particle system is confined to the submanifold \(S\) via imaginary constraint forces, then eq. (4.4) states that the constraint forces do no work and the right-hand side of eq. (4.5) gives a formula for our constraint forces.

We can also observe the constraint effect via the principle of least action and the space of paths. For a path \(\gamma \in \Omega\), let \(\Gamma\) be the subspace of the tangent space \(T_\gamma \Omega\) consisting of vectors \(W \in T_\gamma \Omega\) such that the vectors \(W_t\) are virtual velocities for all \(t\). H"{o}lder’s principle states that the effect of holonomic constraints on a Lagrangian system is given by a restricted principle of least action, and in this way holonomic constraints effectively reduce a system’s degrees of freedom.

**Corollary 4.3 (H"{o}lder’s principle).** A path satisfying the d’Alembert–Lagrange condition (4.4) is a motion of the Lagrangian system with constraints if and only if the functional derivative \(\delta S\) of the action vanishes on the subspace \(\Gamma\).

**Proof.** From Proposition 3.4 we recall that the first variation of the action \(S\) at the path \(q(t)\) in the direction of the fixed-endpoint variation \(h(t)\) is given by

\[
\delta S|_q \left( \frac{\partial H_s}{\partial s} \right) = \int_{t_0}^{t_1} \left( \frac{\partial \mathcal{L}}{\partial q} - \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{q}} \right) \cdot \frac{\partial H_s}{\partial s} \, dt.
\]

Restricting \(\delta S\) to the subspace \(\Gamma\) is equivalent to requiring that \(\frac{\partial H_s}{\partial s}\) is a virtual velocity. Since every virtual velocity could be attained this way, we see from the above variation that \(\delta S|_q\) vanishes on \(\Gamma\) if and only if the d’Alembert–Lagrange principle (4.4) is satisfied. \(\Box\)
4.2. GAUSS’ PRINCIPLE OF LEAST CONSTRAINT

Example 4.4. We can think of the pendulum of 3.8 as being a particle of mass $m$ in a vertical plane $(x, y)$ subject to the downward gravitational force with potential $U(x, y) = mg'y$ and subject to the holonomic constraints

$$x^2 + y^2 - \ell^2 = 0, \quad 2x\dot{x} + 2y\dot{y} = 0$$

where $\ell$ is the length of the pendulum arm. Note that the second condition is a time derivative of the first equation, but is necessary to specify the two-dimensional submanifold $S$ in the four-dimensional tangent space $T\mathbb{R}^2$. If we let $\theta$ denote the angle from the downward vertical as before and $r$ the distance from the pivot, then the Lagrangian is

$$\mathcal{L} = \frac{1}{2}m(r^2\dot{\theta}^2 + \dot{r}^2) + mgr\cos\theta$$

and we can rewrite the holonomic conditions as

$$f_1(r, \theta, \dot{r}, \dot{\theta}) = r^2 - \ell^2 = 0, \quad f_2(r, \theta, \dot{r}, \dot{\theta}) = \dot{r} = 0.$$

The first constraint does not place any restrictions on the virtual velocities $\xi = (\xi_r, \xi_\theta)$ since $\partial f_1 / \partial (\dot{r}, \dot{\theta}) = 0$, but the second condition yields

$$0 = \frac{\partial f_2}{\partial (\dot{r}, \dot{\theta})} \cdot \xi = \xi_r.$$

The d’Alembert–Lagrange principle (4.4) then yields the condition

$$(m\ddot{r} - mg\cos\theta)\xi_r + (mr^2\ddot{\theta} + 2mr\dot{r}\dot{\theta} + mgr\sin\theta)\xi_\theta = 0.$$

Since $\xi_r = 0$ and $\xi_\theta$ is arbitrary, we conclude that the second parenthetical term above vanishes. After using the constraints $f_1$ and $f_2$, we obtain the familiar equation of motion

$$\ddot{\theta} + \frac{g}{\ell}\sin\theta = 0.$$

The holonomic constraints have effectively discarded the $r$ equation and reduced the degrees of freedom from two to one.

4.2 Gauss’ principle of least constraint

Unlike the principle of least action, the d’Alembert–Lagrange principle is not a minimum principle because the virtual velocities are not given by a functional variation. Gauss sought a rephrasing as a minimization problem, in which the actual constrained motion is the constrained motion which deviates the least from the unconstrained motion.

The state of the system at time $t_0$ is given by the position $q_0 = q(t_0)$ and velocity $v_0 = \dot{q}(t_0)$ at a given moment while the acceleration $a(t_0) = \ddot{q}(t_0)$ is determined by the laws of motion and constraints. Consequently, we will consider $q_0$ and $v_0$ as fixed while we vary $a_0$. For a fixed state $(q_0, v_0) \in TM$
and time $t_0$, we will refer to all paths $q(t)$ allowed by the constraints with $q(t_0) = q_0$ and $\dot{q}(t_0) = \dot{v}_0$ as the conceivable motions; these lie in the constraint submanifold $S$ but do not satisfy an equation of motion. A released motion satisfies the unconstrained Euler–Lagrange equations (3.5) and does not lie in $S$, and an actual motion is a conceivable motion satisfying the d’Alembert–Lagrange principle (4.4) (and hence lies in $S$).

By means of a partition of unity, we may reduce to a neighborhood about the point $q_0$ and work in Euclidean coordinates and use the vector notation $q(t) = x(t)$. We will write an actual motion of the system $x_a(t) = x_r(t) + \delta x(t)$ as a deviation from the released motion $x_r(t)$, and we assume that the initial positions $x_a(t_0) = x_r(t_0) = x_0$ and velocities $\dot{x}_a(t_0) = \dot{x}_r(t_0) = \dot{v}_0$ are fixed. Taylor expanding we have

$$x_r(t) = x_0 + v_0(t - t_0) + \frac{1}{2} \ddot{x}_r(t)(t - t_0)^2 + \mathcal{O}((t - t_0)^3),$$

and if we replace $x_r(t)$ by $x_a(t)$ we only change $\delta x(t_0)$ and hence

$$\delta x(t) = \frac{1}{2} \delta \ddot{x}(t_0)(t - t_0)^2 + \mathcal{O}((t - t_0)^3).$$

(4.6)

So far, our considerations have been independent of the constraints.

In our local coordinates the d’Alembert–Lagrange principle requires

$$\left(\frac{d}{dt} \frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x}\right) \cdot \xi = 0$$

(4.7)

for all virtual displacements $\xi$. For simple physical systems $\frac{d}{dt} \frac{\partial L}{\partial \dot{x}}$ is the total force $m \ddot{x}$ and $\frac{\partial L}{\partial x}$ is the system’s force $-\nabla U$ without considering constraints, and so $\frac{d}{dt} \frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x}$ is the force $m \ddot{x}$ due to the constraints. We will now realize this more generally by evaluating $\frac{d}{dt} \frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x}$ at $x_a(t) = x_r(t) + \delta x(t)$. Taylor expanding about $x_r(t)$ we have

$$\left(\frac{d}{dt} \frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x}\right) \bigg|_{x=x_a} = \left(\frac{d}{dt} \frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x}\right) \bigg|_{x=x_r} + \left[\frac{d}{dt} \left(\frac{\partial^2 L}{\partial x^2} \delta \dot{x} + \frac{\partial^2 L}{\partial \dot{x} \partial x} \delta x\right) - \frac{\partial^2 L}{\partial x^2} \delta \dot{x} - \frac{\partial^2 L}{\partial \dot{x} \partial x} \delta x\right] \bigg|_{x=x_r} + \mathcal{O}(\delta x^2).$$

The first term on the right-hand side vanishes since the actual motion $x_a$ solves Lagrange’s equations. For the second term we use eq. (4.6) and take the limit $t \to t_0$ to obtain

$$\left(\frac{d}{dt} \frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x}\right) \bigg|_{x=x_a} \bigg|_{t=t_0} = \frac{\partial^2 L}{\partial x^2} \bigg|_{x=x_r} \bigg|_{t=t_0} \delta \ddot{x}(t_0).$$

Therefore, we interpret eq. (4.7) as saying that

$$\frac{\partial^2 L}{\partial q^2} \bigg|_{q=q_r} (\ddot{q}_a(t_0) - \ddot{q}_r(t_0))$$

(4.8)
is orthogonal to the constraint submanifold \( S \). On the other hand, eq. (4.8) is the gradient of the functional

\[
Z(q(t)) = \frac{1}{2}(\ddot{q} - \ddot{q}_r) \cdot M(\ddot{q} - \ddot{q}_r)\big|_{t=t_0}
\]  

for the Hessian matrix \( M = \frac{\partial^2 L}{\partial \dot{q}^2} \) of the kinetic energy quadratic form at \( x = x_r \) and \( t = t_0 \) (which for conservative systems is a positive definite matrix containing the particle masses). The quantity (4.9) is called Gauss' compulsion and it measures how much the motion \( q(t) \) deviates from the released motion.

The fact that the gradient (4.8) of the compulsion functional (4.9) for the actual motion \( q_a \) is orthogonal to the submanifold \( S \) is enough to conclude that the actual motion \( q_a \) is a critical point for the compulsion functional. This is a general fact regarding Lagrange multipliers, but we will pause now to sketch why this converse of the familiar Lagrange multiplier theorem is also true. Let \( f(q) \) be a functional on \( M \) with a critical point \( q_a \) when restricted to the submanifold \( S \). This happens if and only if \( D(f \circ \phi)|_{q_a} = 0 \) for all coordinates \( \phi \) from Euclidean space to a neighborhood of \( q_a \) in \( S \). Writing \( D(f \circ \phi) = \nabla f \cdot D\phi \) and noting that the columns of the matrix \( D\phi \) span the tangent space \( T_{q_a}S \subset T_qM \), we conclude that \( q_a \) is a critical point on \( S \) if and only if the gradient \( \nabla f(q_a) \in T_qM \) is orthogonal \( T_{q_a}S \).

Altogether, we have proven the following statement.

**Theorem 4.5 (Gauss' principle).** Among the conceivable motions, the actual motion \( q(t) \) is a critical point for the compulsion (4.9) with respect to the released motion. In particular, if \( M \) is positive definite then the actual motion is the global minimum for the compulsion with respect to the released motion.

Gauss’ principle is analogous to the method of least squares in regression analysis, another result discovered by Gauss. In the method of least squares, there is an unknown data correlation function determined by \( n \) parameters and a larger number \( N > n \) of observations, the latter of which deviate slightly from the desired function’s exact values due to observation error and hence the overdetermined system for the function is inconsistent. The remedy is to construct the square sum error between the function and the data, and then find the desired function as a error minimizer in the \( n \) parameters. Here, the compulsion is determined by the \( 2d \) initial conditions \((q_0, v_0) \in TM \) where \( d = \text{dim } M \) is the number of degrees of freedom, and the system is overdetermined due to the extra constraint conditions. Here, the actual motion \( q(t) \) plays the role of the function we seek determined by the smaller number of conditions \( \text{dim } TS < 2d \), the released motion is the measured data which over-determines the function with error, and the compulsion is the square sum error. Moreover, the matrix \( A \) of masses can be interpreted as weights in the method of least squares, which are inserted based on assumed reliability of the data accuracy.

See Exercise 4.1 for an example of Gauss’ principle.
4.3 Integrable constraints

Some authors further require that holonomic constraints be integrable (in the sense of Frobenius). To simplify the formulation of integrability, we will assume the constraints that are time-independent and linear in $\dot{q}$:

\[ a_1(q) \cdot \dot{q} = \cdots = a_k(q) \cdot \dot{q} = 0, \quad (4.10) \]

where the covectors $a_1(q), \ldots, a_k(q) \in T^*_q M$ are linearly independent for all $q \in M$.

Let $\omega_i$ denote the 1-form $a_i(q) \, dq$. We assume that the constraints (4.10) are completely integrable: the 2-forms $d\omega_i$ for $i = 1, \ldots, k$ vanish on the space of velocities satisfying the d’Alembert–Lagrange condition (4.4). It then follows from Frobenius’ theorem that any path $q(t)$ satisfying the d’Alembert–Lagrange condition (4.4) must lie within a smooth integrable submanifold $N$ of dimension $d - k$.

**Corollary 4.6.** A path satisfying the d’Alembert–Lagrange condition (4.4) is a motion of the Lagrangian system with constraints if and only if the path restricted to $N$ is a motion for the Lagrangian restricted to $N$.

**Proof.** From Hölder’s principle (Corollary 4.3) we know that the d’Alembert–Lagrange condition is equivalent to insisting that the action variation $\delta S$ vanishes on a subspace $\Gamma$ of conceivable variations. Here, $\Gamma$ is a subspace of the tangent space $T_{q(t)} \Omega_M$ to the paths $\Omega_M$ on $M$. On the other hand, the motion of the Lagrangian system on $N$ is given by taking the action $S$ on the paths $\Omega_N$ on $N$, and then insisting that its variation vanishes on the tangent space $T_{q(t)} \Omega_N$. The key observation is that for the integrable submanifold $N$, the tangent space $T_{q(t)} \Omega_N$ is equal to the subspace $\Gamma \subset T_{q(t)} \Omega_M$ and so the conditions become identical.

In particular, a holonomic system’s motion is determined by the restriction of the Lagrangian to the constraint submanifold $S$. In this way, a system with holonomic constraints is like a new mechanical system with fewer degrees of freedom, which distinguishes holonomic from non-holonomic constraints.

4.4 Non-holonomic constraints

In the derivation of Lagrange’s equations (Proposition 3.4), the last step relied upon the generalized coordinates $q_j$ being independent. As discussed in the previous section, such $q_j$ can always be chosen for a system subject to holonomic constraints. This is not always true in the non-holonomic case however, and so we wish to develop a new tool for this situation.

We will consider the special case of $m$ non-holonomic constraints that can be expressed as the vanishing of a one-form:

\[ \sum_{k=1}^{n} a_{\ell k} \, dq_k + a_{\ell t} \, dt = 0, \quad \ell = 1, \ldots, m, \quad (4.11) \]
with coefficients $a_{ij}$ dependent on $q_1, \ldots, q_n, t$. Note that velocity-independent holonomic constraints also fit this requirement, since when the condition (4.1) is independent of $\dot{q}$ then taking the differential of both sides gives

$$
\sum_{k=1}^{n} \frac{\partial f}{\partial q_k} dq_k + \frac{\partial f}{\partial t} dt = 0. 
$$

(4.12)

Consider endpoint-fixed variations of the system configurations at time $t_1$ and $t_2$ as in the proof of Proposition 3.4. The displacement $\delta q$ inside the time integral happens for fixed time, and thus the constraint (4.11) for a path variation becomes

$$
\sum_{k=1}^{n} a_{\ell k} dq_k = 0, \quad \ell = 1, \ldots, m. 
$$

(4.13)

If the system’s forces were otherwise conservative, the principle of least action requires

$$
\int_{t_1}^{t_2} \sum_{k=1}^{n} \left( \frac{\partial L}{\partial q_k} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_k} \right) \delta q_k dt = 0. 
$$

(4.14)

To make eq. (4.13) comparable to this, we multiply by the undetermined Lagrange multiplier coefficients $\lambda_\ell$, sum over $\ell$, and integrate from $t_1$ to $t_2$:

$$
\int_{t_1}^{t_2} \sum_{k=1}^{n} \sum_{\ell=1}^{m} \lambda_\ell a_{\ell k} \delta q_k dt = 0. 
$$

(4.15)

Adding this to equation (3.15) we arrive at

$$
\int_{t_1}^{t_2} \sum_{k=1}^{n} \left( \frac{\partial L}{\partial q_k} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_k} + \sum_{\ell=1}^{m} \lambda_\ell a_{\ell k} \right) \delta q_k dt = 0. 
$$

(4.16)

As we have discussed, the generalized coordinates $q_k$ are no longer independent since they are related by the $m$ constraints. However, the first $d - m$ coordinates can be chosen independently, and the remaining $m$ coordinates are determined by the conditions (4.13). Pick the multipliers $\lambda_\ell$ such that

$$
\frac{\partial L}{\partial q_k} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_k} + \sum_{\ell=1}^{m} \lambda_\ell a_{\ell k} = 0, \quad k = d - m + 1, \ldots, d. 
$$

(4.17)

This causes the last $m$ terms of the summation in the variation (4.16) to vanish, leaving us with

$$
\int_{t_1}^{t_2} \sum_{k=1}^{n-m} \left( \frac{\partial L}{\partial q_k} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_k} + \sum_{\ell=1}^{m} \lambda_\ell a_{\ell k} \right) \delta q_k dt = 0. 
$$

(4.18)

Since the first $d - m$ coordinates $q_k$ are independent, then we may conclude

$$
\frac{\partial L}{\partial q_k} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_k} + \sum_{\ell=1}^{m} \lambda_\ell a_{\ell k} = 0, \quad k = 1, \ldots, n - m. 
$$

(4.19)
Rearranging and including our requirements (4.17), we arrive at

\[
\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{q}_k} - \frac{\partial \mathcal{L}}{\partial q_k} = \sum_{\ell=1}^{m} \lambda_{\ell} a_{\ell k}, \quad k = 1, \ldots, d. \tag{4.20}
\]

This is the extension of Lagrange’s equations for non-holonomic constraints. Note that these are \(n\) equations in \(d + m\) unknowns—the \(n\) coordinates \(q_k\) and the \(m\) multipliers \(\lambda_{\ell}\)—and so we must also consider the \(m\) equations of restraint

\[
\sum_{k=1}^{d} a_{\ell k} \dot{q}_k + a_{\ell t} = 0, \quad \ell = 1, \ldots, m, \tag{4.21}
\]

obtained from (4.11) by dividing through by \(dt\). Comparing these new equations of motion (4.20) to the generalization for nonconservative forces (3.22), we observe that the quantities \(\sum_{\ell} \lambda_{\ell} a_{\ell k}\) are a manifestation of the constraint forces.

### 4.5 Exercises

4.1. Show explicitly that for the pendulum of Example 4.4 minimizing the compulsion \(Z\) leads to the familiar equation of motion.

4.2 (Hoop rolling down an inclined plane). Consider a circular disk of mass \(M\) and radius \(r\) rolling without slipping due to gravity down a stationary inclined plane of fixed inclination \(\phi\).

(a) In a vertical plane, the disk requires three coordinates: for example, two Cartesian coordinates \((x, y)\) for the center of mass and an angular coordinate to measure the disk’s rotation. If we pick the origin such that the surface of the inclined plane is the line \(y = r - (\tan \phi)x\), obtain a holonomic constraint of the form (4.1) for the center of mass corresponding to the disk sitting on the plane.

(b) Consequently, we now can pick two generalized coordinates to describe the disk’s motion: let \(x\) denote the distance of the disk’s point of contact and the top of the inclined plane, and \(\theta\) the angle through which the disk has rotated from its initial state. By considering the arc length through which the disk has rolled, show that rolling without slipping poses another holonomic constraint.

(c) In this case, it is easier to treat rolling without slipping as a non-holonomic constraint of the type in section 4.4

\[
r \, d\theta - dx = 0.
\]

Show that the Lagrangian for this system is

\[
\mathcal{L} = \frac{1}{2} M (\dot{\theta}^2 + r^2 \dot{\theta}^2) + Mg x \sin \phi.
\]
(d) Apply Lagrange’s equations of the form \((4.20)\) to determine the equations of motion. Here, \(\lambda\) is the force of friction that causes the disk to roll without slipping. Substituting in the differential equation of restraint
\[
r\dot{\theta} = \dot{x},
\]
conclude that
\[
\ddot{x} = \frac{g}{2} \sin \phi, \quad \ddot{\theta} = \frac{g}{2r} \sin \phi, \quad \lambda = \frac{Mg}{2} \sin \phi.
\]

4.3 (Solving Kepler’s problem using harmonic oscillators [KS65]). Consider the system of section 2.4 in which a particle \(x \in \mathbb{R}^3\) of mass \(m\) under the influence of a central potential
\[
U(|x|) = -\frac{Mm}{|x|}.
\]
The square function on \(\mathbb{C} \simeq \mathbb{R}^2\) given by
\[
u := u_1 + iu_2 = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \mapsto x := u^2 = \begin{pmatrix} u_1^2 - u_2^2 \\ 2u_1u_2 \end{pmatrix}
\]
is conformal except at the origin and maps conic sections centered at the origin to conic sections with one focus at the origin. Since Kepler’s motion is given by conic sections, we are inspired to apply a transformation with similar properties to our system in order to turn the elliptic orbits into simple harmonic oscillation. As a problem solving method this may seem idiosyncratic, but a similar transformation was used by Feynman to solve an open problem in physics.

Attached with the transformation \((4.22)\) is the linear transformation
\[
\begin{pmatrix} dx_1 \\ dx_2 \end{pmatrix} = 2 \begin{pmatrix} u_1 \\ u_2 \\ u_1 \\ u_2 \end{pmatrix} \begin{pmatrix} du_1 \\ du_2 \end{pmatrix}
\]
of differentials. The matrix of this transformation has the following key properties:

(a) the entries are linear homogeneous functions of the \(u_i\);
(b) the matrix is orthogonal, in the sense that
\bullet the dot product of any two different rows vanishes,
\bullet each row has the norm \(u_1^2 + u_2^2 + \cdots + u_n^2\).

We would like to find such a transformation of differentials on \(\mathbb{R}^n\). It turns out such transformations can only exist for \(n = 1, 2, 4,\) or \(8\). Ultimately we would like a transformation \(\mathbb{R}^n \to \mathbb{R}^3\), and so we take \(n = 4\) and choose such a matrix
\[
A = \begin{pmatrix}
u_1 & -u_2 & -u_3 & u_4 \\
u_2 & u_1 & -u_4 & -u_3 \\
u_3 & u_4 & u_1 & u_2 \\
u_4 & -u_3 & u_2 & -u_1
\end{pmatrix}
\]
which satisfies these properties. (This matrix also can be witnessed from quaternion multiplication.) Consequently, we set

$$
\begin{pmatrix}
\frac{dx_1}{dx_2} \\
\frac{dx_3}{dx_2} \\
0
\end{pmatrix} = 2A
\begin{pmatrix}
\frac{du_1}{du_2} \\
\frac{du_2}{du_3} \\
\frac{du_3}{du_4}
\end{pmatrix} = 2
\begin{pmatrix}
u_1 du_1 - u_2 du_2 - u_3 du_3 + u_4 du_4 \\
u_2 du_1 + u_1 du_2 - u_4 du_3 - u_3 du_4 \\
u_3 du_1 + u_4 du_2 + u_1 du_3 + u_2 du_4 \\
u_4 du_1 - u_3 du_2 + u_2 du_3 - u_1 du_4
\end{pmatrix}.
$$

(4.23)

Sadly, only the first three of these are complete differentials for three quantities:

$$
x_1 = u_1^2 - u_2^2 - u_3^2 + u_4^2
$$

$$
x_2 = 2(u_1u_2 - u_3u_4)
$$

$$
x_3 = 2(u_1u_3 + u_2u_4)
$$

(4.24)

and the fourth line of eq. 4.23) yields the non-holonomic constraint:

$$
u_4 du_1 - u_3 du_2 + u_2 du_3 - u_1 du_4 = 0.
$$

(4.25)

Explicit formulas exist for both the one-dimensional kernel and the inverse of the transformation (4.24).

(a) Let \( r = |x| \) denote the distance to the origin in \( \mathbb{R}^3 \). Show that

$$
u_1^2 + u_2^2 + u_3^2 + u_4^2 = r.
$$

Using the orthogonality of \( A \), invert the transformation (4.23) of differentials and conclude that for a fixed point \( u \in \mathbb{R}^4 \) our transformation conformally maps the space orthogonal to the kernel at \( u \) onto \( \mathbb{R}^3 \).

(b) Let \( u, v \in \mathbb{R}^4 \) be two orthonormal vectors which satisfy the condition

$$
u_4 v_1 - u_3 v_2 + u_2 v_3 - u_1 v_4 = 0
$$

derived from the constraint (4.25). From part (a) we know that the plane \( \text{span}\{u, v\} \) is mapped conformally onto a plane of \( \mathbb{R}^3 \). Show that when this transformation is restricted to \( \text{span}\{u, v\} \) onto its image, distances from the origin are squared and angles at the origin are doubled. (The calculation of the image \( x \) of a point in \( \text{span}\{u, v\} \) is rather lengthy, and to reach the conclusion it may help to compare the formula for \( x \) to the formula for Cayley–Klein parameters (see \[Gol51\] Sec. 4.5).)

(c) In particular, it follows that a conic section centered at the origin in the plane \( \text{span}\{u, v\} \subset \mathbb{R}^4 \) gets mapped to another conical section in \( \mathbb{R}^3 \) with one focus at the origin, the latter of which describes Kepler’s motion. Using part (b) and polar coordinates in the plane \( \text{span}\{u, v\} \), show that ellipses and hyperbolas in the plane \( \text{span}\{u, v\} \) centered at the origin are mapped to ellipses and hyperbolas respectively, and from the limit case conclude that a line in \( \text{span}\{u, v\} \) is mapped to a parabola.
(d) We want to take the motion \( x(t) \in \mathbb{R}^3 \) subject to the force \( P = (P_1, P_2, P_3) \) and transpose it into motion \( u(t) \in \mathbb{R}^4 \) subject to the force \( Q = (Q_1, Q_2, Q_3, Q_4) \) and the constraint \((4.25)\). Use the transformation of differentials \((4.23)\) to determine the kinetic energy \( T \) and the force \( Q \) in terms of the coordinates \( u \) and forces \( P \), and use the formula for the force \( Q \) to show that

\[
u_4Q_1 - u_3Q_2 + u_2Q_3 - u_1Q_4 = 0.
\] (4.26)

(e) Apply Lagrange’s equations \((3.22)\) for nonconservative forces to obtain the equations of motion for \( u(t) \). These equations have \( Q_i \) on the right-hand side, so add them together according to the identity \((4.26)\), simplify, and integrate once to obtain

\[
r(u_4\dot{u}_1 - u_3\dot{u}_2 + u_2\dot{u}_3 - u_1\dot{u}_4) = \text{constant}.
\] (4.27)

The parenthetical term (i.e. \( u_4\dot{u}_1 - u_3\dot{u}_2 + u_2\dot{u}_3 - u_1\dot{u}_4 \)) is exactly the constraint \((4.25)\) divided by \( dt \), and so to ensure that the constraint is upheld we pick the initial conditions for \( u \) and \( \dot{u} \) so that the parenthetical term vanishes initially, and hence for all time by the conserved quantity \((4.27)\). (Conversely, it can be shown that the equations of motion for \( u(t) \) and the condition \( u_4\dot{u}_1 - u_3\dot{u}_2 + u_2\dot{u}_3 - u_1\dot{u}_4 = 0 \) yield \( m\ddot{x} = P \).)

(f) The equations of motion for \( u \) have a singularity at \( r = 0 \). To deal with this, substitute the regularizing time \( s \)

\[
s = \int_0^t \frac{dt}{r}, \quad \frac{d}{dt} = \frac{1}{r} \frac{d}{ds}
\]

for the time variable \( t \) to regularize the equations of motion for \( u \). Plug in the Kepler forces \( Q_i = -\partial U(r)/\partial u_i \) and the (signed) semi-major axis

\[
a_0 = \left( \frac{2}{r} - \frac{v^2}{M} \right)^{-1}
\]

to arrive at

\[
\frac{\partial^2 u_i}{\partial s^2} + \frac{M}{4a_0} u_i = 0.
\]

That is, the preimage of bounded orbits \((a_0 > 0)\) under this transformation is simple harmonic motion in \( \mathbb{R}^4 \) with frequency \( \omega = \sqrt{M/4a_0} \). The harmonic motion can be painstakingly transformed into a solution \( u(t) \) by computing and substituting the physical time \( t = \int_0^s r(s) \, ds \).
Chapter 5

Hamilton–Jacobi equation of motion

Rather than \( n \) second-order ODEs, the Hamilton–Jacobi equation is one partial differential equation in \( 2d + 1 \) variables. This perspective is sometimes less powerful when solving for the motion, but it provides a physical interpretation of the previously abstract action functional. The material for this chapter is based on [LL76, Ch. 7] and [Gol51, Ch. 9].

5.1 Hamilton–Jacobi equation

Let us return to the action functional. In section 3.1 we considered the variation of the action \( S(q) \) at the system’s motion \( q(t) \) for paths with fixed endpoints. As in the set up for the proof of the Euler–Lagrange equations (Proposition 3.4) we can reduce to an open subset of Euclidean space, where the variation of the motion \( q(t) \) between times \( t_1 \) and \( t_2 \) is given by \( q(t) + \delta q(t) \). Previously we assumed \( \delta q(t_1) = \delta q(t_2) = 0 \) and then prescribed the resulting variation \( \delta S \) to vanish in accordance with the principle of least action. Now we will consider the action of the true progression of the system \( S[q(t)] \) as a function of the coordinates at time \( t_2 \) by allowing \( \delta q(t_2) \) (and consequently \( \delta S \)) to vary.

Repeating the calculation of the action variation from Proposition 3.4, we now have

\[
\delta S(q; \delta q) = \left[ \frac{\partial L}{\partial \dot{q}} \cdot \delta q \right]_{t_0}^{t_1} + \int_{t_0}^{t_1} \left( \frac{\partial L}{\partial q} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}} \right) \cdot \delta q \, dt = \frac{\partial L}{\partial \dot{q}}(t_2) \cdot q(t_2).
\]

The integral vanishes due to Lagrange’s equations of motion, but now only one of the boundary terms vanish since \( \delta q(t_1) = 0 \). If we let \( t := t_2 \) vary now, then we conclude

\[
\frac{\partial S}{\partial q_i} = \frac{\partial L}{\partial \dot{q}_i} = p_i
\]

(5.1)
after recognizing $\partial L/\partial \dot{q}_i$ as the momentum $p_i$.

By definition of the action we have
\[
\frac{dS}{dt} = L.
\]  

(5.2)

On the other hand we know $S = S[t,q(t)]$, and so the chain rule insists
\[
\frac{dS}{dt} = \frac{\partial S}{\partial t} + \frac{\partial S}{\partial q} \cdot \dot{q} = \frac{\partial S}{\partial t} + p \cdot \dot{q}
\]  

(5.3)

after using the derivative (5.1). Setting (5.2) and (5.3) equal yields
\[
\frac{\partial S}{\partial t} = L - p \cdot \dot{q} = -H,
\]

where $H$ is the Hamiltonian (or total energy) (3.30). Using the derivative (5.1) we recognize this identity as a first order partial differential equation (PDE) for $S(t,q)$:
\[
0 = \frac{\partial S}{\partial t} + H \left( t, q, \frac{\partial S}{\partial q} \right).
\]

(5.4)

This is the Hamilton–Jacobi equation, and for a system with $n$ degrees of freedom there are $n+1$ independent variables $(t,q_1,\ldots,q_n)$. It is often the case in practice that the form of the action $S(t,q)$ is unknown, and cannot be determined from eq. (5.4) alone.

The solution, or complete integral, of this equation has $n+1$ integration constants corresponding to the number of independent variables. Calling these constants $\alpha_1,\ldots,\alpha_n$, and $A$ (which play a physical role as we will see), we can write
\[
S = f(t,q_1,\ldots,q_n,\alpha_1,\ldots,\alpha_n) + A;
\]

we know one of these constants can be additive since the action appears in the PDE (5.4) only through its partial derivatives and hence is invariant under the addition of a constant. Mathematically speaking, a solution $S(t,q,\alpha)$ to the Hamilton–Jacobi equation (5.4) should be required to satisfy
\[
\text{det} \frac{\partial^2 S}{\partial q \partial \alpha} \neq 0
\]
to be a complete integral, in order to avoid incomplete solutions.

The function $f(t,q_1,\ldots,q_n,\alpha_1,\ldots,\alpha_n)$ (called a generating function) induces a change of coordinates (a canonical transformation)—we will develop this idea more from the Hamiltonian perspective. Think of $\alpha_1,\ldots,\alpha_n$ as new momenta, and let $\beta_1,\ldots,\beta_n$ denote new coordinates to be chosen. Note that by the chain rule,
\[
\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial q} \cdot \dot{q} + \frac{\partial f}{\partial \alpha} \cdot \dot{\alpha}.
\]

(5.5)

Set
\[
\beta_i := -\frac{\partial f}{\partial \alpha_i},
\]

(5.6)
5.1. HAMILTON–JACOBI EQUATION

and note that \( p_i = \partial f/\partial q_i \) by eq. (5.1). Consider the new Hamiltonian

\[
H'(\alpha, \beta, t) := H + \frac{\partial f}{\partial t} = H + \frac{\partial S}{\partial t} = 0,
\]

which has corresponding Lagrangian

\[
\mathcal{L}'(\beta, \alpha, t) := \alpha \cdot \dot{\beta} - H' = \alpha \cdot \dot{\beta} = -\frac{df}{dt} + \frac{\partial f}{\partial t} + \frac{\partial f}{\partial q} \cdot \dot{q}
\]

by the chain rule (5.5). Since this new Lagrangian \( \mathcal{L}' \) differs from the old \( \mathcal{L} \) by a complete time derivative, then they generate the same motion by Corollary 3.6.

For any Lagrangian \( \mathcal{L} \), the definition of the Hamiltonian (3.30) requires

\[
\frac{\partial H}{\partial p} = \frac{\partial}{\partial p} (p \cdot \dot{q} - \mathcal{L}(q, \dot{q}, t)) = \dot{q},
\]

\[
\frac{\partial H}{\partial q} = \frac{\partial}{\partial q} (p \cdot \dot{q} - \mathcal{L}(q, \dot{q}, t)) = -\frac{\partial \mathcal{L}}{\partial q} = -\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{q}} = -\dot{p}.
\]

This calculation may appear questionable at first (is the velocity \( \dot{q} \) really independent of the position \( q \) and momentum \( p \)), but it is indeed correct; in fact, we will later give it thorough justification as it is the basis of Hamiltonian mechanics. Applying this to our new Hamiltonian \( H'(\beta, \alpha, t) \), we observe that

\[
\dot{\alpha} = 0, \quad \dot{\beta} = 0.
\]

In other words, we have used the integration constants \( \alpha \) to generate corresponding conserved quantities \( \beta \). Recalling the definition (5.6) of the coordinates \( \beta \), we conclude

\[
\frac{\partial S}{\partial \alpha_i} = \text{constant} \quad (5.7)
\]

in the \( n \) coordinates and time. Note that even if we can obtain a partial solution to the Hamilton–Jacobi equation involving only \( m \) constants \( \alpha_i \), we still get the corresponding \( m \) constants of motion.

Lastly, let us specialize to a conservative system. In this case, the Hamiltonian \( H \) is the total energy \( E \) and is constant and the action integral from times \( 0 \) to \( t \) becomes:

\[
S[q(t)] = \int_{t_0}^t \mathcal{L}(q(t), \dot{q}(t), t) \, dt = \int_{t_0}^t (p \cdot \dot{q} - H) \, dt
\]

\[
= \int_{t_0}^t p \cdot \dot{q} \, dt - \int_{t_0}^t E \, dt = S_0(q_1, \ldots, q_n) - Et
\]

for some \( S_0(q) \) which is only a function of the coordinates. Since \( \partial S_0/\partial t = 0 \), the Hamilton–Jacobi equation then takes the special form

\[
H \left( q, \frac{\partial S}{\partial q}, t \right) = E. \quad (5.8)
\]

In particular, we are no longer required to know the formula for the action.
5.2 Separation of variables

Sometimes we can reduce the Hamilton–Jacobi equation by one coordinate using separation of variables. Suppose that for some system with \( n \) degrees of freedom, we have a coordinate \( q_1 \) with corresponding derivative \( \partial S/\partial q_1 \) that appear in the Hamilton–Jacobi equation only in some particular combination \( \phi(q_1, \partial S/\partial q_1) \). That is, the Hamilton–Jacobi equation (5.4) takes the form

\[
\Phi \left( t, q_2, \ldots, q_n, \frac{\partial S}{\partial q_2}, \ldots, \frac{\partial S}{\partial q_n}, \phi \left( q_1, \frac{\partial S}{\partial q_1} \right) \right) = 0 \tag{5.9}
\]

after rearranging the independent variables. We take the (additive) separation of variables ansatz

\[
S = S'(q_2, \ldots, q_n, t) + S_1(q_1)
\]

for the solution. Plugging this back into our Hamilton–Jacobi equation (5.9) we get

\[
\Phi \left( t, q_2, \ldots, q_n, \frac{\partial S'}{\partial q_2}, \ldots, \frac{\partial S'}{\partial q_n}, \phi \left( q_1, \frac{dS_1}{dq_1} \right) \right) = 0. \tag{5.10}
\]

Note that \( q_1 \) only influences \( \phi \) and is entirely independent from the rest of the expression. Since the variables are independent, we conclude that \( \phi \) must be constant:

\[
\phi \left( q_1, \frac{dS_1}{dq_1} \right) = \alpha_1 \quad \Phi \left( t, q_2, \ldots, q_n, \frac{\partial S'}{\partial q_2}, \ldots, \frac{\partial S'}{\partial q_n}, \alpha_1 \right) = 0. \tag{5.11}
\]

We now have a first order ODE and the Hamilton–Jacobi equation in terms of the remaining \( n - 1 \) coordinates. Since the resulting form of \( \Phi \) is coherent with what we found in the previous section, then we conclude that our ansatz is correct. The ability to remove a coordinate in the Hamilton–Jacobi consideration even when it is not cyclic is a virtue of this approach.

If we do have some cyclic coordinate \( q_1 \) for the system (i.e. \( S \) or \( \mathcal{L} \) is independent of \( q_1 \)), then the Hamilton–Jacobi equation (5.4) becomes

\[
\frac{\partial S}{\partial t} + H(q_2, \ldots, q_n, \frac{\partial S}{\partial q_1}, \ldots, \frac{\partial S}{\partial q_n}, t) = 0. \tag{5.12}
\]

Then \( q_1 \) is of the type (5.9) for the function

\[
\phi \left( q_1, \frac{\partial S}{\partial q_1} \right) = \frac{\partial S}{\partial q_1}. \tag{5.13}
\]

Our result (5.11) yields

\[
\frac{\partial S}{\partial q_1} = \alpha_1, \quad S = S'(q_2, \ldots, q_n, t) + \alpha_1 q_1. \tag{5.14}
\]

Here, \( \alpha_1 = \partial S/\partial q_1 \) is just the conserved momentum corresponding to \( q_1 \).
If this cyclic variable is time, then the system is conservative and we saw in eq. (5.8) that the action is given by

\[ S = W(q_1, \ldots, q_n) - Et. \]  

(5.15)

Here, \(-E\) is the conserved quantity associated with \(t\), although it is not necessarily always the total energy. We have used \(W\) instead of \(S'\) to denote this special case, but they share the same purpose; \(W\) is known as the Hamilton’s characteristic function. The Hamilton–Jacobi equation is now

\[ H \left(q_1, \ldots, q_n, \frac{\partial W}{\partial q_1}, \ldots, \frac{\partial W}{\partial q_n} \right) = -\frac{\partial S}{\partial t} = E, \]  

(5.16)

and \(E\) is just one integration constant \(\alpha_j\) of the motion of which \(S'\) is independent. The corresponding conserved quantity \(\beta_i\) will give the coordinates implicitly as functions of the constants \(\alpha_i, \beta_i, \) and time:

\[ \beta_i = \frac{\partial S}{\partial \alpha_i} = \begin{cases} \frac{\partial W}{\partial \alpha_i} - t & \text{for } i = j, \\ \frac{\partial W}{\partial \alpha_i} & \text{for } i \neq j. \end{cases} \]  

(5.17)

Only the \(j\)th of these equations is time-dependent, and so one of the \(q_i\) can be chosen as an independent variable and the rest will be able to be written in terms of this coordinate. Such solutions for the motion are called orbit equations. For the central forces of section 2.2 we were able to solve for the angle \(\phi\) as a function of the radius \(r\).

### 5.3 Conditionally periodic motion

For this section, we will examine a system of \(n\) degrees of freedom with bounded motion, such that every variable can be separated using the method of the previous section. This means the action takes the form

\[ S = \sum_{i=1}^{d} S_i(q_i) + S'(t), \]  

(5.18)

where each coordinate \(S_i\) is related to the corresponding momentum via

\[ p_i = \frac{\partial S_i}{\partial q_i}, \quad S_i = \int p_i \, dq_i. \]  

(5.19)

The motion is bounded, and so this integral represents the area enclosed by a loop in the phase plane \((q_i, p_i)\) (just as with the period integral \(2.4\)). Every time \(q_i\) returns to a value, \(S_i\) has incremented by an amount \(2\pi I_i\) with

\[ I_i = \frac{1}{2\pi} \oint p_i \, dq_i. \]  

(5.20)
These $I_i$ are called the action variables.

As in section 5.1, the generating function $S_i$ induces a canonical transformation of coordinates with the action variables $I_i$ as the new momenta. We will see this process more systematically from the Hamiltonian perspective. The new position coordinates, called angle variables, are given by

$$w_i = \frac{\partial S}{\partial I_i} = \sum_{k=1}^{n} \frac{\partial S_k(q_k, I_1, \ldots, I_n)}{\partial I_i}$$ (5.21)

since this is the time integral of “$\partial H/\partial p = \dot{q}$”. The generating functions $S_i$ are time-independent, and so the new Hamiltonian $H = E$ is just the old in terms of the new coordinates. The (Hamiltonian) equations of motion require

$$\dot{I}_i = -\frac{\partial H(I_1, \ldots, I_n)}{\partial w_i} = 0, \quad \dot{w}_i = \frac{\partial H(I_1, \ldots, I_n)}{\partial I_i} = \frac{\partial E(I_1, \ldots, I_n)}{\partial I_i},$$ (5.22)

which can be immediately integrated to yield

$$I_i = \text{constant}, \quad w_i = \frac{\partial E}{\partial I_i} t + \text{constant}.$$ (5.23)

As we have already observed, $S_i$ increments by $2\pi I_i$ each time $q_i$ returns to its original value, and so the angle variables $w_i$ also increment by $2\pi$. Consequently the derivative $\partial E/\partial I_i$ is the frequency of motion in the $i$th coordinate, which we were able to identify without solving the entire system.

Since the motion in these variables is periodic, any single-valued function $F(q, p)$ of the system coordinates and momenta will be periodic in the angle variables with period $2\pi$ after being transformed to the canonical variables. Fourier expanding in each of the angle variables, we have

$$F = \sum_{\ell \in \mathbb{Z}^n} A_{\ell} e^{i\ell \cdot w} = \sum_{\ell_1 \in \mathbb{Z}} \cdots \sum_{\ell_n \in \mathbb{Z}} A_{\ell_1 \cdots \ell_n} e^{i(\ell_1 w_1 + \cdots + \ell_n w_n)}.$$ (5.24)

Using eq. (5.23), we may write the angle variables as functions of time. Absorbing the integration constants of the $w_i$ into the coefficients $A_{\ell}$, we get

$$F = \sum_{\ell \in \mathbb{Z}^n} A_{\ell} \exp \left( it \ell \cdot \frac{\partial E}{\partial I} \right).$$ (5.25)

Each term of this sum is periodic with frequency $\ell \cdot (\partial E/\partial I)$. If the frequencies $\partial E/\partial I_i$ are not commensurable however, then the total quantity $F$ is not periodic. In particular, the coordinates $q, p$ are may not be periodic, and the system may not return to any given state that it instantaneously occupies. However, if we wait long enough the system will come arbitrarily close to any given occupied state—this phenomenon is referred to as Poincaré’s recurrence theorem. Such motion is called conditionally periodic.

Two frequencies $\partial E/\partial I_i$ that are commensurable (i.e. their ratio is a rational number) are called a degeneracy of the system, and if all $n$ are commensurable
5.4. GEOMETRIC OPTICS ANALOGY

the system is said to be completely degenerate. In the latter case, all motion is periodic, and so we must have a full set of $2n - 1$ conserved quantities. Only $n$ of these will be independent, and so they can be defined to be the action variables $I_1, \ldots, I_n$. The remaining $n - 1$ constants may be chosen to be $w_i \partial E/\partial I_k - w_k \partial E/\partial I_i$ for distinct $i, k$, since

$$\frac{d}{dt} \left( w_i \frac{\partial E}{\partial I_k} - w_k \frac{\partial E}{\partial I_i} \right) = \dot{w}_i \frac{\partial E}{\partial I_k} - \dot{w}_k \frac{\partial E}{\partial I_i} = \frac{\partial E}{\partial I_i} \frac{\partial E}{\partial I_k} - \frac{\partial E}{\partial I_k} \frac{\partial E}{\partial I_i} = 0.$$  

Note, however, that since the angle variables are not single-valued, neither will be the $n - 1$ constants of motion.

Consider a partial degeneracy, say, of frequencies 1 and 2. This means

$$k_1 \frac{\partial E}{\partial I_1} = k_2 \frac{\partial E}{\partial I_2}$$  \hspace{1cm} (5.26)

for some $k_1, k_2 \in \mathbb{Z}$. The quantity $w_1 k_2 - w_2 k_1$ will then be conserved, since

$$\frac{d}{dt} (w_1 k_1 - w_2 k_2) = \dot{w}_1 k_1 - \dot{w}_2 k_2 = \frac{\partial E}{\partial I_1} k_1 - \frac{\partial E}{\partial I_2} k_2 = 0.$$  \hspace{1cm} (5.27)

Note that this quantity is single-valued modulus $2\pi$, and so a trigonometric function of it will be an actual conserved quantity.

In general, for a system with $n$ degrees of freedom whose action is totally separable and has $n$ single-valued integrals of motion, the system state moves densely in a $n$-dimensional manifold in $2d$-dimensional phase space. For degenerate systems we have more than $n$ integrals of motion, and consequently the system state is confined to a manifold of dimension less than $n$. When a system has less than $n$ degeneracies, then there has fewer than $n$ integrals of motion and the system state travels within a manifold of dimension greater than $n$.

5.4 Geometric optics analogy

In this section we will see that the level sets of the action propagate through configuration space mathematically similar to how light travels through a medium. This is not meant literally, but rather to bring a physical analogy to the formerly abstract notion of the action.

Suppose we have a system for which the Hamiltonian is conserved and equal to the total energy. For simplicity, suppose the system describes just one particle moving in three-dimensional Euclidean space using Cartesian coordinates $q = x = (x, y, z) \in \mathbb{R}^3$; much of the results we will show however hold without these assumptions. Equation \(5.15\) tells us that the action is given by

$$S(q, t) = W(q) - Et.$$  \hspace{1cm} (5.28)

Since we have chosen Cartesian coordinates, rather than discussing the particle’s motion we may discuss the motion of the action level surfaces $S(q, t) = b$ in
time within the same space. If we were to generalize this argument to multiple-particle systems, then instead of the particle’s motion in Cartesian space we must consider the path that the system traces out in configuration space—the space spanned by the coordinates \( q \). At time \( t = 0 \), we have an equation for Hamilton’s characteristic function \( W = b \), and after a time step \( \Delta t \) we then have \( W + \Delta W = b + E\Delta t + \mathcal{O}(\Delta t^2) \).

The propagation of this surface can be thought of as a wavefront. The change in the characteristic function \( W \) during the time interval \( dt \) is \( dW = E dt \). If we call the distance traveled normal to the wavefront \( ds \), then we also have

\[
\frac{\partial W}{\partial s} = |\nabla W|. \tag{5.29}
\]

The velocity \( u \) of the wavefront is then

\[
u = \frac{ds}{dt} = \frac{dW/|\nabla W|}{dW/E} = \frac{E}{|\nabla W|}. \tag{5.30}\]

As a conservative system, the Hamilton–Jacobi equation takes the form (5.16), which in Cartesian coordinates looks like

\[
E = H \left( q, p = \frac{\partial W}{\partial q} \right) = T \left( p = \frac{\partial W}{\partial q} \right) + U = \frac{(\nabla W)^2}{2m} + U,
\]

or after rearranging,

\[
(\nabla W)^2 = 2m(E - U). \tag{5.31}
\]

Plugging this into the velocity (5.29), we have

\[
u = \frac{E}{\sqrt{2m(E - U)}} = \frac{E}{\sqrt{2mT}} = \frac{E}{p}. \tag{5.32}
\]

The faster the particle moves, the slower the action level sets propagate.

The momentum is given by

\[
p = \frac{\partial W}{\partial q} = \nabla W. \tag{5.33}
\]

The gradient is of course normal to the level sets, and so this relation tells us that the particle always moves normal to the level sets of the characteristic function \( W \).

We will now see how the level sets of the action propagate like waves. For some scalar-valued function \( \phi \), the wave equation of optics is

\[
\nabla^2 \phi - \frac{n^2}{c^2} \frac{\partial^2 \phi}{\partial t^2} = 0. \tag{5.34}
\]

If the refractive index \( n \) is constant, then there is a family of plane wave solutions:

\[
\phi(r, t) = \phi_0 e^{i(k \cdot x - \omega t)}, \tag{5.35}
\]
where the vector \( \mathbf{k} \in \mathbb{R}^3 \) is the propagation direction and the magnitude of \( k = 2\pi/\lambda = n\omega/c \) is the wave number.

In geometric optics, the refractive index \( n \equiv n(x) \) is not assumed to be constant, but rather to be changing slowly on the scale of the wavelength. Consequently, we now seek solutions to the wave equation (5.34) with the plane wave ansatz (5.35):

\[
\phi(r,t) = e^{A(x) + ik_0 (L(x) - ct)}.
\] (5.36)

where \( A(x) \) is related to the amplitude of the wave, \( k_0 = \omega/c \) is the wave number in vacuum \( (n = 1) \), and \( L(x) \) is called the optical path length of the wave. Plugging in eq. (5.36), the wave equation (5.34) becomes

\[
\phi \left\{ [\nabla^2 A + (\nabla A)^2 - k_0^2(\nabla L)^2 + k_0^2 n^2] + ik_0 \left( \nabla^2 L + 2\nabla A \cdot \nabla L \right) \right\} = 0.
\] (5.37)

In general, \( \phi \) is nonzero and so the curly-bracketed expression must vanish. We want \( A \) and \( L \) to be real-valued by construction, and so we the real and imaginary parts in square brackets must also vanish:

\[
\nabla^2 A + (\nabla A)^2 + k_0^2 [n^2 - (\nabla L)^2] = 0, \quad \nabla^2 L + 2\nabla A \cdot \nabla L = 0.
\] (5.38)

Now comes the geometric optics approximation: the wavelength \( \lambda = 2\pi/k \) is small compared to the rate of change of the medium. In particular, the wave number in vacuum \( k_0 = 2\pi/\lambda_0 \) must be considerably large compared to the derivative terms in the first equation of (5.38), and thus we require

\[
(\nabla L)^2 = n^2.
\] (5.39)

This is called the eikonal equation.

Returning to eq. (5.31), we see that the characteristic function \( W \) satisfies an eikonal equation for a “medium” of refractive index \( \sqrt{2m(E-U)} = p \) (this equality holds for the single-particle case). This illustrates that the characteristic function is like a wavefront that propagates through the medium of configuration space with refractive index \( p \), in the geometric optics limit.

### 5.5 Exercises

#### 5.1 (Harmonic oscillator).

The harmonic oscillator Hamiltonian is given by

\[
H = \frac{1}{2m} p^2 + \frac{m\omega^2}{2} q^2.
\]

(a) Write down the Hamilton–Jacobi equation (5.4) for this system. Since this system is conservative, we expect a solution (up to an arbitrary additive constant) of the form

\[
S(q, \alpha, t) = W(q, \alpha) - at
\]
where the constant $\alpha$ is the total energy. Plug this ansatz into the Hamilton–Jacobi equation and conclude that

$$W = \pm m\omega \int \sqrt{\frac{2\alpha}{m\omega^2} - q^2} \, dq.$$  

This integral can be evaluated further, but it is not necessary for our purposes.

(b) The quantity $\beta$ will implicitly give us the equation of motion $q(\alpha, \beta, t)$. Using the definition (5.6), show that

$$t + \beta = \frac{1}{\omega} \left\{ \sin^{-1} \left( \sqrt{\frac{2\alpha}{m\omega^2}} q \right) \right\} \text{(plus sign)} + \text{constant.}$$

The integration constant may be absorbed into $\beta$ which has yet to be determined, and since $\sin^{-1}(x) = \pi/2 - \cos^{-1}(x)$ we can take the second case and absorb the $\pi/2$ into $\beta$ also. We should pick up a factor of $-1$ on one side of the equation due to this last step, but this will not matter since we have chosen cosine which is even. Altogether, we arrive at the familiar solution

$$q(t) = \sqrt{\frac{2\alpha}{m\omega^2}} \cos \left( \omega(t + \beta) \right).$$

(c) To find the constants we must apply the initial conditions $q(0) = q_0$, $p(0) = p_0$. Determine $\alpha$ and $\beta$ using $p_0 = (\partial S/\partial q)|_{t=0}$ and $q(0) = q_0$ respectively, and obtain the solution as a function of the initial values:

$$q(t) = \sqrt{\frac{q_0^2 + p_0^2}{m^2\omega^2}} \cos \left[ \omega t + \cos^{-1} \left( \frac{q_0}{\sqrt{q_0^2 + p_0^2 / m^2\omega^2}} \right) \right].$$

5.2 (Central field). Consider the motion of a particle in a central field as in section 2.2. In polar coordinates, we have

$$T = \frac{m}{2} (r^2 + r^2 \dot{\phi}^2), \quad U = U(r),$$

$$p_r = \frac{\partial L}{\partial \dot{r}} = mr, \quad p_\phi = \frac{\partial L}{\partial \dot{\phi}} = mr^2 \dot{\phi},$$

$$H = \frac{1}{2m} \left( p_r^2 + \frac{p_\phi^2}{r^2} \right) + U(r).$$

(a) This Hamiltonian is both time-independent and cyclic in $\phi$, and so we expect an action of the form

$$S = W_1(r) + \alpha_\phi \phi - Et.$$
Plug this into the Hamilton–Jacobi equation (5.16) for conservative systems and integrate to arrive at

\[ W = W_r(r) + \alpha \phi = \pm \int \sqrt{2m(E - U(r)) - \frac{\alpha^2}{r^2}} \, dr + \alpha \phi. \]

(b) Use eq. (5.17) to obtain the implicit equations of motion

\[ \beta_1 = \pm \int \frac{\frac{1}{2} 2m \, dr}{\sqrt{2m(E - U(r)) - \frac{\alpha^2}{r^2}}} - t, \]

\[ \beta_2 = \pm \int \frac{-\frac{1}{2} \alpha \phi \, dr}{\sqrt{2m(E - U(r)) - \frac{\alpha^2}{r^2}}} + \phi. \]

These match what we found in equations (2.13) and (2.14), where the constant \( \alpha \phi = M \) is the angular momentum associated to the cyclic coordinate \( \phi \).

5.3 (Kepler’s problem). We will find the frequency of oscillations for Kepler’s problem using action variables without solving the equations of motion. Consider a particle of mass \( m \) in an inverse-square central force field, as in section 2.4

\[ T = \frac{m}{2} (\dot{r}^2 + r^2 \dot{\theta}^2 + r^2 \sin^2 \theta \dot{\phi}^2), \quad U = -\frac{k}{r}, \]

\[ p_r = \frac{\partial L}{\partial \dot{r}} = m \dot{r}, \quad p_\theta = \frac{\partial L}{\partial \dot{\theta}} = m r^2 \dot{\theta}, \quad p_\phi = \frac{\partial L}{\partial \dot{\phi}} = m r^2 \sin^2 \theta \dot{\phi}, \]

\[ H = \frac{1}{2m} \left( p_r^2 + \frac{p_\theta^2}{r^2} + \frac{p_\phi^2}{r^2 \sin^2 \theta} \right) - \frac{k}{r}. \]

The constant \( k \) is positive, since we must have an attractive force for bounded motion.

(a) Write down the Hamilton–Jacobi equation (5.16) for conservative systems. Notice that all of the coordinates are separable, and so the characteristic function is of the form

\[ W = W_r(r) + W_\theta(\theta) + W_\phi(\phi). \]

(b) The Hamiltonian is cyclic in \( \phi \), and so the derivative \( \partial W/\partial \phi = \partial W_\phi/\partial \phi = \alpha_\phi \) is the constant that is angular momentum about the \( z \)-axis. Plug this in, group the terms involving only \( \theta \), and conclude that

\[ \left( \frac{\partial W_\theta}{\partial \theta} \right)^2 + \frac{\alpha_\phi^2}{\sin^2 \theta} = \alpha_\phi^2, \quad \frac{1}{2m} \left[ \left( \frac{\partial W_r}{\partial r} \right)^2 + \frac{\alpha_\theta^2}{r^2} \right] = \frac{k}{r} = E. \]
The equations for $W_\phi$, $W_\theta$, and $W_r$ demonstrate the conservations of angular momentum about the $z$-axis $p_\theta$, total angular momentum $p$, and total energy $E$, and from here they could be integrated to obtain the equations of motion.

(c) Use the three differential equations of part (b) to obtain the action variables:

\[ I_\phi = \alpha_\phi = p_\phi, \]
\[ I_\theta = \frac{1}{2\pi} \oint \sqrt{\alpha_\phi^2 - \frac{\alpha_\phi^2}{\sin^2 \theta}} \, d\theta, \]
\[ I_r = \frac{1}{2\pi} \oint \sqrt{2mE + \frac{2mk}{r} - \frac{\alpha_\theta^2}{r^2}} \, dr. \]

(d) Let us look at the second action variable $I_\theta$. We know from section 2.2 that this motion is coplanar, so let $\psi$ denote the angle in the plane of orbit. Set the momentum in the $(r, \theta, \phi)$ variables and $(r, \psi)$ variables equal, and conclude that $p_\theta \, d\theta = p \, d\psi - p_\phi \, d\phi$. Conclude that

\[ I_\theta = p - p_\phi = \alpha_\theta - \alpha_\phi. \]

(e) Now for the third action variable $I_r$. The integral for $I_r$ is evaluated between two turning points $r_1, r_2$ for which the integrand $p_r = m \dot{r}$ must vanish. We can therefore integrate from $r_2$ to $r_1$ and back to $r_2$, for which the integrand is first negative then positive, corresponding to the sign of the momentum $p_r = m \dot{r}$. In the complex plane, this integrand is analytic everywhere but $r = 0$ and along the segment on the real axis connecting $r_1$ and $r_2$. Integrate around a counterclockwise simple path $\gamma$ encompassing $r_1$ and $r_2$ to obtain

\[ I_r = -\alpha_\theta^2 + \frac{mk}{\sqrt{-2mE}} = -(I_\theta + I_\phi) + \frac{mk}{\sqrt{-2mE}}. \]

The energy

\[ E(I) = -\frac{mk^2}{2(I_r + I_\theta + I_\phi)^2} \]

is symmetric in the three action variables, and so the frequency of oscillations in each coordinate $r, \theta, \phi$ is the same:

\[ \frac{\partial E}{\partial I_r} = \frac{\partial E}{\partial I_\theta} = \frac{\partial E}{\partial I_\phi} = \frac{mk^2}{(I_r + I_\theta + I_\phi)^3}, \]

as is expected since the force is central.
Part III

Hamiltonian Mechanics

Rather than working on the tangent bundle over configuration space, Hamiltonian mechanics separates the position and momentum variables and is naturally phrased on the cotangent bundle. The equations of motion are a system of first-order differential equations with two equations for each degree of freedom, and is often the preferred formulation in mathematics. The position and momentum are of course not truly independent coordinates, and so an acceptable change of variables must satisfy the criterion of a canonical transformation. This relationship between the position and momenta induces a geometric structure on the tangent bundle, which is abstracted into symplectic geometry in the case of conservative mechanics and contact geometry for dissipative mechanics.
Chapter 6

Hamilton’s equations of motion

We will develop the equations of motion and key tools of the Hamiltonian perspective, with a primary focus on Euclidean space before exploring general geometries. The material for this chapter is based on [Arn89, Ch. 3], [Gol51, Ch. 7–8], and [LL76, Ch. 7].

6.1 Hamilton’s equations

Recall that for a system with coordinates \(q = (q_1, q_2, \ldots, q_n)\) on an \(n\)-dimensional manifold \(M\) the Hamiltonian

\[
H(q, p, t) := \dot{q} \cdot p - \mathcal{L}(q, \dot{q}, t) = \sum_{i=1}^{n} \dot{q}_i \frac{\partial \mathcal{L}}{\partial \dot{q}_i} - \mathcal{L}
\]

is often the total energy of the system, and the components of \(p\) are the momenta \(p_i := \partial \mathcal{L} / \partial \dot{q}_i\). In general, the Hamiltonian is a smooth function function \(H(q, p, t) : T^*M \times I \rightarrow \mathbb{R}\) where \(I \subset \mathbb{R}\) is an interval, since the momentum is a covector and not a vector. In Proposition 3.11 we saw that the Hamiltonian determines the system’s motion, and that the quantity \(H\) is conserved whenever the Lagrangian \(\mathcal{L}\) is time-independent.

When we defined the Hamiltonian \(H\) in section 3.5, we showed that it is equal to the total energy \(T + U\) of the system in Cartesian coordinates. We will now show that the Hamiltonian \(H(q, p, t)\) is equal to the total energy \(T + U\) for any system with velocity-independent potential energy \(U \equiv U(q, t)\). As in eq. (3.11) the kinetic energy is a quadratic form in the velocity variables:

\[
T = \frac{1}{2} \sum_{i,j} a_{ij}(q, t) \dot{q}_i \dot{q}_j
\]
So $T$ is a homogeneous function of order 2 in $\dot{q}$: for any constant $k$ we have
\[ T(q, k\dot{q}, t) = \frac{1}{2} \sum_{i,j} a_{ij}(q, t) \frac{d}{dt} (kq_i) \frac{d}{dt} (kq_j) = k^2 T(q, \dot{q}, t). \] (6.3)

This determines the derivative of $T$, since
\[ 2kT = \frac{d}{dk}(k^2 T) = \frac{d}{dk} T(q, k\dot{q}, t) = \left( \frac{\partial}{\partial(kq)} T(q, k\dot{q}, t) \right) \cdot \left( \frac{\partial k\dot{q}}{\partial k} \right) = \frac{\partial T(k\dot{q})}{\partial(k\dot{q})} \cdot \dot{q}, \]
and so taking $k = 1$, we conclude
\[ \frac{\partial T}{\partial \dot{q}} \cdot \dot{q} = 2T. \]

(This result is known as Euler’s homogeneous function theorem.) Since the potential energy is velocity independent, then we have
\[ H = p \cdot \dot{q} - \mathcal{L} = \frac{\partial \mathcal{L}}{\partial \dot{q}} \cdot \dot{q} - \mathcal{L} = \frac{\partial T}{\partial \dot{q}} \cdot \dot{q} - (T - U) = 2T - T + U = T + U \]
as desired.

Treating the position $q$, velocity $\dot{q}$, and momentum $p$ as independent variables, the differential of the Hamiltonian must satisfy
\[ \frac{\partial H}{\partial p} \, dp + \frac{\partial H}{\partial q} \, dq + \frac{\partial H}{\partial t} \, dt = dH = d(p \cdot \dot{q} - \mathcal{L}) \]
\[ = \dot{q} \, dp + \left( p - \frac{\partial \mathcal{L}}{\partial \dot{q}} \right) \, dq - \frac{\partial \mathcal{L}}{\partial q} \, dq - \frac{\partial \mathcal{L}}{\partial t} \, dt \]
\[ = \dot{q} \, dp + 0 \, dq - \dot{p} \, dq - \frac{\partial \mathcal{L}}{\partial t} \, dt \]
where in the last equality we used Lagrange’s equations. Of course the momentum and velocity are not actually independent, but the coefficient of $dq$ will appear with a factor in the term $dp$ and the same cancellation occurs. Matching terms we conclude
\[ \frac{\partial H}{\partial p} = \dot{q}, \quad \frac{\partial H}{\partial q} = -\dot{p}. \] (6.4)

These $2n$ first-order differential equations are called Hamilton’s equations. Note that equating the terms containing $dt$ provides no information since this equality follows immediately from the definition of the Hamiltonian.

In Lagrangian mechanics we saw that a cyclic coordinate $q_k$ resulted in the conservation of $p_k = \frac{\partial \mathcal{L}}{\partial \dot{q}_k}$, and thus $p_k$ entered the remaining $n - 1$ equations as a constant parameter. For Hamilton’s equations, a cyclic variable $q_k$ yields
\[ \dot{p}_k = \frac{\partial H}{\partial q_k} = 0, \quad \dot{q}_k = \frac{\partial H}{\partial p_k}. \]
The first of these equations expresses the conservation of \( p_k \), and the second is independent of \( q_k \) and is thus easily integrable. Consequently, we are left with \( 2n - 2 \) equations in terms of the constant parameter \( p_k \).

To conclude this section, we will observe Hamilton’s equations as a consequence of the principle of least action. Although we have defined momentum and thus Hamilton’s equations themselves in terms of the Lagrangian and the generalized coordinates \( q_i \), we can also consider the momenta \( p_i \) as variables since we can still extract the Euler–Lagrange equations as long as we have at least \( n \) independent coordinates (cf. the remark at the end of section 3.1). For the Lagrangian of a system with \( n \) degrees of freedom, we must specify \( 2n \) initial conditions—\( n \) for the coordinates \( q_i \) and another \( n \) for the \( \dot{q}_i \)—and so deriving Hamilton’s equations without the Lagrangian equations of motion allow us to regard the \( q_i \) and \( p_i \) both as independent coordinates linked only by the equations of motion. To demonstrate this, suppose we have a system whose motion is described by the coordinates \((q(t), p(t))\) from time \( t_0 \) to \( t_1 \). Within a coordinate patch on Euclidean space, we may write any other time progression of the system as \( q(t) + \delta q(t) \) with \( \delta q(t_0) = \delta q(t_1) = 0 \). The principle of least action insists

\[
0 = \delta S = \delta \int_{t_0}^{t_1} \mathcal{L} \, dt = \delta \int_{t_0}^{t_1} (p \cdot \dot{q} - H(q, p, t)) \, dt
\]

We may integrate the second term by parts:

\[
\int_{t_0}^{t_1} p_i \delta \dot{q}_i \, dt = \int_{t_0}^{t_1} \frac{d}{dt} (p_i \delta q_i) \, dt = p_i \delta q_i \bigg|_{t_0}^{t_1} - \int_{t_0}^{t_1} \dot{p}_i \delta q_i \, dt
\]  

(6.5)

The boundary term vanishes by the fixed-endpoint condition \( \delta q_i(t_0) = \delta q_i(t_1) = 0 \). Plugging this back in to the action variation yields

\[
0 = \int_{t_0}^{t_1} \left[ \left( \dot{q} - \frac{\partial H}{\partial p} \right) \cdot \delta p - \left( \dot{p} + \frac{\partial H}{\partial q} \right) \cdot \delta q \right] \, dt
\]

Since this integral vanishes for any choice of \( \delta p_i(t) \) and \( \delta q_i(t) \), and we are always able to pick paths such that these parameters are varied independently, then we conclude that the two parenthetical terms must vanish:

\[
\dot{q}_i - \frac{\partial H}{\partial p_i} = 0, \quad \dot{p}_i + \frac{\partial H}{\partial q_i} = 0.
\]

This agrees with Hamilton’s equations (6.4).

6.2 Legendre transformation

The Legendre transformation is an involution on the space of real-valued strictly convex functions, which naturally joins the Lagrangian and Hamiltonian per-
spectives. As is the case in physical application, the Lagrangian and Hamiltonian are quadratic in \( \dot{q} \) due to the kinetic energy, and are Legendre transforms of each other in the variable \( \dot{q} \).

Throughout this section, \( f: U \to \mathbb{R} \) will denote a smooth function on an open convex set \( U \subseteq \mathbb{R}^d \) that is (strictly) convex in the sense that the Hessian matrix is positive definite, which is written as \( f''(x) > 0 \), and is defined by

\[
f''(x)y \cdot y = \left( \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right) y \cdot y > 0 \quad \text{for all } 0 \neq y \in \mathbb{R}^d
\]

at every point \( x \in U \). We will now define the Legendre transform \( f^*(\xi) \) of such a function \( f(x) \). Consider the distance \( F(x, \xi) = x \cdot \xi - f(x) \) between the hyperplane \( x \cdot \xi \) of “slope” \( \xi \) and the graph of the function \( f(x) \). Since \( f \) is convex, then for fixed \( \xi \) the function \( F \) will be a concave function of \( x \), and so there is a unique point \( x^*(\xi) \) which maximizes \( F(x, \xi) \). We define the Legendre transform of \( f \) by \( f^*(\xi) = F(x^*(\xi), \xi) \), which mathematically is the function

\[
f^*(\xi) = \sup_{x \in U} (x \cdot \xi - f(x)) \quad (6.6)
\]

on the domain \( U^* \) where this supremum is finite:

\[
U^* = \sup \left\{ \xi \in \mathbb{R}^d : \sup_{x \in U} (x \cdot \xi - f(x)) < \infty \right\}. \quad (6.7)
\]

**Example 6.1.** Consider the quadratic function

\[
f(x) = \frac{1}{2}Ax \cdot x + b \cdot x + c
\]

on \( U = \mathbb{R}^d \), where \( A \) is a real symmetric positive definite matrix, \( b \in \mathbb{R}^d \), and \( c \in \mathbb{R} \). In order to maximize the distance \( F(x, \xi) \), we differentiate:

\[
0 = \frac{\partial F}{\partial x}(x^*) = \frac{\partial}{\partial x} [x \cdot \xi - \frac{1}{2}Ax \cdot x - b \cdot x - c]_{x=x^*} = \xi - Ax^* - b.
\]

This equation has one critical point \( x^* = A^{-1}(\xi - b) \), which must be our maximum. Plugging this back in, we get

\[
f^*(\xi) = F(x^*(\xi), \xi) = A^{-1}(\xi - b) \cdot \xi - \frac{1}{2}(\xi - b) \cdot A^{-1}(\xi - b) - b \cdot A^{-1}(\xi - b) - c = \frac{1}{2}A^{-1}(\xi - b) \cdot (\xi - b) - c
\]

on \( U^* = \mathbb{R}^d \). In particular, we can take \( A = mI \), \( b = 0 \), and \( c = 0 \) so that \( f(x) = \frac{1}{2}m|\dot{x}|^2 \) is the kinetic energy, and we get that \( f^*(\xi) = |\xi|^2/2m \) is also the kinetic energy if we recognize \( \xi \) as the momentum.

Next, we will demonstrate the transformation’s key properties.

**Theorem 6.2 (Involutive property).** The Legendre transform \( f^*: U^* \to \mathbb{R} \) of \( f: U \to \mathbb{R} \) is a convex function, whose Legendre transform \( (f^*)^* \) is \( f \). In other words, the Legendre transform is an involution.
6.2. LEGENDRE TRANSFORMATION

Proof. First we check that the transform \( f^* \) is also convex. As the unique maximizer of the distance \( F(x, \xi) \), we know that \( x^* = x^*(\xi) \) is the unique solution to

\[
0 = \frac{\partial F}{\partial x}(x^*) = \xi - Df(x^*)
\]

(6.8)

where \( Df(x^*) = \nabla f(x^*) \) is the gradient. The function \( Df \) between open sets of \( \mathbb{R}^d \) is invertible and so we can write \( x^* = (Df)^{-1}(\xi) \). By the inverse function theorem, the derivative of this function \( x^*(\xi) \) is

\[
D\xi x^*(\xi) = D\xi (Df)^{-1}(\xi) = (D^2f)^{-1}(x^*(\xi))
\]

(6.9)

where \( D^2f(x) = f''(x) \) is the Hessian matrix whose entries are functions of \( x \). The first derivative of \( f^*(\xi) \) is

\[
Df^*(\xi) = D\xi \{ \xi \cdot (Df)^{-1}(\xi) - f[(Df)^{-1}(\xi)] \}
= (Df)^{-1}(\xi) + \xi \cdot (D^2f)^{-1}(x^*(\xi)) - Df[(Df)^{-1}(\xi)] \cdot (D^2f)^{-1}(x^*(\xi))
= (Df)^{-1}(\xi).
\]

Therefore, the second derivative of our Legendre transform is

\[
(f^*)''(\xi) = D^2f^*(\xi) = (D^2f)^{-1}(x^*(\xi)) > 0,
\]

(6.10)

which is positive definite since \( f''(x) \) is. That is, the Legendre transform \( f^* \) is also convex.

Now that we know

\[
g(\xi) := f^*(\xi) = x^*(\xi) \cdot \xi - f(x^*(\xi)) = Df^{-1}(\xi) \cdot \xi - f(Df^{-1}(\xi))
\]

(6.11)

is also convex, we may now consider its Legendre transform. Let \( \xi^* \) be the point which maximizes the distance function \( G(\xi, x) \) for \( g \). This point is determined by the condition

\[
0 = \frac{\partial G}{\partial \xi}(\xi^*) = \frac{\partial}{\partial \xi} [\xi \cdot x - g(\xi)]_{\xi = \xi^*} = x - Dg(\xi^*),
\]

(6.12)

which we can solve as

\[
\xi^* = (Dg)^{-1}(x) = Df(x)
\]

(6.13)

(using our previous calculation of \( Dg = Df^* \)). Consequently, the transform of \( g(\xi) \) is given by

\[
g^*(x) = G(\xi^*, x) = \xi^* \cdot x - g(\xi^*)
= Df(x) \cdot x - Df^{-1}(Df(x)) \cdot Df(x) + f(Df^{-1}(Df(x))) = f(x)
\]

as desired. For fixed \( x_0 \), \( G(\xi, x_0) \) geometrically represents the “\( y \)-coordinate” of the intersection of the line tangent to \( f(x) \) with slope \( \xi \). Since \( f \) is convex, all of the tangent lines of \( f \) lie below the curve, and so the maximum of \( G(\xi, x) \) will be at \( f(x_0) \). Therefore, the transformation of an already transformed function \( g(\xi) \) of some smooth \( f(x) \) must reproduce the original function \( f(x) \).
Returning to the context of mechanics, we make the following simple calculation which is a consequence of the kinetic energy being quadratic.

**Proposition 6.3.** Let \( M \) be a real symmetric positive definite matrix. The Legendre transformation of the conservative Lagrangian

\[
L(q, \dot{q}) = \frac{1}{2} M \dot{q} \cdot \dot{q} - U(q)
\]

in the velocity \( \dot{q} \) variables with dual variable \( p \) yields the corresponding Hamiltonian

\[
H(q, p) = \frac{1}{2} M^{-1} p \cdot p + U(q).
\]

Since the velocity \( \dot{q} \) is a tangent vector, then this duality and the corresponding natural paring \( x \cdot \xi = \langle \dot{q}, p \rangle \) shows us that momentum \( p \) is naturally a covector.

**Proof.** Take \( A = M, \ b = 0, \ \text{and} \ c = U(q) \)—which is constant with respect to the velocity \( \dot{q} \) in Example 6.1. \( \square \)

### 6.3 Liouville’s theorem

Liouville’s theorem states that the density of trajectories in phase space surrounding a system configuration as it evolves in time is constant.

Suppose the Hamiltonian \( H(q, p) \) for a system with \( n \) degrees of freedom does not depend explicitly on time. We are interested in time progressions of the system in phase space, the \( 2n \)-dimensional space of the coordinates \((q, p) = (q_1, \ldots, q_n, p_1, \ldots, p_n) \in \mathbb{R}^{2n} \). Consider the vector field given by Hamilton’s equations: at each point \((q, p)\) in phase space we know that the trajectory of the system at this configuration is tangent to the vector \((-\partial H/\partial q, \partial H/\partial p)\) which is tangent to the trajectory passing through \((q, p)\). We will (as always) suppose that the solution \((q(t), p(t))\) for any initial condition \((q(0), p(0)) = (q_0, p_0)\) of Hamilton’s equations can be extended for all time. The **phase flow** is then the set of transformations \( \phi^t : (q_0, p_0) \mapsto (q(t), p(t)) \) with \( t \in \mathbb{R} \) of phase space to itself. The phase flow forms a group under function composition \( \circ \) parameterized by the additive group \( t \in \mathbb{R} \) and a group action on phase space \( \mathbb{R}^{2d} = \mathbb{R}_q \times \mathbb{R}_p \):

1. **At** \( t = 0 \) the phase flow is the identity map: \( \phi^0(q_0, p_0) = (q_0, p_0) \) for all \((q_0, p_0) \in \mathbb{R}^{2n} \).

2. **Composition** corresponds to addition in \( t \in \mathbb{R} \): \( \phi^0 \circ \phi^t(q_0, p_0) = \phi^{t_1}(q_0, p_0) = \phi^{0 + t_1}(q_0, p_0) \).

3. **Inversion** corresponds to negation in \( t \in \mathbb{R} \): \( \phi^0 \circ \phi^{-t}(q_0, p_0) = \phi^{0 + (-t)}(q_0, p_0) = \phi^0(q_0, p_0) \) and vice versa.

4. **Function composition** is associative and the group action is as well, because addition is associative in \( \mathbb{R} \).
6.3. LIOUVILLE’S THEOREM

Theorem 6.4 (Liouville’s theorem). The density of system points in phase space \( \mathbb{R}^{2n} = \mathbb{R}^n_p \times \mathbb{R}^n_q \) around a phase curve remains constant along the curve. That is, for any measurable set \( D \subset \mathbb{R}^{2n} \) in phase space, the volume of \( \phi^t(D) \) is equal to that of \( D \).

Proof. First, let us simplify notation using the point \( x := (q, p) \in \mathbb{R}^{2n} \) and the Hamiltonian vector field \( f : \mathbb{R}^{2n} \to \mathbb{R}^{2n}, f(x) := (-\partial H/\partial q, \partial H/\partial p) \), so that

\[
\dot{x} = f(x), \quad \phi^t(x) = x + f(x)t + \mathcal{O}(t^2).
\]

(6.14)

Consider a region \( D(0) \subset \mathbb{R}^{2n} \), and let \( D(t) = \phi^t(D(0)) \) denote its flow and \( v(t) \) denote volume of \( D(t) \). We will prove the more general statement

if \( \nabla \cdot f \equiv 0 \) then \( v(t) = v(0) \),

which is sufficient since for our vector field we have

\[
\nabla \cdot f = -\frac{\partial^2 H}{\partial p \partial q} + \frac{\partial^2 H}{\partial q \partial p} = 0
\]

by the symmetry of second derivatives.

The volume \( v(t) \) may be expressed as the integral

\[
v(t) = \int_{\phi^t(D(t_0))} dx = \int_{D(t_0)} \det \left[ \frac{\partial \phi^t_{t_0}}{\partial x_j} \right] dx \tag{6.15}
\]

by changing variables. Using the Taylor expansion (6.14), we can write

\[
\left[ \frac{\partial \phi^t_{t_0}}{\partial x_j} \right] = I + (t - t_0) \left[ \frac{\partial f_i}{\partial x_j} \right] + \mathcal{O}((t - t_0)^2).
\]

It is true for any matrix \( A \) and scalar \( k \) that

\[
\det(I + hA) = 1 + h \operatorname{tr}(A) + \mathcal{O}(h^2),
\]

which is just a rewriting of Jacobi’s formula from matrix calculus. With this, the volume formula (6.15) becomes

\[
v(t) = \int_{D(t_0)} \left[ 1 + (t - t_0) \operatorname{tr} \left[ \frac{\partial f_i}{\partial x_j} \right] + \mathcal{O}((t - t_0)^2) \right] dx
\]

\[
= \int_{D(t_0)} \left[ 1 + (t - t_0) (\nabla \cdot f) + \mathcal{O}((t - t_0)^2) \right] dx.
\]

Differentiating this, we conclude

\[
\frac{dv}{dt} \bigg|_{t=t_0} = \int_{D(t_0)} \nabla \cdot f \, dx \tag{6.16}
\]

The integrand is zero by premise, and so \( v(t) \) is constant in time: \( v(t) \equiv v(0) \). \( \square \)
Liouville’s theorem has many important consequences. For example, we now know that for a Hamiltonian system there can be no asymptotically stable equilibrium points or asymptotically stable closed trajectories in phase space, since these would require that the density of phase curves to increase around such phenomena. We also have the following phenomenon.

**Corollary 6.5. (Poincare’s recurrence theorem)** Fix $t \in \mathbb{R}$ and $D \subset \mathbb{R}^{2n}$ a bounded region of phase space, and let $\phi := \phi^t$ denote a phase flow group element. Then for any positive measure set $\mathcal{U} \subset D$ there exists $x_0 \in \mathcal{U}$ and a positive integer $n$ such that $\phi^n(x_0) \in \mathcal{U}$.

For bounded motion—as is the case for a conservative system with potential energy $U(x) \to +\infty$ as $|x| \to \infty$, in which case a particle is confined to the region $D = \{(q, p) \in \mathbb{R}^{2d} : T + U \leq E\}$—this means that the system will return to an arbitrary vicinity of any given possible configuration $(q, p)$ infinitely often, given enough time. If we were to open a connection between a chamber of gas and a chamber of vacuum then the gas molecules will eventually all return to the initial chamber, seemingly in violation of the second law of thermodynamics. Although it may appear that Poincare’s theorem contradicts Liouville’s, the time scales are often quite large—for a gas chamber it is longer than the age of the universe—and so there is no conflict.

**Proof.** Since for a smooth Hamiltonian function Hamilton’s equations are a smooth system of ODEs, then the subsequent flow $\phi : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is injective by uniqueness of solutions and continuous by well-posedness. Liouville’s theorem tells us that the special form of this system that $g$ is volume-preserving.

Consider the collection of sets $\mathcal{U}, \phi(\mathcal{U}), \phi^2(\mathcal{U}), \phi^3(\mathcal{U}), \ldots \subset D$. Since $\phi$ is volume-preserving, all of these sets must have the same volume. $D$ is bounded and thus has finite volume, and so it is impossible for all of these sets to be disjoint. That is, there exists some distinct $0 < j < k$ such that $\phi^j(\mathcal{U}) \cap \phi^k(\mathcal{U}) \neq \emptyset$. Since $\phi$ is injective, this requires $\phi^{k-j}(\mathcal{U}) \cap \mathcal{U} \neq \emptyset$. Namely, we can pick some $x_0 \in \mathcal{U}$ in this intersection, which gives $\phi^{k-j}(x_0) \in \mathcal{U}$.

In fact, we can conclude that the set of points in $\mathcal{U}$ which do not return to $\mathcal{U}$ infinitely often has measure zero. \qed

Together, a measure space $X$ with a finite measure $\mu$ and a measurable function $\phi : X \to X$ that is measure preserving (in the sense that $\mu(\phi^{-1}(A)) = \mu(A)$ for all measurable $A \subset X$) constitute a measure preserving space, which is the fundamental object of study in discrete dynamical systems.

### 6.4 Poisson bracket

For any two functions $f$ and $g$ of position $q$, momenta $p$, and time $t$ we define the Poisson bracket of $f$ and $g$ to be

$$\{f, g\} = \sum_{k=1}^{n} \left( \frac{\partial f}{\partial p_k} \cdot \frac{\partial g}{\partial q_k} - \frac{\partial f}{\partial q_k} \cdot \frac{\partial g}{\partial p_k} \right).$$  \hspace{1cm} (6.17)
6.4. POISSON BRACKET

(Note that there is another popular convention which differs by a factor of \(-1\).) It arises naturally as the time evolution of any function \(f(p,q,t)\); by the chain rule and Hamilton’s equations we have

\[
\frac{df}{dt} = \frac{\partial f}{\partial t} + \sum_k \left( \frac{\partial f}{\partial q_k} \dot{q}_k + \frac{\partial f}{\partial p_k} \dot{p}_k \right)
\]

\[
= \frac{\partial f}{\partial t} + \sum_k \left( \frac{\partial f}{\partial q_k} \frac{\partial H}{\partial p_k} - \frac{\partial f}{\partial p_k} \frac{\partial H}{\partial q_k} \right) = \frac{\partial f}{\partial t} + \{H,f\}. \tag{6.18}
\]

Consequently, a time-independent quantity \(f(q,p)\) is conserved exactly when its bracket with the Hamiltonian \(\{H,f\}\) vanishes. Taking the quantity \(f\) to be a position or momentum coordinate, we see that Hamilton’s equations may be written as

\[
\dot{q}_i = \frac{\partial H}{\partial p_i} = \{H,q_i\}, \quad \dot{p}_i = -\frac{\partial H}{\partial q_i} = \{H,p_i\}. \tag{6.19}
\]

This perspective generalizes nicely to the Heisenberg formulation of quantum mechanics, with the Poisson brackets replaced by operator commutators; an observable without explicit time dependence form a Lax pair with the Hamiltonian.

Next, we record some properties. It is easy to see that the Poisson bracket is bilinear (linear in both entries) and antisymmetric (\(\{f,g\} = -\{g,f\}\)) from the definition (6.17). The brackets also posses a Leibniz rule,

\[
\{fg,h\} = f\{g,h\} + g\{f,h\}, \tag{6.20}
\]

which can be seen by using the product rule for \(fg\) and expanding. They also behave nicely with the coordinates and momenta:

\[
\{q_i,q_k\} = 0, \quad \{p_i,p_k\} = 0, \quad \{p_i,q_k\} = \delta_{ik}. \tag{6.21}
\]

Also, if \(f(p,q,t)\) is a function as before and \(g(p,q,t) \equiv G(f(q,p,t))\), then their bracket

\[
\{f,G(f)\} = \sum_{k=1}^d \left( \frac{\partial f}{\partial p_k} G'(f) \frac{\partial f}{\partial q_k} - \frac{\partial f}{\partial q_k} G'(f) \frac{\partial f}{\partial p_k} \right) = 0 \tag{6.22}
\]

vanishes by the chain rule.

The Poisson bracket also satisfies the Jacobi identity:

\[
\{f,\{g,h\}\} + \{g,\{h,f\}\} + \{h,\{f,g\}\} = 0, \tag{6.23}
\]

which we will now verify. Since \(\{f,g\}\) is a linear function where each term contains a first derivative of \(f\) and \(g\), then \(\{h,\{f,g\}\}\) (for example) is a linear function in which each term contains a second derivative. Let \(D_g(\phi) = \{g,\phi\}\) and \(D_h(\phi) = \{h,\phi\}\). Note that the first term of eq. (6.23) does not contain any second derivatives of \(f\), and so all of the second derivatives of \(f\) are contained within

\[
\{g,\{h,f\}\} + \{h,\{f,g\}\} = \{g,\{h,f\}\} - \{h,\{g,f\}\}
\]

\[
= D_g(D_h(f)) - D_h(D_g(f)) = (D_gD_h - D_hD_g)f
\]
To simplify notation and observe some cancellation, we write the operators \( D_g \) and \( D_h \) as

\[
D_g = \sum_j \left( \frac{\partial g}{\partial p_j} \frac{\partial}{\partial q_j} - \frac{\partial g}{\partial q_j} \frac{\partial}{\partial p_j} \right) = \sum_k \xi_k \frac{\partial}{\partial x_k}
\]

\[
D_h = \sum_j \left( \frac{\partial h}{\partial p_j} \frac{\partial}{\partial q_j} - \frac{\partial h}{\partial q_j} \frac{\partial}{\partial p_j} \right) = \sum_k \eta_k \frac{\partial}{\partial x_k}
\]

where \( x = (q,p) \), \( \xi = \left( \frac{\partial g}{\partial p}, -\frac{\partial g}{\partial q} \right) \), and \( \eta = \left( \frac{\partial h}{\partial p}, -\frac{\partial h}{\partial q} \right) \). Therefore

\[
D_g D_h = \sum_k \xi_k \frac{\partial}{\partial x_k} \left( \sum_l \frac{\eta_l}{\partial x_l} \right) = \sum_{k,l} \left( \xi_k \frac{\partial \eta_l}{\partial x_k} + \xi_l \frac{\partial^2}{\partial x_k \partial x_l} \right)
\]

\[
D_h D_g = \sum_k \eta_k \frac{\partial}{\partial x_k} \left( \sum_l \frac{\xi_l}{\partial x_l} \right) = \sum_{k,l} \left( \eta_k \frac{\partial \xi_l}{\partial x_k} + \eta_l \frac{\partial^2}{\partial x_k \partial x_l} \right)
\]

In taking the difference of these, we see that the second terms in these last equalities cancel and we are left with

\[
D_g D_h - D_h D_g = \sum_{k,l} \left( \xi_k \frac{\partial \eta_l}{\partial x_k} - \eta_k \frac{\partial \xi_l}{\partial x_k} \right) \frac{\partial}{\partial x_l}
\]

That is, all of the second derivatives of \( f \) cancel, leaving only first derivatives.

Symmetrically, this must also be true for \( g \) and \( h \), and since every term of eq. (6.23) contains only terms containing exactly one second derivative, then all the terms must cancel.

We will record one final property of the Poisson bracket.

**Proposition 6.6.** (Poisson’s theorem) If \( f(q,p) \) and \( g(q,p) \) are constants of motion for a Hamiltonian system, then their Poisson bracket \( \{ f, g \} \) is also conserved.

Although powerful, it should be noted that the new quantity \( \{ f, g \} \) may be trivial, especially considering that a system with \( n \) degrees of freedom can only have up to \( 2n - 1 \) conserved quantities.

**Proof.** Since \( f \) is conserved and time-independent we have

\[
0 = \frac{df}{dt} = \{ H, f \},
\]

and similarly for \( g \). The Jacobi identity (6.23) yields

\[
0 = \{ H, \{ f, g \} \} + \{ f, \{ g, H \} \} + \{ g, \{ H, f \} \} = \{ H, \{ f, g \} \} - \{ f, \{ H, g \} \} + \{ g, \{ H, f \} \} = \{ H, \{ f, g \} \}.
\]
Rearranging, we get
\[ \frac{d}{dt} \{ f, g \} = \{ H, \{ f, g \} \} = 0, \]
and so \( \{ f, g \} \) is an integral of the motion. \( \square \)

**Example 6.7.** If \( L_1 = x_2 p_3 - x_3 p_2 \) and \( L_2 = x_3 p_1 - x_1 p_3 \) are the first and second components of the angular momentum \( \mathbf{L} = \mathbf{x} \times \mathbf{p} \) of a particle \( \mathbf{x} \in \mathbb{R}^3 \), then their bracket
\[ \{ L_1, L_2 \} = \{ r_2 p_3, r_3 p_1 \} + \{ r_3 p_2, r_1 p_3 \} = r_2 p_1 - p_2 r_1 = -L_3 \]
is the remaining component of the angular momentum.

### 6.5 Canonical transformations

For a system with \( n \) degrees of freedom, any choice of coordinates \( q = (q_1, q_2, \ldots, q_n) \) is of course far from unique. In fact, we may choose any set of \( n \) independent quantities \( Q_i(q,t), i = 1, \ldots, n \) that describe the system’s state and Lagrange’s equations must still hold in terms of the new coordinates. For the Hamiltonian approach, we would like to consider the more general transformation of \( (q, p) \) to the \( 2n \) independent quantities \( Q_i(q,p,t), P_i(q,p,t), \) and \( T(q,p,t) \) for \( i = 1, \ldots, n \) so that on the extended phase space \( T^* M \times I, I \subset \mathbb{R} \) the one-forms
\[ p \, dq - H \, dt = P \, dQ - K \, dT + dS \]  \hspace{1cm} (6.24)
are equal, for some smooth functions \( K(Q,P,T) \) and \( S(Q,P,T) \). A transformation satisfying eq. (6.24) is called **canonical**.

The definition (6.24) is made so that Hamilton’s equations
\[ \frac{dQ_i}{dT} = \frac{\partial K}{\partial P_i}, \quad \frac{dP_i}{dt} = -\frac{\partial K}{\partial Q_i} \]
hold for the new variables and the new Hamiltonian \( K(Q,P,t) \) (sometimes referred to as the *Kamiltonian*); in particular, if the canonical transformation is time-independent then we have \( K(Q,P) = H(q,p) \). As remarked at the end of section 3.1, this is a consequence of the Lagrangians differing by a total time derivative:
\[ p \cdot \dot{q} - H(q,p,t) = P \cdot \dot{Q} - K(Q,P,t) + \frac{d}{dt} [F(q(t),p(t),Q(t),P(t),t)] . \]

In physics the function \( F \) is called the **generating function** for the transformation.

**Remark.** Some authors define canonical transformations as any diffeomorphism which obeys Hamilton’s equations (6.5), and even falsely claim the equivalence of the two definitions (e.g. [LL76, Sec. 45]). The condition that a transformation preserves Hamilton’s equations (6.5) is much weaker—for example \( P = 2p, Q = q \) satisfies section 6.5 but not eq. (6.24)—but our definition has deeper geometric significance.
The first type of generating function we will consider is \( F \equiv F(q,Q,t) \). Applying the chain rule and rearranging eq. (6.24) yields

\[
p \cdot \dot{q} - P \cdot \dot{Q} + K(Q,P,t) - H(q,p,t) = \frac{\partial F}{\partial q} \cdot \dot{q} + \frac{\partial F}{\partial Q} \cdot \dot{Q} + \frac{\partial F}{\partial t}.
\]

Matching terms we find

\[
p_i = \frac{\partial F}{\partial q_i}, \quad P_i = -\frac{\partial F}{\partial Q_i}, \quad K = H + \frac{\partial F}{\partial t}. \tag{6.25}
\]

In this way, the generating function does indeed characterize the transformation.

Now consider a generating function of the form \( \Phi \equiv \Phi(q,P,t) \). Note that

\[
\frac{d}{dt} \left( \sum_{i=1}^{d} P_i Q_i \right) = \sum_{i=1}^{d} \dot{P}_i Q_i + \sum_{i=1}^{d} P_i \dot{Q}_i.
\]

We add this to eq. (6.24) to solve for the remaining variables \( p \) and \( Q \):

\[
p \cdot \dot{q} - P \cdot \dot{Q} + \dot{P} \cdot Q + \dot{P} \cdot \dot{Q} + K(Q,P,t) - H(q,p,t)
\]

\[
= \frac{d}{dt} \left[ P \cdot Q + F(q,P,t) \right] = \frac{\partial \Phi}{\partial q} \cdot \dot{q} + \frac{\partial \Phi}{\partial P} \cdot \dot{P} + \frac{\partial \Phi}{\partial t}
\]

which requires that our generating function be of the form \( \Phi(q,P,t) = P \cdot Q + F(q,P,t) \) and

\[
p_i = \frac{\partial \Phi}{\partial q_i}, \quad Q_i = \frac{\partial \Phi}{\partial P_i}, \quad K = H + \frac{\partial \Phi}{\partial t}. \tag{6.26}
\]

Similar procedures yield the relations corresponding to generating functions of \( p,Q \) or of \( p,P \). Notice that in matching terms we need the \( 2n \) generating function coordinates to be independent. For example, if the new coordinates \( Q_i \equiv Q_i(q,t) \) do not depend on the old momenta, then the coordinates \( Q \) and \( q \) are not independent, and so a generating function \( F(q,Q,t) \) would not work. Canonical transformations highlight the point that the Hamiltonian perspective regards both \( q \) and \( p \) equally as independent coordinates. In particular, the transformation \( Q_i = q_i, \quad P_i = -q_i \) leaves the Hamiltonian formulation unchanged, which would not be legal in the Lagrangian framework.

Lastly, we can characterize time-independent canonical transformations with Poisson brackets. Statement (c) is often how a transformation is verified as canonical in practice.

**Proposition 6.8.** For a time-independent transformation \((q,p) \rightarrow (Q,P)\) the following conditions are equivalent.

(a) The transformation is canonical, i.e. \( p \, dq = P \, dQ + dS \).
6.5. CANONICAL TRANSFORMATIONS

(b) The transformation preserves the Poisson bracket:
\[ \{ f, g \}_{p,q} = \{ f, g \}_{P,Q} \]
for all smooth functions \( f \) and \( g \) on phase space, where the subscripts denote the respective variables of differentiation.

(c) The new variables satisfy the coordinate relations with respect to the old Poisson brackets:
\[ \{ Q_i, Q_j \}_{p,q} = 0, \quad \{ P_i, P_j \}_{p,q} = 0, \quad \{ P_i, Q_j \}_{p,q} = \delta_{ij}. \]

Proof. (b) implies (c): This is immediate, since we can replace the differentiation variables \( q, p \) with \( Q, P \) by assumption—for example,
\[ \{ Q_i, Q_j \}_{p,q} = \{ Q_i, Q_j \}_{P,Q} = 0. \]

(c) implies (b): By premise, we compute
\[
\left( \begin{array}{cc}
\frac{\partial P}{\partial q} & \frac{\partial P}{\partial p} \\
\frac{\partial Q}{\partial q} & \frac{\partial Q}{\partial p}
\end{array} \right) \left( \begin{array}{cc}0 & -I \\
I & 0 \end{array} \right) \left( \begin{array}{cc}
\frac{\partial P}{\partial q} & \frac{\partial P}{\partial p} \\
\frac{\partial Q}{\partial q} & \frac{\partial Q}{\partial p}
\end{array} \right) = \left( \begin{array}{cc}
\frac{\partial P}{\partial q} & \frac{\partial P}{\partial p} \\
\frac{\partial Q}{\partial q} & \frac{\partial Q}{\partial p}
\end{array} \right) \left( \begin{array}{cc}0 & -I \\
I & 0 \end{array} \right) \left( \begin{array}{cc}
\frac{\partial P}{\partial q} & \frac{\partial P}{\partial p} \\
\frac{\partial Q}{\partial q} & \frac{\partial Q}{\partial p}
\end{array} \right) \left( \begin{array}{cc}0 & -I \\
I & 0 \end{array} \right).
\]

Plugging this in, we have
\[
\{ f, g \}_{p,q} = -\left( \begin{array}{cc}
\frac{\partial f}{\partial q} & \frac{\partial f}{\partial p}
\end{array} \right) \left( \begin{array}{cc}0 & -I \\
I & 0 \end{array} \right) \left( \begin{array}{cc}
\frac{\partial g}{\partial q} & \frac{\partial g}{\partial p}
\end{array} \right)
= -\left( \begin{array}{cc}
\frac{\partial f}{\partial P} & \frac{\partial f}{\partial Q}
\end{array} \right) \left( \begin{array}{cc}0 & -I \\
I & 0 \end{array} \right) \left( \begin{array}{cc}
\frac{\partial g}{\partial P} & \frac{\partial g}{\partial Q}
\end{array} \right) = \{ f, g \}_{P,Q}
\]

(a) implies (c): The time-independence requires that the new Hamiltonian \( K(Q, P, t) = H(q, p, t) \) is the same as the old Hamiltonian. On the one hand, the chain rule and Hamilton’s equations insist
\[
\dot{Q}_i = \frac{\partial Q_i}{\partial q} \cdot \dot{q} + \frac{\partial Q_i}{\partial p} \cdot \dot{p} = \frac{\partial Q_i}{\partial q} \cdot \frac{\partial H}{\partial q} - \frac{\partial Q_i}{\partial p} \cdot \frac{\partial H}{\partial p} = \{ H, Q_i \}_{p,q}.
\]

On the other hand, the new Hamilton’s equations (6.5) tell us that this should be equal to
\[
\frac{\partial H}{\partial P_i} = \frac{\partial H}{\partial q} \cdot \frac{\partial q}{\partial P_i} + \frac{\partial H}{\partial p} \cdot \frac{\partial p}{\partial P_i}.
\]
CHAPTER 6. HAMILTON'S EQUATIONS OF MOTION

Equating these, we have

$$\frac{\partial Q_i}{\partial p_j} = -\frac{\partial q_j}{\partial P_i}, \quad \frac{\partial Q_i}{\partial q_j} = \frac{\partial p_j}{\partial P_i}. \quad \text{(6.25)}$$

The analogous computation for $\dot{P}_i$ lead to the relations

$$\frac{\partial P_i}{\partial p_j} = \frac{\partial q_j}{\partial Q_i}, \quad \frac{\partial P_i}{\partial q_j} = -\frac{\partial p_j}{\partial Q_i}. \quad \text{(6.26)}$$

Together these are exactly the conditions of (c).

(c) implies (a): The same computation from the previous implication shows that $d\mathbf{p} \wedge d\mathbf{q} = d\mathbf{P} \wedge d\mathbf{Q}$. Using Stokes’ theorem to pass to the boundary, we deduce that for any closed contour $\gamma$ contractible to a point we have

$$\oint_{\gamma} p \, dq = \oint_{\gamma} P \, dQ. \quad \text{(6.27)}$$

Therefore we conclude that the difference form $p \, dq - P \, dQ$ is closed since its antiderivative is well-defined.

6.6 Infinitesimal canonical transformations

A Hamiltonian flow can be thought of as a continuous canonical transformation. Consider a Hamiltonian system undergoing a canonical transformation

$$Q_i = q_i + \delta q_i, \quad P_i = p_i + \delta p_i, \quad i = 1, \ldots, d, \quad \text{(6.28)}$$

which is infinitesimal in the sense that we will be sending $\delta q_i$ and $\delta p_i$ to zero later to obtain a statement about derivatives. We expect the corresponding generating function to be a perturbation of that which generates the identity transformation. It is the generating function

$$\Phi_0(q, P, t) = \sum_{i=1}^n q_i P_i \quad \text{(6.29)}$$

which yields the identity transformation, since according to eq. (6.26) we have

$$p_i = \frac{\partial \Phi_0}{\partial q_i} = P_i, \quad Q_i = \frac{\partial \Phi_0}{\partial P_i} = q_i, \quad K(Q, P) = H(q, p). \quad \text{(6.27)}$$

Consequently, we seek a generating function for the transformation (6.27) of the form

$$\Phi(q, P, t) = \sum_{i=1}^n q_i P_i + \epsilon G(q, P) \quad \text{(6.28)}$$
for some small parameter $\epsilon$. We will be considering the first order behavior as $\epsilon \to 0$, and $\delta q_i$ and $\delta p_i$ will depend linearly on $\epsilon$. By eq. (6.26), the function $\Phi$ generates the transformation

\[ p_i = \frac{\partial \Phi}{\partial q_i} = P_i + \epsilon \frac{\partial G}{\partial q_i}, \quad Q_i = \frac{\partial \Phi}{\partial P_i} = q_i + \epsilon \frac{\partial G}{\partial p_i}. \]

The leading term of the derivative $\frac{\partial p_i}{\partial P_i} = \frac{\partial}{\partial P_i} (P_i - \delta p_i) = 1 + \mathcal{O}(\delta p_i)$ is one, and with the factor of $\epsilon$ in front this derivative does not make a first order contribution. We made this change of variables so that we may take the function $G = G(q,p)$ to be dependent only on the old coordinates and momenta $q, p$.

Solving for $\delta q_i$ and $\delta p_i$ in their definition (6.27), we get

\[ \delta p_i = P_i - p_i = -\epsilon \frac{\partial G}{\partial q_i} + \mathcal{O}(\epsilon^2), \quad \delta q_i = Q_i - q_i = \epsilon \frac{\partial G}{\partial p_i} + \mathcal{O}(\epsilon^2). \] (6.29)

If we take the function $G = H(q,p)$ to be the system’s Hamiltonian and let the parameter $\epsilon$ be an infinitesimal time interval $\delta t$, this becomes

\[ \delta q_i = \delta t \frac{\partial H}{\partial p_i} + \mathcal{O}(\delta t^2), \quad \delta p_i = -\delta t \frac{\partial H}{\partial q_i} + \mathcal{O}(\delta t^2). \] (6.30)

Taking $\delta t \to 0$ we recover Hamilton’s equations, which indicates that the Hamiltonian is the instantaneous generator of the system’s motion with time.

We will now consider how the Hamiltonian is affected by this transformation. Let $u(q,p)$ be a smooth function, which is perturbed as the result of the transformation (6.27) by

\[ u(q_i + \delta q_i, p_i + \delta p_i) = u(q_i, p_i) + \delta u \] (6.31)

Taylor expanding, we have

\[ \delta u = \sum_{i=1}^d \left( \frac{\partial u}{\partial q_i} \delta q_i + \frac{\partial u}{\partial p_i} \delta p_i \right) + \mathcal{O}(\delta q_i^2 + \delta p_i^2) \]

\[ = \sum_i \left( \frac{\partial u}{\partial q_i} \frac{\partial G}{\partial p_i} - \frac{\partial u}{\partial p_i} \frac{\partial G}{\partial q_i} \right) + \mathcal{O}(\epsilon^2) = -\epsilon \{u, G\} + \mathcal{O}(\epsilon^2) \] (6.32)

after plugging in for $\delta q_i$ and $\delta p_i$ using eq. (6.29). If we take the function $u(q,p) = H(q,p)$ to be the Hamiltonian, then

\[ \left. \frac{dH}{d\epsilon} \right|_{\epsilon=0} = -\{H, G\}. \] (6.33)

Since $G$ is time-independent, then this quantity vanishes if and only if $G$ is a constant of the motion, and so constants of the motion are instantaneous generators of those transformations which leave the Hamiltonian invariant. This is another appearance of Noether’s theorem (Proposition 3.12).
6.7 Canonical variables

For any system with bounded periodic motion there is always a change of variables so that the motion is dictated by a linearly increasing angular variable and an energy parameter. In this section we will consider a conservative system with one degree of freedom and bounded motion. As in section 5.3 this analysis also applies coordinate-wise to a system with periodic motion in each variable (but not necessarily jointly periodic motion).

Since the motion is bounded and necessarily periodic, we may consider the action variable

\[ I = \frac{1}{2\pi} \oint p \, dq \]  

(6.34)

which was first glimpsed in the period integral (2.4) and we later identified in eq. (5.20). The integral is evaluated over one period, around a loop in the phase plane around which \( q \) returns to an extreme value once. We may also write

\[ I = \frac{1}{2\pi} \iint dp \, dq \]  

(6.35)

where the double integral is evaluated over the region enclosed by the trajectory \((q(t), p(t))\) in phase space.

We would like to treat \( I \) as a momentum variable in order to obtain the Hamiltonian and Hamilton’s equations in terms of this geometric quantity. Consider the generating function

\[ \Phi = \int p(q, E) \, dq \]  

(6.36)

where \( E \equiv H(q(t), p(t)) \) is the constant energy along a trajectory and \( p(q, E) \) is implicitly determined. For our conservative system \( I \equiv I(E) \) is strictly a function of the energy, and so the generating function \( \Phi \equiv \Phi(q, I) \) may be considered a function of \( I \) rather than \( E \). The first equation of eq. (6.26) requires that

\[ p = \frac{\partial \Phi}{\partial q} \]  

(6.37)

which agrees with the fundamental theorem of calculus applied to the definition (6.36). The new coordinate—which we will denote by \( w \)—is dictated by the second equation of eq. (6.26):

\[ w = \frac{\partial \Phi}{\partial I} \]  

(6.38)

Together, \( w \) and \( I \) are called the canonical variables of the system, and individually \( I \) is the action variable and \( w \) is the angle variable.

Since the generating function \( \Phi(q, I, \lambda) \) does not explicitly depend on time, the last equation of eq. (6.26) tells us that the Hamiltonian is unaffected:

\[ K = H + \frac{\partial \Phi}{\partial t} = H, \]  

(6.39)
and is just the old one in terms of the new variables. However, since the action variable \( I \equiv I(E) \) is determined by the energy then in the new variables \( H \equiv E \equiv E(I) \). Hamilton’s equations are now

\[
\dot{I} = -\frac{\partial H}{\partial w} = 0, \quad \dot{w} = \frac{\partial H}{\partial I} = \frac{dE}{dI} \tag{6.40}
\]

The first equation is expected, since \( I \) is given by the system’s energy which is constant for a conservative system. The latter can be easily integrated to obtain

\[
w(t) = \frac{dE}{dI} t + \text{constant} = \omega(I) t + \text{constant}, \tag{6.41}
\]

and so \( w \) is the phase of the oscillations.

From their definitions (6.34) and (6.36) we can see that after each oscillation the generating function \( \Phi \) increments by \( 2\pi I \). In turn, the equation of motion (6.41) tells us that the angle variable \( w \) increases by \( 2\pi \) with each oscillation, much like an actual angle coordinate. Conversely, the original variables \( q \) and \( p \) are unchanged after an integer number of periods.

### 6.8 Exercises

#### 6.1 (Harmonic oscillator).

The one-dimensional harmonic oscillator has the kinetic and potential energies

\[
T = \frac{1}{2}mv^2, \quad U = \frac{1}{2}kx^2.
\]

Since this system is conservative, the Hamiltonian \( H = E \) is given by the total energy

\[
H = T + U = \frac{1}{2}m\dot{p}^2 + \frac{m\omega^2}{2}x^2
\]

where \( \omega^2 = k/m \).

(a) Directly apply Hamilton’s equations to obtain \( \ddot{x} + \omega^2 x = 0 \), which has solution

\[
x(t) = x_0 \cos [\omega(t - t_0)]
\]

as in Example 1.11. Hamilton’s equations may not make solving the motion easier than using Lagrange’s or Newton’s equations, but rather provide a new interpretation of the system’s motion.

(b) By considering the shape of trajectories in phase space, find the angle variable \( I \) and show that the energy as a function of the angle variable \( I \) is \( E(I) = \omega I \).

(c) Write down (but do not evaluate) the integral for the generating function \( \Phi \), and show that the angle variable \( w \) is

\[
w = \tan^{-1} \left( \frac{x}{\sqrt{\frac{2I}{m\omega} - x^2}} \right) + \text{constant}.
\]
(d) By design, the angle variable $w$ is cyclic for the new Hamiltonian $E(I) = \omega I$. From the equality

$$w(t) = \frac{dE}{dI} t + \text{constant} = \omega t + \text{constant},$$

invert this equation to conclude

$$x(t) = x_0 \cos [\omega (t - t_0)]$$

as in part (a). This method is unwieldy in application, and is intended for the purpose of physical understanding rather than solving for the equations of motion.

6.2 (Charged particle in an electromagnetic field). In Exercise 3.8 we found the Lagrangian for a charged particle in $\mathbb{R}^3$ with charge $q$ and mass $m$ moving through an electromagnetic field to be

$$\mathcal{L} = \frac{1}{2} mv^2 - q\phi + \frac{q}{c} v \cdot A$$

where $\phi$ and $A$ are the scalar and vector potentials for the electric and magnetic fields:

$$B = \nabla \times A, \quad E = -\nabla \phi - \frac{1}{c} \frac{\partial A}{\partial t}.$$

(a) Show that the Hamiltonian for this system is

$$H = \frac{1}{2m} \left(p - \frac{q}{c} A\right)^2 + q\phi.$$

(b) Although the electric and magnetic fields are uniquely expressed, the scalar and vector potentials that $\phi$ and $A$ are not unique and they appear explicitly in the Hamiltonian. In fact, substituting $A' = A + \nabla f$ for any smooth function $f(x,y,z,t)$ leaves $B$ unchanged since the curl of a gradient is also zero. Show that for $E$ to remain unchanged we need to also substitute

$$\phi' = \phi - \frac{1}{c} \frac{\partial f}{\partial t}.$$

Together, replacing $(A,\phi) \rightarrow (A',\phi')$ is called a gauge transformation, and a specific pair $(A,\phi)$ is called a choice of gauge.

(c) Since the electric and magnetic fields are unaffected by the choice of gauge, then any physical laws in terms of the potentials should also be invariant. Show that under a gauge transformation the Hamiltonian becomes

$$H' = H - \frac{q}{c} \frac{\partial f}{\partial t},$$

and that Hamilton’s equations still hold in the new variables.
6.3 (Young’s inequality). Show that for any Legendre transform pair \( f(x) \) and \( f^*(\xi) \) we have
\[ x \cdot \xi \leq f(x) + f^*(\xi). \]
Apply this inequality to the function \( f(x) \) of Example 6.1.

6.4 (Example of Poisson’s theorem). Show that for the angular momentum \( L = x \times p \) of a particle \( x \in \mathbb{R}^3 \) we have
\[ \{ L, L \cdot n \} = L \times n \]
for any unit vector \( n \). Do so by picking a choice of Cartesian coordinates with respect to \( n \), writing down the infinitesimal canonical transformation \([6.27]\) given by rotation about the \( n \) axis by an angle \( d\theta \), and apply the calculation \([6.32]\) with \( u \) replaced by \( L \). Conclude that \( \{ L_i, L_j \} = -L_k \) for \( i, j, k \in \{1, 2, 3\} \) in cyclic order.
Chapter 7

Symplectic geometry

Symplectic geometry is the differential geometric generalization of the time-independent Hamiltonian structure of phase space to general manifolds, and the physical results for conservative systems from the previous chapter can be lifted to this perspective. The material for this section is based on [Arn89, Ch. 8] chapter 8 and [Lee13, Ch. 22], and we will try to meld the physics and mathematical notations.

7.1 Symplectic structure

Let \( M \) be a smooth even-dimensional manifold of dimension \( 2n \). A \textit{symplectic form} or \textit{symplectic structure} on \( M \) is a two-form \( \omega \) on \( M \) that is closed (\( d\omega = 0 \)) and \textit{nondegenerate}: for each \( x \in M \) the map \( T_xM \to T^*_xM \) which takes \( \xi \mapsto \xi \lrcorner \omega = \omega(\xi, \cdot) \) is invertible. The pair \((M, \omega)\) is called a \textit{symplectic manifold}.

There are other more easily verifiable criteria for a two-form \( \omega \) (cf. Exercise 7.1), but conceptually this is the classification we will rely upon. Heuristically, the symplectic form \( \omega \) is an identification of the tangent and cotangent spaces at each point in \( M \), which bears resemblance to Riemannian structure but as we shall see exhibits drastically different behavior.

Example 7.1. Consider the cotangent bundle \( T^*N \) of a \( n \)-dimensional manifold \( N \), or simply the Euclidean space \( T^*\mathbb{R}^n = \mathbb{R}^n_q \times \mathbb{R}^n_p \). Then the two-form

\[
\omega = dp \wedge dq = \sum_{i=1}^{n} dp^i \wedge dq^i \tag{7.1}
\]

is a symplectic form on \( T^*N \), and is called the \textit{canonical symplectic form} on the cotangent bundle; we will later see how this form naturally encapsulates Hamilton’s equations. This form is closed, since it is the exterior derivative of the one-form

\[
p dq = \sum_{i=1}^{n} p^i \ dq^i, \tag{7.2}
\]
which is in turn called the tautological one-form on $T^*M$; we first saw this form in the period integral (2.4) and then again later in defining the action variable (5.20). To see that $\omega$ is nondegenerate, fix $(q, p) \in T^*N$ and let $(v_1, \ldots, v_n, a_1, \ldots, a_n) \in T_{(q, p)}(T^*N)$ denote the dual basis to $(dq, dp)$ so that

$$dq^i(v_j) = \delta_{ij}, \quad dp^j(a_j) = \delta_{ij}, \quad dq^i(a_j) = dp^j(v_j) = 0$$

for $i, j = 1, \ldots, n$. Often in differential topology we write $\partial q_i$ for the dual vector to $dq_i$, but physically we like to think of the tangent vector to a point in phase space as the velocity and (mass-times-)acceleration. The action of $\omega$ on these basis vectors is

$$\omega(v_i, v_j) = \omega(a_i, a_j) = 0, \quad \omega(v_i, a_j) = -\omega(a_j, v_i) = \delta_{ij}$$

for $i, j = 1, \ldots, n$. To check nondegeneracy, fix $\xi \in T_{(q, p)}(T^*N)$ and suppose we have $\omega(\xi, \eta) = 0$ for all $\eta \in T_{(q, p)}(T^*N)$. Expanding $\xi = \sum_{i=1}^n (b^i v_i + c^i a_i)$ for constants $b^i, c^i \in \mathbb{R}$, we see that

$$0 = \omega(\xi, v_i) = -c^i, \quad 0 = \omega(\xi, a_i) = b^i$$

and so $\xi = 0$ as desired.

In fact, the form of Example 7.1 is the fundamental example in the following sense.

**Theorem 7.2** (Darboux’s theorem). Let $(M, \omega)$ be a $2n$-dimensional symplectic manifold. For any $\xi \in M$ there exists local coordinates $(q, p)$ centered at $\xi$ with respect to which $\omega$ has the representation (7.1).

For a proof see [Lee13, Th. 22.13].

**Remark.** Symplectic manifolds are automatically orientable because $\omega^n$ is a nonvanishing 2n-form on $M$. Indeed, if we fix $x \in M$ and write $\omega = \sum_i dp^i \wedge dq^i$ at $x$, then the $n$-fold wedge product

$$\omega^n = \sum_I dp^{i_1} \wedge dq^{i_1} \wedge \cdots \wedge dp^{i_n} \wedge dq^{i_n}$$

where the sum ranges over all multi-indices $I = (i_1, \ldots, i_n)$ of length $n$. Any term where $I$ contains a repeated index is zero because $dq^i \wedge dq^i = 0$, and since 2-forms commute under the wedge product then we can rewrite this as

$$\omega^d = d! dp^1 \wedge dq^1 \wedge \cdots \wedge dp^d \wedge dq^d.$$

This is nonvanishing at $x$ by definition of the nondegeneracy of $\omega$.

### 7.2 Hamiltonian vector fields

Next let us see how to use the symplectic structure to obtain the familiar dynamics from a Hamiltonian. Let $(M, \omega)$ be a symplectic manifold and $H : M \to \mathbb{R}$
be a smooth function. Then $dH$ is a differential one-form which associates a covector to each point in $M$. On the other hand, for Hamilton’s equations we need to specify a vector field for the right-hand side of the differential equation. By definition, for each $x \in M$ the map $T_xM \to T^*_xM$ which takes $\xi \mapsto \xi \omega$ is invertible, and so let $J : T^*_xM \to T_xM$ denote its inverse. Then $JdH$ is a vector field on $M$, called the Hamiltonian vector field associated to the Hamiltonian $H$. The induced flow $e^{tJdH} : M \to M$, $x_0 \mapsto x(t)$ defined to solve the ODE

$$\dot{x}(t) = JdH(x(t)), \quad x(0) = x_0$$

(7.3)

is called a Hamiltonian flow on $M$. In differential geometry the notation $\hat{\omega} : T^*_xM \to T_xM$ is used for the map $\xi \mapsto \xi \omega$, and so in place of $J$ the notations $\hat{\omega}^{-1}$ and

$$X_H := \hat{\omega}^{-1}(dH) \iff X_H \omega = dH$$

(7.4)

are often used.

**Example 7.3.** Consider the canonical symplectic form (7.1) on Euclidean phase space $T^*\mathbb{R}^n = \mathbb{R}^{nq} \times \mathbb{R}^{np}$. On Euclidean space we already have a natural identification of vectors and covectors via the dot product, and we can express $J$ as a linear transformation in terms of this identification. Given a vector field

$$X = \sum_{i=1}^n \left( b^i \frac{\partial}{\partial q^i} + c^i \frac{\partial}{\partial p^i} \right)$$

for some smooth coefficients $b^i$ and $c^i$, we compute

$$X \omega = \sum_{i=1}^n \left( c^i dq^i - b^i dp^i \right) = \begin{pmatrix} c \\ -b \end{pmatrix} = J \begin{pmatrix} b \\ c \end{pmatrix}$$

and hence

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$$

(7.5)

for $I$ the identity matrix. Writing

$$dH = \frac{\partial H}{\partial q} dq + \frac{\partial H}{\partial p} dp = \sum_{i=1}^n \left( \frac{\partial H}{\partial q^i} dq^i + \frac{\partial H}{\partial p^i} dp^i \right) = \begin{pmatrix} \frac{\partial H}{\partial q} \\ -\frac{\partial H}{\partial p} \end{pmatrix},$$

we have

$$\begin{pmatrix} \dot{q} \\ \dot{p} \end{pmatrix} = JdH = \begin{pmatrix} \frac{\partial H}{\partial p} \\ -\frac{\partial H}{\partial q} \end{pmatrix}$$

which agrees with Hamilton’s equations (6.4).

Symplectic geometry is the generalization of time-independent Hamiltonian dynamics since the Hamiltonian function $H$ is automatically conserved under its Hamiltonian flow.
Proposition 7.4 (Conservation of energy). Let \((M, \omega)\) be a symplectic manifold and let \(H\) be a smooth function on \(M\). Then \(H\) is constant along the integral curves of \(J \, dH\), and when \(dH \neq 0\) the vector field \(dH\) is tangent to the level sets of \(H\).

Proof. Both are a consequence of
\[
(J \, dH)(H) = dH(J \, dH) = \omega(J \, dH, J \, dH) = 0,
\]
which holds since \(\omega\) is alternating.

\[\square\]

7.3 Integral invariants

Consider a smooth diffeomorphism \(g : M \to M\). A \(k\)-form \(\eta\) is said to be an integral invariant for \(g\) if
\[
\int_{g(N)} \eta = \int_N \eta
\]
for all orientable \(k\)-dimensional submanifolds \(N\) with boundary.

Theorem 7.5. A Hamiltonian flow \(g_t = e^{tJ \, dH}\) preserves the symplectic structure: \(g_t^* \omega = \omega\). In other words, the symplectic form \(\omega\) is an integral invariant of \(g_t\).

Proof. The map \(g_t\) is smoothly homotopic to the identity via the family of diffeomorphisms \(g_s\), \(s \in [0, t]\), in the sense that at time \(s = 0\) the map \(g_0 : M \to M\) is the identity, and at time \(s = t\) the map \(g_t : M \to M\) is what we are given. Fix \(N\) a smooth orientable 2-dimensional submanifold, and let \(\Omega_N := \{g_s(N) : s \in [0, t]\}\) denote the image of \(N\) under the homotopy. We can think of \(\omega_N\) as an orientable 3-manifold in \([0, t] \times M\) or as being immersed in \(M\). With this choice of orientation we have
\[
\partial \Omega_N = g_t(N) \cup (-N) \cup (-\partial N). \tag{7.7}
\]

We claim that for any smooth curve \(\gamma\) in \(M\) we have
\[
\frac{d}{dt} \int_{\Omega_\gamma} \omega = \int_{g^*(\gamma)} dH, \tag{7.8}
\]
where \(H\) is the Hamiltonian for the flow \(g^t\). Let \(\phi : [a, b] \to M\) be a parametrization of \(\gamma\). Then \(\Omega_\gamma\) is parameterized by \(\Phi(s, x) := g^*(\phi(x))\) and we have
\[
\int_{\Omega_\gamma} \omega = \int_a^b \int_0^t \omega \left( \frac{\partial \Phi}{\partial x}, \frac{\partial \Phi}{\partial s} \right) dx \, ds
\]
where \(\frac{\partial \Phi}{\partial s}(s, x)\) is in \(T_{\Phi(s, x)} M\). By definition of the Hamiltonian phase flow \(g_t = e^{tJ \, dH}\) the tangent vector \(\frac{\partial \Phi}{\partial s}\) points in the direction of \(J \, dH\), and so we have \(\omega(\frac{\partial \Phi}{\partial x}, \frac{\partial \Phi}{\partial s}) = dH(\frac{\partial \Phi}{\partial x})\). Therefore
\[
\int_{\Omega_\gamma} \omega = \int_0^t \int_a^b dH \left( \frac{\partial \Phi}{\partial x} \right) dx \, ds = \int_0^t \left( \int_{g^t(\gamma)} dH \right) ds,
\]
and the identity (7.8) follows from the fundamental theorem of calculus.

For a closed curve like \( \partial N \) we note that

\[
\int_{g'(\partial N)} dH = \int_{g'(\emptyset)} H = 0,
\]

and so the identity (7.8) implies

\[
\int_{\Omega_{\partial N}} \omega = 0. 
\]  

(7.9)

Since \( \omega \) is closed by definition, then by Stokes' theorem we have

\[
0 = \int_{\Omega_N} d\omega = \int_{\partial \Omega_N} \omega. 
\]

Decomposing the boundary \( \partial \Omega_N \) according to (7.7) we obtain

\[
0 = \int_{g(N)} \omega - \int_{\Omega_{\partial N}} \omega. 
\]

From (7.9) we know the last integral vanishes, and so we conclude that \( \omega \) is an integral invariant of \( g_t \).

In the previous section we saw that \( \omega^n \) defines a volume form on \( M \) and so we immediately obtain an analog of Theorem 6.4.

**Corollary 7.6 (Liouville’s theorem).** Each of the forms \( \omega^2, \omega^4, \ldots \) is preserved by a Hamiltonian flow. In particular, every Hamiltonian flow preserves the volume form \( \omega^n \).

Recall from section 6.5 that on a cotangent bundle \( M = T^*Q \) with canonical symplectic form \( \omega = dp \wedge dq \), a (time-independent) canonical transformation \( g : T^*Q \to T^*Q \) satisfies

\[
g^*(p \, dq) = p \, dq + dS. 
\]  

(7.10)

This definition does not make \( p \, dq \) an integral invariant for a canonical transformation, since the condition (7.6) only holds for closed curves \( N \). Instead, we employ the following useful observation.

**Proposition 7.7.** Let \( g : M \to M \) be a smooth diffeomorphism. If \( \eta \) is a k-form such that (7.6) holds for only closed orientable k-submanifolds \( N \), then \( d\eta \) is an integral invariant.

**Proof.** Let \( N \) be an orientable \((k+1)\)-submanifold. Then by Stokes’ theorem we have

\[
\int_N d\eta = \int_{\partial N} \eta, \quad \int_{g(N)} d\eta = \int_{\partial g(N)} \eta = \int_{g(\partial N)} \eta.
\]

Since \( \partial N \) is closed then these two right-hand sides are equal by premise. Therefore we conclude that (7.6) holds for the form \( d\eta \), and hence \( d\eta \) is an integral invariant. \( \Box \)
Noting that \( d(p\,dq) = dp \wedge dq \), we conclude:

**Corollary 7.8.** Canonical transformations preserve the symplectic form \( \omega \) and the volume form \( \omega^n \).

**Remark.** In section 6.6 we saw that for Euclidean phase space \( T^*\mathbb{R}^n \) Hamiltonian flows are canonical transformations. However, the converse of Proposition 7.7 is not true in general, and so Theorem 7.5 does not imply the identity (7.10). We must assume that \( M \) is simply connected in order for Hamiltonian flows to be canonical transformations, since then Theorem 7.5 implies (7.10).

### 7.4 Poisson bracket

In section 6.4 we saw how to phrase Hamiltonian dynamics in terms of the Poisson bracket, and in this section we will see how this notion manifests in terms of symplectic structure. For smooth functions \( f, g \in C^\infty(M) \) on a symplectic manifold \( (M, \omega) \) we define the *Poisson bracket* of \( f \) and \( g \) to be

\[
\{f, g\} = \omega(Jdf, Jdg) = dg(Jdf) = (Jdf)(g). \tag{7.11}
\]

As in (6.18) the Poisson bracket \( \{H, f\} \) is also the evolution of the quantity \( f \) under the Hamiltonian flow \( H \), since the Lie derivative of the function \( f \) along \( J\,dH \) is given by

\[
\mathcal{L}_{J\,dH}f = \left. \frac{d}{dt} \right|_{t=0} f \circ e^{t\,J\,dH} = (J\,dH)(f) = \{H, f\} \tag{7.12}
\]

according to the definition (7.3) of the Hamiltonian flow. In particular, we again have that \( f \) is conserved by the Hamiltonian flow of \( H \) if and only if \( \{H, f\} = 0 \).

**Example 7.9.** Let us check that the new definition (7.11) agrees with the phase space definition (6.17) on Euclidean space \( T^*\mathbb{R}^n = \mathbb{R}^n_q \times \mathbb{R}^n_p \). Using the calculation of \( J \) from Example 7.3 we have

\[
J\,df = J \left( \frac{\partial f}{\partial q} \,dq + \frac{\partial f}{\partial p} \,dp \right) = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \left( \frac{\partial f}{\partial q} \begin{array}{c} \frac{\partial}{\partial p} \\ \frac{\partial}{\partial q} \end{array} \right) = \frac{\partial f}{\partial p} \frac{\partial}{\partial q} - \frac{\partial f}{\partial q} \frac{\partial}{\partial p}.
\]

Consequently, the definition (7.11) yields

\[
\{f, g\} = (J\,df)(g) = \frac{\partial f}{\partial p} \frac{\partial g}{\partial q} - \frac{\partial f}{\partial q} \frac{\partial g}{\partial p},
\]

which agrees with the first definition (6.17).

The properties listed in section 6.4 resemble those of the commutator because the Poisson bracket at the level of functions corresponds to the commutator of the respective Hamiltonian vector fields.
7.5. TIME-DEPENDENT SYSTEMS

**Proposition 7.10.** If \((M, \omega)\) is a symplectic manifold, then the Poisson bracket is the Hamiltonian of the commutator of the corresponding Hamiltonian vector fields:

\[
J \mathrm{d}\{f, g\} = [J \mathrm{d}f, J \mathrm{d}g].
\]  

(7.13)

In particular, the vector space \(C^\infty(M)\) is a Lie algebra under the Poisson bracket: the Poisson is bilinear, antisymmetric, and satisfies the Jacobi identity.

**Proof.** In fact, it suffices to prove that the Poisson bracket satisfies the Jacobi identity

\[
\{\{f, g\}, h\} + \{\{g, h\}, f\} + \{\{h, f\}, g\} = 0
\]  

(7.14)

for arbitrary smooth functions \(f, g, h \in C^\infty(M)\). This is sufficient because we can use the antisymmetry of \(\omega\) to write

\[
\{\{f, g\}, h\} = \{f, \{g, h\}\} - \{g, \{f, h\}\}.
\]

and recognizing the Poisson bracket as a Lie derivative via (7.12), we have

\[
\mathcal{L}_{J \mathrm{d}\{f, g\}} h = \mathcal{L}_{J \mathrm{d}f} \mathcal{L}_{J \mathrm{d}g} h - \mathcal{L}_{J \mathrm{d}g} \mathcal{L}_{J \mathrm{d}f} h = \mathcal{L}_{\{[J \mathrm{d}f, J \mathrm{d}g]\}} h.
\]

Since \(h \in C^\infty(M)\) is arbitrary, the identity (7.13) must hold.

It remains to demonstrate the Jacobi identity (7.14). The left-hand side of (7.14) is a combination of terms which each have at least one second order derivative. Isolating the terms containing second order derivatives of \(f\) and recognize the Poisson bracket as a Lie derivative via (7.12), we have

\[
\{\{f, g\}, h\} + \{\{g, h\}, f\} - \{\{h, f\}, g\} = \mathcal{L}_{J \mathrm{d}h} \mathcal{L}_{J \mathrm{d}g} f - \mathcal{L}_{J \mathrm{d}g} \mathcal{L}_{J \mathrm{d}h} f.
\]

However, we know \(\mathcal{L}_{J \mathrm{d}h} \mathcal{L}_{J \mathrm{d}g} - \mathcal{L}_{J \mathrm{d}g} \mathcal{L}_{J \mathrm{d}h} = \mathcal{L}_{\{J \mathrm{d}h, J \mathrm{d}g\}}\) is only a first order differential operator, and so we conclude that the terms with second order derivatives of \(f\) vanish. Lastly, since the left-hand side of (7.14) is symmetric in \(f, g,\) and \(h,\) we conclude that there can be no second order derivatives and hence must vanish. \(\square\)

### 7.5 Time-dependent systems

Recall that Hamilton’s equations (6.4) remain valid for time-dependent Hamiltonian systems, while so far we have only allowed \(H\) to be a smooth function on phase space \(M\). This section is dedicated to extending the symplectic geometry developed thus far to time-dependent Hamiltonian systems.

Let \((M, \omega)\) be a 2n-dimensional symplectic manifold, and define the extended phase space \(M \times \mathbb{R}\). Given a possibly time-dependent Hamiltonian \(H\), we define the Poincaré–Cartan one-form locally in terms of the canonical coordinates \((q, p)\) on \(M\) guaranteed by Theorem 7.2 as

\[
\tau = p \mathrm{d}q - H \mathrm{d}t = \sum_{i=1}^{n} p^i \mathrm{d}q^i - H \mathrm{d}t.
\]  

(7.15)
CHAPTER 7. SYMPLECTIC GEOMETRY

Note that the first term above is the tautological one-form $pdq$ on $M$, the
differential of which yields the symplectic form $\omega$. This is the form that we
insisted be preserved by a canonical transformation in eq. (6.24), and the same
notion of canonical transformations holds on a general extended phase space
$M \times \mathbb{R}$ in terms of the canonical coordinates $(q, p, t)$.

Remark. The extended phase space $M \times \mathbb{R}$ together with the Poincaré–Cartan
one-form $\tau$ do indeed define a contact manifold, but as we will see in the next
chapter it is not the natural contact extension of $M$ since $\tau$ depends on the
system’s Hamiltonian $H$.

On $M \times \mathbb{R}$ we define the extended Hamiltonian vector field

$$Y_H = X_H + \frac{\partial}{\partial t},$$

(7.16)

where $X_H(t)$ is the Hamiltonian vector field on $M \times \{t\}$ defined by (7.3). In
analogy with the second condition of (7.4), the vector field (7.16) is the unique
solution to

$$Y_H \cdot d\tau = 0.$$  

(7.17)

The flow of $Y_H$ is given by

$$\dot{q}^i = \frac{\partial H}{\partial p^i}, \quad \dot{p}^i = -\frac{\partial H}{\partial q^i}, \quad i = 1,$$

(7.18)

which is just Hamilton’s equations (6.4) joined with the trivial equation $\dot{t} = 1$. It
follows that any smooth time-dependent function $f$ on $M \times \mathbb{R}$ evolves according to

$$\frac{df}{dt} = \{H, f\}_{p,q} + \frac{\partial f}{\partial t},$$

as was the case in section 6.4. In particular, a time-dependent Hamiltonian $H$
is no longer conserved under its own flow.

Example 7.11. [Cal41, Kan48] showed that in addition to naturally occurring
time-dependent Hamiltonian systems, this framework can also be applied to some
systems which are time-independent and dissipative by introducing arti-
tficial time dependence into the Hamiltonian. Consider the one-dimensional
Hamiltonian system

$$H(Q, P, t) := e^{-\gamma t} \frac{P^2}{2m} + e^{\gamma t} U(Q)$$

(7.19)
on the extended phase space $\mathbb{R}_Q \times \mathbb{R}_P \times \mathbb{R}_t$, where $\gamma \geq 0$ is a constant and the
coordinates $Q, P$ are related to the physical coordinates $q, p$ by the non-canonical
transformation

$$P = e^{\gamma t} p, \quad Q = q.$$  

Then the equations of motion (7.18) yield

$$m \ddot{q} + m \gamma \dot{q} + U'(q) = 0.$$

This represents a Newtonian system with a friction force that depends linearly
on the velocity, like the damped harmonic oscillator of Example 1.13.
7.6 Locally Hamiltonian vector fields

In section 7.1 we called a vector field $V$ on a symplectic manifold $(M,\omega)$ Hamiltonian if it is equal to $JdH$ for some smooth function $H \in \mathcal{C}^\infty(M)$, and in Theorem 7.5 we saw that $\omega$ is necessarily invariant under the flow of $JdH$. In fact, any smooth vector field $V$ on $M$ whose flow leaves $\omega$ invariant (i.e. $(e^{tV})^* \omega = \omega$) is called symplectic. However, a vector field being Hamiltonian is a global condition—in the sense that the corresponding Hamiltonian $H$ must extend smoothly to all of $M$—while being symplectic is a pointwise condition and is hence only local.

Consequently, a smooth vector field $V$ on $M$ is called locally Hamiltonian (as opposed to globally Hamiltonian) if for each point $p$ there exists a neighborhood on which $V$ is Hamiltonian. As an extension of Theorem 7.5 locally Hamiltonian vector fields are exactly those which are symplectic.

**Proposition 7.12.** Let $(M,\omega)$ be a symplectic manifold. A smooth vector field $V$ on $M$ is symplectic if and only if it is locally Hamiltonian. If $M$ is also simply connected, then every locally Hamiltonian vector field is globally Hamiltonian.

**Proof.** Differentiating the condition $(e^{tV})^* \omega = \omega$ with respect to $t$ at $t = 0$, we see that a vector field $V$ is symplectic if and only if $\mathcal{L}_V \omega = 0$. From Cartan’s magic formula we have

$$\mathcal{L}_V \omega = d(V_\omega) + V \lrcorner d\omega = d(V_\omega)$$

since $\omega$ is closed. Therefore $V$ is symplectic if and only if $V_\omega$ is closed.

If $V$ is locally Hamiltonian, then in a neighborhood of any point there exists a function $f$ so that $V = Jd(f)$, and hence $V_\omega = df$ which is certainly closed. Conversely, if $V$ is a symplectic vector field, then $V_\omega$ is closed and hence exact in a neighborhood of any point, and writing $V_\omega = df$ we deduce that $V = Jd(f)$ on a neighborhood as desired.

Now suppose $M$ is also simply connected. Then every closed one form is exact, and so $V_\omega$ is closed if and only if $V_\omega = df$ and $V = Jd(f)$ for a smooth function $f$ defined on all of $M$.

A smooth vector field $V$ is called an infinitesimal symmetry of the Hamiltonian $H$ if both $\omega$ and $H$ are invariant under the flow $e^{tV}$ of $V$. The second condition $(e^{tV})^* H = H$ can be recast as $V(H) = 0$, as can be seen by differentiating $H \circ e^{tV} = H$ with respect to $t$ at $t = 0$. With this notion, we can prove the following analogue of Proposition 3.12.

**Proposition 7.13 (Noether’s theorem).** Let $(M,\omega)$ be a symplectic manifold and $H$ a fixed Hamiltonian. If $f$ is a conserved quantity, then the Hamiltonian vector field $Jdf$ is an infinitesimal symmetry. Conversely, if $M$ is also simply connected then each infinitesimal symmetry is the Hamiltonian vector field of a conserved quantity, and the quantity is unique up to the addition of a function that is constant on each component of $M$. 

CHAPTER 7. SYMPLECTIC GEOMETRY

Proof. First suppose $f$ is a conserved quantity. Then from the identity (7.12) we know that $\{H, f\} = 0$. However, since the Poisson bracket is antisymmetric, we deduce that $0 = \{f, H\} = (J df)(H)$ as well. This demonstrates that $H$ is conserved by $J df$ and from Theorem 7.5 we know that $\omega$ is also conserved by $J df$, and hence $J df$ is an infinitesimal symmetry.

Now suppose that $M$ is simply connected and $V$ is an infinitesimal symmetry. Then $V$ is symplectic by definition, and by Proposition 7.12 we know $V$ is globally Hamiltonian. Writing $V = J df$, since $H$ is conserved by $f$ then we have $0 = (J df)(H) = \{f, H\} = -\{H, f\}$, and therefore $f$ is a conserved quantity. If $g$ is another function with $J dg = V = J df$, then by definition of $J$ we have $d(f - g) = (J df - J dg) \omega = 0$ and hence $f - g$ is constant on the components of $M$.

7.7 Exercises

7.1. Show that the following criteria for a nondegenerate two-form $\omega$ are equivalent.

(a) For each $x \in M$ the map $T_x M \to T^*_x M$ which takes $\xi \mapsto \omega(\xi, \cdot)$ is invertible.

(b) For each $x \in M$ and $\xi \in T_x M$ nonzero there exists $\eta \in T_x M$ such that $\omega(\xi, \eta) \neq 0$.

(c) The local matrix representation of $\omega$ in terms of some (hence every) basis is invertible.

7.2. (a) Show that rotation on the two-dimensional sphere $S^2$ is a Hamiltonian flow.

(b) Show that translations $g_t(q, p) = (q + t, p)$ on the torus $\mathbb{R}^2/\mathbb{Z}^2$ is a locally Hamiltonian flow, but is not globally Hamiltonian.

7.3. Let $(M, \omega)$ be a $2n$-dimensional compact symplectic manifold.

(a) Show that the $n$-fold wedge product $\omega^n$ is not exact.

(b) Show that the de-Rham cohomology groups $H^{2k}_{dR}(M)$ are nontrivial for $k = 1, \ldots, n$.

(c) Conclude that the two-dimensional sphere $S^2$ is the only sphere that admits a symplectic structure.

7.4. (Real symplectic matrices) A linear transformation $\mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is called symplectic if it preserves the canonical symplectic form (7.1). The set of all symplectic matrices forms a group with matrix multiplication and is often denoted by $Sp(2n, \mathbb{R})$. 
7.7. EXERCISES

(a) Prove that the determinant of any symplectic matrix is equal to 1.

(b) Show that a matrix \( A \) is symplectic if and only if \( A^T J A = J \) for the matrix \( J \) defined in (7.5).

(c) Show that the Jacobian of a canonical transformation (7.10) is symplectic. (Note that this calculation implicitly appears in the proof of Proposition 6.8.)

7.5. (Symplectic and complex structure) Identify Euclidean phase space \( \mathbb{R}^{2n} = \mathbb{R}^n_q \times \mathbb{R}^n_p \) with the complex space \( \mathbb{C}^n \) via \( z_j := q_j + ip_j \).

(a) Show that matrix multiplication by matrix \( J \) (7.5) on \( \mathbb{R}^{2n} \) corresponds to multiplication by \( -i \) on \( \mathbb{C}^n \). This is why the matrix (7.5) is called \( J \), and in fact some authors (e.g. [Arn89]) choose the opposite sign for the canonical symplectic form (7.1) and use the notation \( I \) in place of \( J \).

(b) Show that the Hermitian inner product \( (z, w) \mapsto \sum z_j \overline{w}_j \) on \( \mathbb{C}^n \) corresponds to \( (\xi, \eta) \mapsto \xi \cdot \eta + i \xi \cdot J \eta \) on \( \mathbb{R}^{2n} \). In other words, the scalar product and symplectic product are the real and imaginary parts of the Hermitian inner product, respectively.

7.6. (Hamiltonian PDE [Gar71]) The Hamiltonian mechanics we have developed thus far originates from the time derivative of trajectories in phase space and hence provides a class of ODEs on a finite dimensional manifold; however, this can be extended to PDE when we consider the time derivative of trajectories in an infinite dimensional function space. Given a smooth functional \( F \) of, say, smooth real-valued periodic functions \( C^\infty(\mathbb{R}/\mathbb{Z}) \) we denote the Fréchet derivative kernel by \( \delta F/\delta q(x) \), so that

\[
\frac{dF}{dq}(f) = \left. \frac{d}{ds} \right|_{s=0} F(q + sf) = \int \frac{\delta F}{\delta q}(x)f(x)\,dx.
\]

The Fréchet space \( C^\infty(\mathbb{R}/\mathbb{Z}; \mathbb{R}) \) will now take the role of the manifold \( M \). Specifically, for two smooth functionals \( F(q) \) and \( G(q) \) on \( C^\infty(\mathbb{R}/\mathbb{Z}) \) we define the Poisson bracket

\[
\{F, G\} = \int \left( \frac{\delta F}{\delta q} \right)'(x) \frac{\delta G}{\delta q}(x)\,dx.
\]

In other words, we have replaced the dot product on the Euclidean phase space \( \mathbb{R}^{2n} \) and the matrix \( J \) (7.5) with the \( L^2 \) real inner product and the skew-symmetric operator \( J = \frac{\partial}{\partial x} \). Given a smooth functional \( H(q) \) on \( C^\infty(\mathbb{R}/\mathbb{Z}) \), we define the associated Hamiltonian PDE

\[
\frac{\partial q}{\partial t} = \frac{\partial}{\partial x} \left( \frac{\partial H}{\partial q} \right)
\]

in the spirit of (7.3) and (7.12).

It turns out that many of the results of finite dimensional Hamiltonian mechanics have analogs in this infinite dimensional setting. For example, the global
minimum of a classical Hamiltonian is a (Liapunov) stable equilibrium, and the
global minimum of a Hamiltonian functional is orbitally stable (i.e. concentration compactness).

(a) Check that this Poisson bracket is bilinear, antisymmetric, and satisfies
the Jacobi identity.

(b) Given a Hamiltonian functional \( H(q) \), show that a smooth functional \( F(q) \)
is constant for solutions \( q \) to the PDE associated to \( H \) if and only if
\( \{ H, F \} = 0 \) as a functional on \( C^\infty(\mathbb{R}/\mathbb{Z}) \).

(c) Show that for any Hamiltonian functional \( H(q) \) both the Hamiltonian \( H \)
and the mass functional
\[
M(q) := \int_0^1 q(x) \, dx
\]
are automatically conserved for solutions \( q \) to the PDE associated to \( H \).

(d) Show that the momentum functional
\[
P(q) := \frac{1}{2} \int_0^1 q(x)^2 \, dx
\]
does indeed generate translations, in the sense that the solution to the
associated PDE with initial data \( q(0, x) = f(x) \) is \( q(t, x) = f(x + t) \).

(e) The Korteweg–de Vries (KdV) equation is the PDE associated to the
Hamiltonian
\[
H_{KdV}(q) := \int_0^1 \left( \frac{1}{2} q'(x)^2 - q(x)^3 \right) \, dx,
\]
and arises as the long-wavelength and shallow-water limit for unidirectional water waves of height \( q \) from the undisturbed water level \( q = 0 \).
Show that the mass \( M(q) \), momentum \( P(q) \), and energy \( H_{KdV}(q) \) are all constant for solutions \( q \) to KdV. In fact, these are the first three of an
infinite hierarchy of conserved quantities for KdV.
Chapter 8
Contact geometry

Just as symplectic geometry extends the structure of conservative Hamiltonian
dynamics on phase space, contact geometry is the natural generalization of
nonconservative dynamics on the product of phase space with the time axis.
The material for this chapter is based on [BCT17] and [Lee13, Ch. 22].

8.1 Contact structure

A contact manifold is a smooth $(2n+1)$-dimensional manifold $M$ paired with a
contact form $\eta$. A contact form $\eta$ is a one-form required to satisfy the following
nondegeneracy condition: for each $x \in M$ the restriction of $d\eta_x$ to the subspace
$\ker(\eta_x) \subset T_x M$ is nondegenerate (i.e. $d\eta_x$ is a symplectic tensor for all $x \in M$).
The rank-$2n$ distribution $N \subset TM$ satisfying $N_x = \ker(\eta_x)$ for each $x \in M$ is
called a contact structure on $M$; it plays a fundamental role and is sometimes
taken in the literature to be the defining geometric concept instead of $\eta$.

Similar to how the symplectic form nondegeneracy condition is equivalent to
the nonvanishing of the $n$-fold wedge product $(\omega)^n$, this nondegeneracy condition
turns out to be equivalent to

$$\eta \wedge (d\eta)^n \neq 0; \quad (8.1)$$

the proof of this equivalence is Exercise 8.1. Consequently, $\eta \wedge (d\eta)^n$ defines a
volume form on $M$, and so in particular $M$ must be orientable. Although the
condition (8.1) is sometimes easily verifiable in practice, we will conceptually
be relying on the first condition.

Example 8.1. On the Euclidean contact space $\mathbb{R}S \times T^*\mathbb{R}^n = \mathbb{R}S \times \mathbb{R}_q^n \times \mathbb{R}_p^n$ we have the canonical contact form

$$\eta = dS - pdq = dS - \sum_{i=1}^n p^i \, dq^i. \quad (8.2)$$
CHAPTER 8. CONTACT GEOMETRY

Note that this is the combination of $dS$ and the tautological one-form \textup{(7.2)} on $T^*\mathbb{R}^n$. A straightforward computation shows that $\eta \wedge (d\eta)^n = dS \wedge dq \wedge dp$ is the Euclidean volume form on $\mathbb{R}_t \times T^*\mathbb{R}^n$, but we can also check that $d\eta$ is a symplectic tensor. Note that $d\eta = -dp \wedge dq$, and so the rank-2 distribution $N \subset T\mathbb{R}^{2n+1}$ annihilated by $\eta$ is spanned by the vector fields

\begin{align*}
X_i &= \frac{\partial}{\partial q^i} + p^i \frac{\partial}{\partial S}, \\
Y_i &= \frac{\partial}{\partial p^i}
\end{align*}

for $i = 1, \ldots, n$. Moreover, we have

\begin{align*}
&d\eta(X_i, X_j) = 0, \quad d\eta(Y_i, Y_j) = 0, \quad d\eta(X_i, Y_j) = \delta_{ij}
\end{align*}

for $i, j = 1, \ldots, n$, and it follows that $d\eta|_N$ is nondegenerate as in Example \textup{7.1}.

This example is again fundamental in the following sense.

**Theorem 8.2** (Contact Darboux theorem). If $(M, \eta)$ is a $(2n+1)$-dimensional contact manifold, then for each $x \in M$ there exist local coordinates $(q, p, S)$ centered at $x$ with respect to which $\eta$ has the representation \textup{(8.2)}.

For a proof see [Lee13, Th. 22.31].

The contact structure automatically induces an associated vector field called the Reeb field, which heuristically points orthogonally to the distribution $N$ and plays the role of $S$-axis.

**Proposition 8.3** (The Reeb field). If $(M, \eta)$ is a contact manifold, then there exists a unique smooth vector field $\xi$ on $M$ called the Reeb field satisfying

\begin{align*}
\xi \lrcorner d\eta = 0, \quad \eta(\xi) = 1.
\end{align*}

**Proof.** The map $\Phi$ which takes $X \mapsto X \lrcorner d\eta$ defines a smooth bundle homomorphism $\Phi : TM \to T^*M$, and for each $x \in M$ it reduces to a linear map $\Phi_x : T_xM \to T^*_xM$. Since $d\eta_x$ restricted to the subspace $N_x$ is nondegenerate by definition, then $\Phi_x|_{N_x}$ is injective and hence $\Phi_x$ has rank at least $2n$. On the other hand, we know that $\Phi_x$ cannot have rank $2n + 1$ because then $d\eta_x$ would be nondegenerate and contradict that $T_xM$ is odd-dimensional. Therefore, we conclude that $\ker \Phi_x$ is one-dimensional. Moreover, since $\ker(\Phi_x)$ is not contained in $N_x = \ker(\eta_x)$ by definition, we know there exists a unique $\xi \in \ker(\Phi_x)$ with $\eta_x(\xi_x) = 1$; these correspond to the two conditions \textup{(8.3)} respectively.

The smoothness of $\xi$ follows from the smoothness of $\eta$. Note that $\ker \Phi \subset TM$ is a smooth rank-one subbundle, and so around any $x \in M$ we can pick a smooth nonvanishing section $X$ of $\ker \Phi$ near $x$. Since $\eta(X) \neq 0$, then we can write $\xi = \eta(X)^{-1} X$ as a composition of smooth maps near $x$.

**Example 8.4.** For the Euclidean contact space of Example \textup{8.1} we see that Reeb field is

\begin{align*}
\xi = \frac{\partial}{\partial S}
\end{align*}

as the two conditions \textup{(8.3)} are easily verified.
8.2 Hamiltonian vector fields

Given a smooth function \( H \) on a contact manifold \((M, \eta)\), the associated contact Hamiltonian vector field \( X_H \) is uniquely determined by the two conditions

\[
\eta(X_H) = H, \quad (X_H \cdot d\eta)|_N = -dH|_N. \tag{8.4}
\]

Since \( d\eta|_N \) is nondegenerate by definition, there is a unique vector field \( Y \) on \( N \) satisfying the second condition of (8.4). The vector field \( X_H := Y + H\xi \) is then the unique solution of the conditions (8.4).

In comparison to symplectic Hamiltonian vector fields, the first condition of (8.4) looks like the primitive of \( \omega(X_H) = dH \) (which we wrote as \( X_H = JdH \)) and the second condition is like \( X_H \cdot \omega = dH \); in the symplectic case these conditions were redundant (cf. (7.4)), but now we need both in order to determine \( X_H \) on and off of the kernel \( N_x \) of \( \eta_x \).

**Example 8.5.** Let us see what the contact Hamiltonian vector field \( X_H \) looks like for the Euclidean contact space of Example 8.1 (and hence also the expression of \( X_H \) in the local coordinates guaranteed by Theorem 8.2). Given a smooth function \( H(s, q, p) \) it is easily verified that

\[
X_H = \sum_{i=1}^{n} \left( p^i \frac{\partial H}{\partial p^i} - H \right) \frac{\partial}{\partial S} + \frac{\partial H}{\partial p^i} \frac{\partial}{\partial q^i} - \left( \frac{\partial H}{\partial q^i} + p^i \frac{\partial H}{\partial S} \right) \frac{\partial}{\partial p^i} \tag{8.5}
\]

satisfies the two conditions (8.4), from which we obtain the differential equation system

\[
\frac{dS}{dt} = \sum_{i=1}^{n} p^i \frac{\partial H}{\partial p^i} - H, \quad \frac{dq^i}{dt} = \frac{\partial H}{\partial p^i}, \quad \frac{dp^i}{dt} = -\frac{\partial H}{\partial q^i} - p^i \frac{\partial H}{\partial S}. \tag{8.6}
\]

When \( H \) is independent of \( s \), the second two sets of equations reduce to Hamilton’s equations of motion (6.4). We also recognize the quantity \( S \) as the action \( S(q, t) \); indeed, if we use the formula (5.1) for the momentum and use the Hamilton–Jacobi equation (5.4) we see that the time derivative of the action is

\[
\frac{dS}{dt} = \sum_{i=1}^{n} \frac{\partial S}{\partial q^i} \frac{dq^i}{dt} + \frac{\partial S}{\partial t} = \sum_{i=1}^{n} p^i \frac{dq^i}{dt} - H.
\]

To conclude the section, let us consider a one-dimensional physical example for to illustrate how contact geometry encapsulates time-independent nonconservative dynamics.

**Example 8.6.** Take \( n = 1 \) in the Euclidean contact space of Example 8.1 and consider the Hamiltonian

\[
H(q, p, S) = \frac{p^2}{2m} + U(q) + \gamma S \tag{8.7}
\]
where $\gamma$ is a real constant. The contact Hamilton’s equations (8.6) read

$$
\dot{q} = \frac{p}{m}, \quad \dot{p} = -U'(q) - \gamma p, \quad \dot{S} = \frac{p^2}{2m} - U(q) - \gamma S.
$$

This represents a Newtonian system with a friction force that depends linearly on the velocity, like the damped harmonic oscillator of Example 7.11. Note that, as opposed to Example 7.11, in this approach the momentum coordinate still coincides with the physical momentum defined via the velocity.

### 8.3 Dynamics

Using the expression (8.5) of a contact Hamiltonian vector field $X_H$ in terms of the local coordinates $(q,p,S)$ of Theorem 8.2, we see that a smooth function $F(q,p,S)$ on a contact manifold $(M,\eta)$ evolves according to

$$
\frac{dF}{dt} = X_H(F) = -H\frac{\partial F}{\partial S} + \sum_{i=1}^{n} p^i \{H,F\}_{p^i,S} + \{H,F\}_{p,q}. \quad (8.8)
$$

The last term above we recognize from (6.18) and the first two terms are corrections introduced by the contact structure. The notation $\{H,F\}_{p^i,S}$ should be interpreted as a convenient shorthand defined via the above formula with no deeper meaning.

Taking $F = H(q,p,S)$ to be the Hamiltonian for the vector field $X_H$ in the formula (8.8) above, we see that the Hamiltonian evolves according to

$$
\frac{dH}{dt} = -H\frac{\partial H}{\partial S}. \quad (8.9)
$$

From this we see that $H$ is a conserved quantity if and only if $H$ is independent of $S$ or $H = 0$. In particular, for a conservative Hamiltonian $H = H(q,p)$ we recover the conservation of the Hamiltonian. However, in general the rate of decrease of the Hamiltonian is proportional to the system’s energy $H$ and its dissipation $\partial H/\partial S$.

Specifically, let us consider a Hamiltonian of the form

$$
H(q,p,S) = H_0(q,p) + f(S) \quad (8.10)
$$

where $H_0(q,p)$ is the mechanical energy of the system (e.g. $H_0(q,p) = p^2/2m + U(q)$), then according to the formula (8.8) the mechanical energy obeys

$$
\frac{dH_0}{dt} = -\sum_{i=1}^{n} p^i \frac{\partial H_0}{\partial p^i} f'(S). \quad (8.11)
$$
8.4. CONTACT TRANSFORMATIONS

We interpret this as saying that \( f(S) \) is the potential for the system’s dissipative force. Moreover, the evolution \([8.9]\) of the Hamiltonian can be integrated to obtain

\[
H(t) = H(0) \exp \left[ - \int_0^t f'(S) \right].
\]

Plugging in the Hamiltonian \([8.10]\), we obtain an implicit equation which in principle determines the action \( S = S(q, p, t) \); thus the equations of motion \([8.6]\) reduce to the \( 2n \) equations for the positions \( q^i \) and momenta \( p^i \).

**Example 8.7.** If we take \( d = 1 \) and the dissipative potential \( f(S) = \gamma S \) to be linear as in Example \([8.6]\) then equation \([8.11]\) becomes

\[
\dot{H}_0 = -m\gamma \dot{q}^2
\]

which agrees with what we found for the damped harmonic oscillator of Example \([1.13]\). The energy of this system decays exponentially to zero, according to

\[
H(t) = H(0) e^{-\gamma t}.
\]

Solving for the action \( S \) we obtain

\[
S(q, p, t) = \frac{1}{\gamma} \left[ H(0) e^{-\gamma t} - \frac{p^2}{2m} - U(q) \right].
\]

### 8.4 Contact transformations

A contact transformation is a smooth transformation which leaves the contact structure invariant. As opposed to canonical transformations, we allow the contact form \( f\eta \) in the new coordinates to differ by a smooth nonvanishing factor \( f \in \mathcal{C}^\infty(M) \) since we are interested in the contact structure \( N \). In terms of the canonical coordinates \((q, p, S)\) guaranteed by Theorem \([8.2]\), the new coordinates \((\tilde{q}, \tilde{p}, \tilde{S})\) must satisfy

\[
\tilde{\eta} = d\tilde{S} - \tilde{p} \, d\tilde{q} = f(dS - p \, dq) = f\eta.
\]

Writing \((\tilde{S}, \tilde{q}, \tilde{p})\) as functions of \((q, p, S)\), this is equivalent to the conditions

\[
f = \frac{\partial \tilde{S}}{\partial S} - \sum_{j=1}^n \tilde{p}^j \frac{\partial \tilde{q}^j}{\partial S}, \quad -fp^i = \frac{\partial \tilde{S}}{\partial q^i} - \sum_{j=1}^n \tilde{p}^j \frac{\partial \tilde{q}^j}{\partial q^i}, \quad 0 = \frac{\partial \tilde{S}}{\partial p^i} - \sum_{j=1}^n \tilde{p}^j \frac{\partial \tilde{q}^j}{\partial p^i}
\]

for \( i = 1, \ldots, d \). Note that canonical transformations are independent of \( S, \tilde{S} \) and are defined by the condition \([8.12]\) with \( f \equiv 1 \).

Since we allow for a conformal factor \( f \) in the definition \([8.12]\), the volume form is also rescaled. Indeed, if \( \tilde{\eta} = f\eta \) then \( d\tilde{\eta} = df \wedge \eta + f \, d\eta \), and so

\[
\tilde{\eta} \wedge (d\tilde{\eta})^n = f^{d+1} \eta \wedge (d\eta)^n
\]
is the new volume form. In the case of canonical transformations we have $f \equiv 1$, and hence we recover volume preservation.

As with canonical transformations in section 6.5 we may consider a contact transformation as being generated by a generating function. For an example, assume that the coordinates $(q, \tilde{q}, S)$ are independent. Then the differential of the generating function $\tilde{S} = \tilde{S}(q, \tilde{q}, S)$ may be written

$$d\tilde{S} = \sum_{i=1}^{n} \frac{\partial \tilde{S}}{\partial q^i} dq^i + \sum_{i=1}^{n} \frac{\partial \tilde{S}}{\partial \tilde{q}^i} d\tilde{q}^i.$$  

Plugging this into (8.12), we see that the remaining coordinates are determined in terms of $\tilde{S}$ by

$$f = \frac{\partial \tilde{S}}{\partial S}, \quad f p^i = -\frac{\partial \tilde{S}}{\partial q^i}, \quad \tilde{p}^i = \frac{\partial \tilde{S}}{\partial \tilde{q}^i}.$$  

Taking $f \equiv 1$ and comparing to eq. (6.25), we see that $\tilde{S}$ is related to the canonical transformation generating function $F(q, \tilde{q})$ via

$$\tilde{S} = S - F(q, \tilde{q}).$$  

In particular, we conclude that all physicist’s time-independent canonical transformations (cf. the remark of section 6.5) are a special case of contact transformations.

Recall that for symplectic structures the Hamiltonian dynamics generate instantaneous canonical transformations; this was first seen in section 6.6 computationally, and then abstractly as an equivalence in Proposition 7.12. In analogy with symplectic vector fields, we will call a smooth vector field $V$ on $M$ if a contact vector field if its flow $e^{tV}$ preserves the contact structure $N$,

$$d(e^{tV})_x(N_x) = N_{e^{tV}x}, \quad (8.13)$$

for all $t \in \mathbb{R}$ and $x \in M$ in the domain of definition of the flow $e^{tV}$.

**Proposition 8.8.** Let $(M, \eta)$ be a contact manifold. A smooth vector field $V$ on $M$ is a contact vector field if and only if it is a contact Hamiltonian vector field.

**Proof.** From the condition (8.13) and the definition of the Lie derivative, we see that $V$ is a contact vector field if and only if $\mathcal{L}_V \eta = 0$ on $N$.

First assume that $V$ is a contact Hamiltonian vector field, and write $V = X_H$ for a Hamiltonian $H$. Then from Cartan’s magic formula and the first condition (8.4) defining $X_H$ we have

$$\mathcal{L}_{X_H} \eta = d[\eta(X_H)] + X_H \lrcorner \, d\eta = dH + X_H \lrcorner \, d\eta.$$  

From the second condition of (8.4) we know that $X_H \lrcorner \, d\eta$ is equal to $-dH$ on $N$. By the definition (8.3) of the Reeb field it then follows that

$$\mathcal{L}_{X_H} \eta = -\xi(H)\eta = -\frac{\partial H}{\partial S} \eta, \quad (8.14)$$
where the last equality is merely the expression in terms of the local canonical coordinates. Comparing this to the definition (8.12) of a contact transformation, we see that the flow is a contact transformation with \( f = -\xi(H) = -\partial H/\partial S \).

In particular, if we restrict to the contact structure \( \mathcal{N} \), we have \( \eta = 0 \) and hence \( \mathcal{L}_{X_H}\eta = 0 \) as desired.

Conversely, assume that \( V \) is contact vector field. Then Cartan’s magic formula reads

\[
0 = (\mathcal{L}_V\eta)|_N = d\eta(X_H)|_N + (V \cdot d\eta)|_N.
\]

Consider the smooth function \( H = \eta(V) \), defined so that the first condition of the definition (8.4) holds. Then we obtain the second condition of (8.4) from the above equality, and so we conclude that \( V = X_H \) is the contact Hamiltonian vector field for \( H \).

Heuristically, we do not expect the volume form to be preserved by a general contact Hamiltonian flow since contact dynamics includes dissipative systems. Using eq. (8.14) we see that the volume form evolves according to

\[
\mathcal{L}_{X_H}[\eta \wedge (d\eta)^n] = (\mathcal{L}_{X_H}\eta) \wedge (d\eta)^n + \sum_{i=0}^{n-1} \eta \wedge (d\eta)^i \wedge [d(\mathcal{L}_{X_H}\eta)] \wedge (d\eta)^{n-1-i}
\]

\[
= -(n+1) \frac{\partial H}{\partial S} \eta \wedge (d\eta)^n.
\]

This illustrates the connection between the Hamiltonian’s \( S \)-dependence to the system’s dissipation, and consequently systems for which \( \frac{\partial H}{\partial S} \) is nonvanishing are called dissipative.

Instead, we have a variant of Liouville’s theorem due to [BT], in which a rescaled volume form is preserved away from the zero set \( H^{-1}(0) \).

**Proposition 8.9** (Canonical measure for dissipative contact systems). Let \((M,\eta)\) be a \((2n+1)\)-dimensional contact manifold and \( H \) a smooth function on \( M \). Then the volume form

\[
|H|^{-(n+1)} \eta \wedge (d\eta)^n
\]

is an invariant measure for the contact Hamiltonian flow for \( H \) along orbits outside of \( H^{-1}(0) \). Moreover, up to scalar multiplication it is the unique such measure whose density with respect to the standard volume form depends only on \( H \).

**Proof.** For a smooth function \( \rho \) on \( M \), a computation using (8.14) shows that

\[
\mathcal{L}_{X_H}[\rho \eta \wedge (d\eta)^n] = (\mathcal{L}_{X_H}\rho) \eta \wedge (d\eta)^n - (n+1) \frac{\partial H}{\partial S} \rho \eta \wedge (d\eta)^n
\]

\[
= \left[ X_H(\rho) - (n+1) \frac{\partial H}{\partial S} \rho \right] \eta \wedge (d\eta)^n.
\]
If we assume \( \rho = \rho(H) \) then
\[
X_H(\rho) = -H\rho'(H)\frac{\partial H}{\partial S},
\]
and so the vanishing of \( \mathcal{L}_{X_H}[\rho \eta \wedge (d\eta)^n] \) occurs exactly when \( \rho \) solves
\[
\rho'(H) = -(n + 1)H^{-1}\rho.
\]
This equation has the solution \( \rho(H) = |H|^{-(n+1)} \) and it is unique up to scalar multiplication.

\section{8.5 Time-dependent systems}

Thus far we have allowed \( H \) to be a function on the contact manifold \( M \), and hence have only considered time-independent dissipative systems. In this section, we present the extension introduced in \cite{BCT17} of contact Hamiltonian systems to include time-dependence.

For \((M, \eta)\) a \((2n + 1)\)-dimensional contact manifold, we define the \textit{extended manifold} \( M \times \mathbb{R} \). In analogy with the Poincaré–Cartan one-form (7.15), given a possibly time-dependent Hamiltonian \( H \) we extend the contact form to
\[
\theta = \,dS - p\,dq + H\,dt = dS - \sum_{i=1}^n p_i\,dq_i + H\,dt \tag{8.15}
\]
in terms of the canonical coordinates \((q, p, S)\) on \( M \) guaranteed by Theorem 8.2.

On \( M \times \mathbb{R} \) we define the \textit{extended contact Hamiltonian vector field}
\[
Y_H = X_H + \frac{\partial}{\partial t}.
\]
In place of the conditions (8.4), it can be checked that this vector field is uniquely determined by
\[
\theta(Y_H) = 0, \quad Y_H \,d\theta = -\frac{\partial H}{\partial S}\theta. \tag{8.16}
\]
Here, the first condition is analogous to how (7.17) replaced (7.4) for time-dependent symplectic systems, and the second condition is the analog of (8.14) (which serves as a rephrasing for the second condition of (8.4) that does not involve \( N \)).

The flow of \( Y_H \) is given by
\[
\frac{dS}{dt} = \sum_{i=1}^n p_i \frac{\partial H}{\partial p^i} - H, \quad \frac{dq^i}{dt} = \frac{\partial H}{\partial p^i}, \quad \frac{dp^i}{dt} = -\frac{\partial H}{\partial q^i} - p^i \frac{\partial H}{\partial S}, \quad i = 1,
\]
which are the old equations of motion (8.6) joined with the trivial equation \( \dot{t} = 1 \). It follows that any smooth time-dependent function \( F \) on \( M \times \mathbb{R} \) evolves according to
\[
\frac{dF}{dt} = -H \frac{\partial F}{\partial S} + \sum_{i=1}^n p^i \{H, F\}_{p^i,S} + \{H, F\}_{p,q} + \frac{\partial F}{\partial t}. \tag{8.17}
\]
8.5. TIME-DEPENDENT SYSTEMS

using the notation of eq. 8.8. In particular, we see that under its own flow
the Hamiltonian now changes according to both its dissipation \( \partial H/\partial S \) and its
time-dependence.

Lastly, let us extend the notion of contact transformations to our extended
manifold \( M \times \mathbb{R} \). In terms of canonical coordinates, a time-dependent contact
transformation \((q, p, S, t) \mapsto (\tilde{q}, \tilde{p}, \tilde{S}, \tilde{t})\) must satisfy

\[
\tilde{\theta} = d\tilde{S} - \tilde{p} d\tilde{q} + K \ dt = f(dS - p dq + H dt) = f \theta
\] (8.18)

for a smooth nonvanishing factor \( f \in C^\infty(M \times \mathbb{R}) \) and a new Hamiltonian
\( K \in C^\infty(M \times \mathbb{R}) \). Expanding \( d\tilde{q} \) and \( d\tilde{S} \), we see that the new Hamiltonian
must satisfy

\[
fH = \frac{\partial \tilde{S}}{\partial t} - \sum_{i=1}^{n} \tilde{p}_i \frac{\partial \tilde{q}_i}{\partial t} + K.
\] (8.19)

As before, we may consider a contact transformation as being generated
by a generating function. For example, let us assume that the coordinates
\((q, \tilde{q}, S, t)\) are independent. After substituting the differential of the generating
function \( \tilde{S} = \tilde{S}(q, \tilde{q}, S, t) \) into (8.18), we see that the remaining coordinates are
determined in terms of \( \tilde{S} \) by

\[
f = \frac{\partial \tilde{S}}{\partial S}, \quad fp^i = -\frac{\partial \tilde{S}}{\partial \tilde{q}^i}, \quad \tilde{p}_i = \frac{\partial \tilde{S}}{\partial \tilde{q}_i}, \quad fH = \frac{\partial \tilde{S}}{\partial t} + K.
\] (8.19)

The first three conditions are unchanged and the last condition defines the new
Hamiltonian \( K = K(q, \tilde{q}, S, t) \). Taking \( f \equiv 1 \) and comparing to eq. (6.25), we see
that \( \tilde{S} \) is related to the canonical transformation generating function \( F(q, \tilde{q}, t) \)
via

\[
\tilde{S} = S - F(q, \tilde{q}, t).
\]

However, now there is an additional constraint on \( \tilde{S} \) imposed by the invariance
of the second condition of (8.16). After the transformation we must have

\[
Y_H \cdot d\theta = -\frac{\partial K}{\partial \tilde{S}} \tilde{\theta}.
\]

Using \( \tilde{\theta} = f \theta, \ Y_H = Y_H \), and the extended contact Hamiltonian vector field
conditions [8.16], this yields

\[
f \frac{\partial K}{\partial \tilde{S}} = f \frac{\partial H}{\partial S} + df(Y_H).
\]

In the special case \( f \equiv 1 \) we note that if \( H \) is independent of \( S \) then \( K = 0 \)
is a solution, in which case the last condition of (8.19) becomes the familiar
Hamilton–Jacobi equation (5.4). However, in general \( f \) may be \( S \)-dependent
and so the notion of contact transformations is strictly more general than even the
physicist’s notion of canonical transformations (cf. the remark of section 6.5).
Example 8.10. In Example 7.11 we introduced a non-canonical coordinate transformation \((q,p) \mapsto (q,e^{\gamma t})\) to describe a dissipative system using time-dependent Hamiltonian dynamics. However, it is easily checked that the transformation
\[
\tilde{q} = q, \quad \tilde{p} = e^{\gamma t} p, \quad \tilde{S} = e^{\gamma t} S, \quad \tilde{t} = t
\]
satisfies eq. (8.15) and is hence a time-dependent contact transformation. Here, the conformal factor is \(f = e^{\gamma t}\), and the Hamiltonians are expressed in their respective coordinates: \(H\) is given by eq. (8.7) and \(K\) is given by eq. (7.19).

8.6 Exercises

8.1. Show that a smooth one-form \(\eta\) on a \((2n+1)\)-dimensional manifold \(M\) satisfies the nondegeneracy condition of a contact form if and only if it satisfies the nonvanishing top form condition (8.1).

8.2 (Contact structure on \(S^{2n+1}\)). On the Euclidean space \(\mathbb{R}^{2n+2}\) consider the coordinates \((x^1,\ldots,x^{n+1},y^1,\ldots,y^{n+1})\) and define the one-form
\[
\theta := \sum_{i=1}^{n+1} (x^i \, dy^i - y^i \, dx^i).
\]
The standard contact form on the sphere \(S^{2n+1}\) is \(\eta := \iota^* \theta\), where \(\iota := S^{2n+1} \hookrightarrow \mathbb{R}^{2n+2}\) is the inclusion map.

(a) Show that the vector fields
\[
V = \sum_{i=1}^{n+1} \left( x^i \frac{\partial}{\partial x^i} + y^i \frac{\partial}{\partial y^i} \right), \quad W = \sum_{i=1}^{n+1} \left( x^i \frac{\partial}{\partial y^i} - y^i \frac{\partial}{\partial x^i} \right)
\]
satisfy \(V \cdot d\theta = 2\theta\) and \(W \cdot d\theta = -d(x^2 + y^2)\).

(b) Let \(S \subset T(\mathbb{R}^{2n+1} \setminus \{0\})\) denote the subbundle spanned by \(V\) and \(W\), and let
\[
S^\perp = \bigcup_{p \in S^{2n+1}} \{ X \in T_p \mathbb{R}^{2n+2} : d\theta_p(V_p,X_p) = d\theta_p(W_p,X_p) = 0 \}
\]
denote its symplectic complement. Show that \(\theta\) is indeed a contact form with respect to the contact structure \(S^\perp\).

(c) Show that the corresponding Reeb field is given by \(W\) restricted \(S^{2n+1}\).

8.3. (Solving the damped parametric oscillator via expanding coordinates) Consider the one-dimensional damped parametric oscillator
\[
H = \frac{p^2}{2m} + \frac{m}{2} \omega^2(t)q^2 + \gamma S
\]
with time-dependent frequency $\omega(t)$ and damping parameter $\gamma$. Show that the new expanding coordinates
\[
\tilde{q} = e^{\gamma t / 2} q, \quad \tilde{p} = e^{\gamma t / 2} (p + \frac{1}{2} m \gamma q), \quad \tilde{S} = e^{\gamma t} (S + \frac{1}{2} m \gamma q^2), \quad \tilde{t} = t
\]
define a contact transformation, with respect to which the Hamiltonian takes the form
\[
K = \frac{\tilde{p}^2}{2m} + \frac{m}{2} \left( \omega^2(t) - \frac{1}{4} \gamma^2 \right) \tilde{q}^2.
\]
This new Hamiltonian $K$ now corresponds to an undamped parametric oscillator with the new frequency $\sqrt{\omega^2(t) - \frac{1}{4} \gamma^2}$. The undamped oscillator has been extensively studied and solutions for the equations of motion can be obtained. In the harmonic oscillator case $\omega(t) \equiv \omega_0$ the Hamiltonian $K$ is a conserved quantity, and the coordinates expand exponentially in time so that the trajectories form closed orbits at a slower frequency.

8.4. (Solving the damped parametric oscillator via conserved quantities) Consider the one-dimensional damped parametric oscillator
\[
H = \frac{p^2}{2m} + \frac{m}{2} \omega^2(t)q^2 + \gamma S
\]
with time-dependent frequency $\omega(t)$ and damping parameter $\gamma$.

(a) We seek a solution $F$ to eq. (8.17) with vanishing left-hand side. Substituting the quadratic ansatz
\[
F(q, p, S, t) := \beta(t)p^2 - 2\xi(t)qp + \eta(t)q^2 + \zeta(t)S,
\]
obtain a system of first-order equations for $\beta$, $\eta$, $\xi$, and $\zeta$, where the $\dot{\zeta}$ has solution $\zeta(t) = \zeta_0 e^{\gamma t}$.

(b) Using the substitution $\beta(t) := \frac{1}{2m} e^{\gamma t} \alpha^2(t)$, show that the $\beta$ equation is solved if and only if $\alpha$ satisfies the Ermakov equation
\[
\dot{\alpha} + [\omega^2(t) - \frac{1}{4} \gamma^2] \alpha = \alpha^{-3},
\]
and the remaining equations become
\[
\eta(t) = \frac{1}{2} m e^{\gamma t} \left\{ [\dot{\alpha} - \frac{1}{2} \gamma \alpha]^2 + \alpha^{-2} \right\}, \quad \xi(t) = \frac{1}{2} e^{\gamma t}[\dot{\alpha} - \frac{1}{2} \gamma \alpha] + \frac{1}{4}.
\]

(c) Conclude that the quantity $F(q, p, S, t) = I(q, p, t) + \zeta_0 G(q, p, S, t)$ is conserved, with
\[
I = \frac{m}{2} e^{\gamma t} \left\{ \left[ \frac{p}{m} \alpha - \left( \frac{1}{2} \alpha - \frac{1}{2} \gamma \alpha \right) q \right]^2 + \left( \frac{q}{\alpha} \right)^2 \right\}, \quad G = e^{\gamma t} (S - \frac{1}{2} qp),
\]
and $\alpha(t)$ solves the Ermakov equation. Moreover, since $F$ is invariant for all initial conditions and $\zeta_0$ is determined solely by initial conditions, then $I$ and $G$ must be separately conserved.
(d) Show that the new coordinates

$$\tilde{q} = \arctan \left( \frac{\dot{\alpha} - \frac{\gamma}{2} \alpha - \alpha^2 p}{mq} \right),$$

$$\tilde{p} = I(q, p, t),$$

$$\tilde{S} = G(q, p, S, t),$$

(and $\tilde{t} = t$) define a contact transformation, with respect to which the new Hamiltonian is simply $K = I(\alpha)^{-2}$. Solve the new equations of motion for $\tilde{q}, \tilde{p}, \tilde{S}$ and obtain the solution

$$q(t) = \sqrt{\frac{2I}{m}} e^{\gamma t} \alpha(t) \cos \phi(t),$$

$$S(t) = Ge^{-\gamma t} + \frac{1}{2} q(t)p(t),$$

$$p(t) = \sqrt{2mIe^{\gamma t}} \left( \frac{\dot{\alpha} - \frac{\gamma}{2} \alpha}{\alpha} \right) \cos \phi(t) - \frac{1}{\alpha} \sin \phi(t),$$

$$\phi(t) = \int_{t_0}^{t} \frac{d\tau}{\alpha^2(\tau)}.$$

Here, $\alpha(t)$ solves the Ermakov equation and the conserved quantities $I$ and $G$ are determined by the initial conditions.
Bibliography


[BT] Alessandro Bravetti and Diego Tapias. Liouville’s theorem and the canonical measure for nonconservative systems from contact geometry.


