Classical Mechanics

Babis Anastasiou[a], Vittorio del Duca[b]
Institute for Theoretical Physics, ETH Zurich,
8093 Zurich, Switzerland
[a]E-mail: babis@ethz.ch
[b]E-mail: Vittorio.Del.Duca@cern.ch
Typesetting: Patrik Weber

December 17, 2015

Abstract

The subject of the course is classical mechanics. The following topics are discussed:

- Galileian transformations and Newtonian mechanics
- Variational methods
- Principle of least action
- Lagrangian mechanics
- Symmetries and conservation laws
- Two body systems
- Oscillations
- Rigid body dynamics
- Hamiltonian mechanics
- Hamilton-Jacobi equation
- Special Relativity

References

[1] Landau and Lifshitz, Mechanics, Course of Theoretical Physics Vol. 1., Pergamon Press

[2] Classical Mechanics, 3rd Edition, Goldstein, Poole and Safko, Addison Wesley

[3] F. Gantmacher, Lectures in Analytical Mechanics, Mir Publications.

Contents

1 Principle of relativity and Galileian transformations

We will assume the existence of frames of reference (a system of coordinates) in which the motion of a free particle (on which no force is applied) is uniform, i.e. it moves with a constant velocity.

$$\text{free particle:} \quad \vec{v} = \text{constant}. \tag{1.1}$$

Such reference frames are called inertial.

The *principle of relativity* states that **the laws of physics take the exact same form in all inertial frames.**

Let us assume that we know the trajectory (position as a function of time) of all the particles which constitute a physical system as they are observed in one inertial reference frame S. Can we compute the trajectories in a different inertial frame S'? To achieve that, we need to transform the space and time coordinates $(\vec{r}, t)$ in the frame S to the corresponding coordinates $(\vec{r}\,', t')$ in the frame S'. A good guess for (Cartesian) coordinate transformations are the so called Galilei transformations:

$$\vec{r}\,' = \mathbf{R}\,\vec{r} - \vec{r}_0 - \overrightarrow{V}t, \quad t' = t_0 + t. \tag{1.2}$$

or, equivalently,

$$r_i' = \sum_{j=1}^{3} R_{ij} r_j - r_{0,i} - V_i t. \tag{1.3}$$

where $\overrightarrow{V}$ is the apparent velocity for a point in the frame S' according to an observer in the frame S. $\vec{r}_0$ is the position of the origin of S' as seen in S at a time $t = 0$ and t_0 is the time shown in the clock of S' at $t = 0$. The 3×3 matrix R describes the rotation of the axes of S' with respect to the axes of S and it satisfies:

$$\mathbf{R}^T\mathbf{R} = \mathbf{1}_{3\times 3}\,, \tag{1.4}$$

where $\mathbf{1}_{3\times 3}$ is the 3×3 identity matrix. Equivalently, we write

$$R_{ij}R_{ik} = \delta_{jk}, \tag{1.5}$$

where δ_{ij} is the Kronecker delta symbol. In addition, the determinant of the rotation matrix is:

$$\det\mathbf{R} = 1. \tag{1.6}$$

It turns out that Galileo's transformations are only accurate for small velocities in comparison to the speed of light: $V \ll c$. We will discuss the correct generalisation of Galileo's transformations (the so called Lorentz transformations) towards the end of this course. In the mean time, we will assume that Eq 1.2 is valid; this is an excellent approximation for a plethora of phenomena.

1.1 Time intervals and distances in two inertial frames

Consider two events:

$$\text{Event A: } (t_A, \vec{r}_A),$$

$$\text{Event B: } (t_B, \vec{r}_B),$$

as observed in the reference frame S.

In the reference frame S' these events are described by the time and space coordinates:

$$\text{Event A: } (t'_A, \vec{r}'_A) = (t_A + t_0, \mathbf{R}\vec{r}_A + \vec{r}_0 - \vec{V}t_A),$$

$$\text{Event B: } (t'_B, \vec{r}'_B) = (t_B + t_0, \mathbf{R}\vec{r}_B + \vec{r}_0 - \vec{V}t_B).$$

The time difference of the two events is equal in the two frames S and S':

$$t'_A - t'_B = t_A - t_B. \tag{1.7}$$

Therefore, time intervals are measured to be the same in all inertial frames. Space distances are also the same in all inertial frames. For two events occurring at the same time t, their space distance is

$$\begin{aligned} \Delta r' &= |\vec{r}'_A - \vec{r}'_B| = |\mathbf{R}\,\vec{r}_A - \mathbf{R}\vec{r}_B| = \sqrt{(\vec{r}_A - \vec{r}_B)^T \mathbf{R}^T \mathbf{R} (\vec{r}_A - \vec{r}_B)} \\ &= \sqrt{(\vec{r}_A - \vec{r}_B)^T (\vec{r}_A - \vec{r}_B)} = |\vec{r}_A - \vec{r}_B| = \Delta r. \end{aligned} \tag{1.8}$$

1.2 Vectors, scalars and rotations

The position of a particle $\vec{r}$, under a rotation, transforms as:

$$\vec{r} \to \vec{r}' = \mathbf{R}\,\vec{r}, \quad \det \mathbf{R} = 1, \quad \mathbf{R}^T\mathbf{R} = 1. \tag{1.9}$$

The position is one of many other objects with the same transformation under rotations. We can formally define a vector $\vec{A}$ to be a set of three

numbers which transforms in the same way as the position vector does under a rotation:

$$\vec{A} \to \vec{A}' = \mathbf{R}\,\vec{A}, \quad \det \mathbf{R} = 1, \quad \mathbf{R}^T\mathbf{R} = 1. \tag{1.10}$$

For example, time derivatives of the position vector (velocity, acceleration, ...) are also vectors:

$$\frac{d\vec{r}}{dt} \to \frac{d\vec{r}'}{dt'} = \frac{d(\mathbf{R}\,\vec{r})}{dt} = \mathbf{R}\,\frac{d\vec{r}}{dt} \tag{1.11}$$

We can define a dot-product of two vectors as

$$\vec{A} \cdot \vec{B} = \sum_{i=1}^{3} A_i B_i. \tag{1.12}$$

The dot product is invariant under rotations:

$$\begin{aligned} \vec{A} \cdot \vec{B} \;\; &\to \;\; \vec{A}' \cdot B' = \sum_i A'_i B'_i = \sum_{ijk} (R_{ij}A_j)(R_{ik}B_k) \\ &= \;\; \sum_{ijk} R^T_{ji} R_{ik} A_j B_k = \sum_{jk} \delta_{jk} A_j B_k = \sum_k A_k B_k \\ &= \;\; \vec{A} \cdot \vec{B}. \end{aligned} \tag{1.13}$$

All objects which are invariant under rotation transformations are called *scalars.*

The cross-product of two vectors (such as the angular momentum) is also a vector. Indeed, if $\vec{A}, \vec{B}$ are vectors then their cross-product transforms as:

$$\vec{A} \times \vec{B} \to \vec{A}' \times \vec{B}' = \mathbf{R}(\vec{A} \times \vec{B}). \tag{1.14}$$

To prove the above, consider the dot product

$$(\vec{A} \times \vec{B}) \cdot C = \sum_i \left(\sum_{jk} \epsilon_{ijk} A_j B_k \right) C_i = \det\left(\ \vec{A}\,,\ \ \vec{B}\,,\ \ \vec{C}\ \right)$$

Under a rotation, this dot product is invariant,

$$\begin{aligned} \det\left(\ \vec{A}\,,\ \ \vec{B}\,,\ \ \vec{C}\ \right) \;\; &\to \;\; \det\left(\ \vec{A}'\,,\ \ \vec{B}'\,,\ \ \vec{C}'\ \right) = \det\left(\ \mathbf{R}\,\vec{A}\,,\ \ \mathbf{R}\,\vec{B}\,,\ \ \mathbf{R}\,\vec{C}\ \right) \\ &= \det\left\{\mathbf{R}\ \left(\ \vec{A}\,,\ \ \vec{B}\,,\ \ \vec{C}\ \right)\right\} \\ &= \det(\mathbf{R})\,\det\left(\ \vec{A}\,,\ \ \vec{B}\,,\ \ \vec{C}\ \right) \\ &= \det\left(\ \vec{A}\,,\ \ \vec{B}\,,\ \ \vec{C}\ \right) \end{aligned} \tag{1.15}$$

and it is therefore a scalar:

$$\begin{aligned}
(\vec{A}' \times \vec{B}') \cdot \vec{C}' &= (\vec{A} \times \vec{B}) \cdot \vec{C} \\
\rightsquigarrow (\vec{A}' \times \vec{B}') \cdot (\mathbf{R}\,\vec{C}) &= (\vec{A} \times \vec{B}) \cdot \vec{C} \\
\rightsquigarrow \left[\mathbf{R}^T (\vec{A}' \times \vec{B}')\right] \cdot \vec{C} &= (\vec{A} \times \vec{B}) \cdot \vec{C} \\
\rightsquigarrow (\vec{A}' \times \vec{B}') \cdot \vec{C} &= \left[\mathbf{R}\,(\vec{A} \times \vec{B})\right] \cdot \vec{C}
\end{aligned} \tag{1.16}$$

Since the above is valid for any vector $\vec{C}$, we conclude that

$$\vec{A}' \times \vec{B}' = \mathbf{R}\,(\vec{A} \times \vec{B}). \tag{1.17}$$

Finally, the gradient differential operator

$$\vec{\nabla} \equiv \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}\right) \tag{1.18}$$

is also a vector. To see this, we first note that

$$x'_i = \sum_j R_{ij} x_j \rightsquigarrow \frac{\partial x'_i}{\partial x_j} = R_{ij} \tag{1.19}$$

and

$$x_i = \sum_j R^T_{ij} x'_j \rightsquigarrow \frac{\partial x_i}{\partial x'_j} = R^T_{ij} = R_{ji}. \tag{1.20}$$

Then

$$\vec{\nabla} \to \vec{\nabla}' = \frac{\partial}{\partial x'_i} = \sum_j \frac{\partial x_j}{\partial x'_i} \frac{\partial}{\partial x_j} = \sum_j R_{ij} \frac{\partial}{\partial x_j} = \mathbf{R}\,\vec{\nabla}. \tag{1.21}$$

The rotation matrix $\mathbf{R}$ is a particular case of a *Jacobian matrix*, in this case the one for a change of variables. In general, given a vector-valued function, $\vec{f} : \mathbb{R}^n \to \mathbb{R}^m$ which maps a vector $\vec{x} \in \mathbb{R}^n$ onto a vector $\vec{f}(\vec{x}) \in \mathbb{R}^m$, that is, given m functions, $f_1, \ldots, f_m$, each depending on n variables, $x_1, \ldots, x_n$, the *Jacobian matrix* is defined as,

$$\mathbf{J}_{ij} = \frac{\partial f_i}{\partial x_j}\,, \qquad i = 1, \ldots, m\,, \quad j = 1, \ldots, n\,. \tag{1.22}$$

Note that the gradient (1.18) is the particular case of a Jacobian matrix with $m = 1$.

Jacobian matrices are also needed to invert functions. In fact, noting that a function is continuously differentiable if its derivative is also a continuous function, a vector-valued function is continuously differentiable if the entries of the Jacobian matrix are continuous functions. If $m = n$, we can define the determinant of the Jacobian matrix, usually called the *Jacobian*,

$$|\mathbf{J}_{ij}| = \left| \frac{\partial f_i}{\partial x_j} \right| , \qquad i, j = 1, \ldots, n \, .$$

The inverse function theorem states that if a continuously differentiable function $\vec{f} : \mathbb{R}^n \to \mathbb{R}^n$ has a non-vanishing Jacobian at a point $\vec{x}$, then $\vec{f}$ is invertible near $\vec{x}$, and the inverse function is also a continuously differentiable function.

1.3 The form of laws in classical mechanics

The Galileian principle of relativity requires that all laws of physics are the same in inertial reference frames. If we can cast our physics laws as vector equalities, it is then guaranteed that they will have the same form for all frames which are related by a rotation transformation.

$$\begin{aligned} \text{Physics law in frame } S\text{: } \overrightarrow{A} &= \overrightarrow{B} \\ \rightsquigarrow \mathbf{R}\,\overrightarrow{A} &= \mathbf{R}\,\overrightarrow{B} \\ \rightsquigarrow \text{Physics law in frame with rotated axes } S'\text{: } \overrightarrow{A}' &= \overrightarrow{B}' \end{aligned} \tag{1.23}$$

The principle of relativity requires that the laws of physics take the same form not only under rotation transformations but also under "boosts" where two reference frames appear to move with a relative velocity with respect to each other:

$$\vec{r} \to \vec{r}\,' = \vec{r} + \overrightarrow{V} t \tag{1.24}$$

Notice that the acceleration vector

$$\vec{a} \equiv \frac{d^2\vec{r}}{dt^2} = \frac{d^2\vec{r}\,'}{dt^2} = \vec{a}\,' \tag{1.25}$$

is the same for all inertial observers with parallel axes. Newton's law:

$$\overrightarrow{F} = m\vec{a} \tag{1.26}$$

connects the acceleration, a boost invariant quantity, to the force. As long as the latter is also boost invariant then the Galileian principle of relativity is satisfied.

You will soon figure out that the acceleration vector is not boost invariant when using the correct principle of relativity as it is formulated in Einstein's special relativity. Therefore, we should not expect that the physical forces of nature are boost invariant either. Having said that, we will be able to ignore special relativity in most of our classical mechanics investigations, in the same way that Newton and many others were oblivious about it when they set the foundations of modern physics. In particular, our Newtonian understanding of gravity is consistent with the Galileian principle of relativity.

Consider an isolated system of N-particles which interact with each other through the gravitational force only. The force acting on the i-th particle in a reference frame S is

$$m_i \frac{d^2\vec{r}_i}{dt^2} = -G \sum_{j\neq i} m_i m_j \frac{\vec{r}_i - \vec{r}_j}{|\vec{r}_i - \vec{r}_j|^3} \tag{1.27}$$

According to the principle of relativity, a Galileian transformation

$$(t, \vec{r}) \to (t', \vec{r}') = (t + t_0, \mathbf{R}\,\vec{r} + \vec{V}t + \vec{r}_0)$$

should leave Eq. 1.27 invariant. Indeed, this is satisfied. In the new frame,

$$\begin{aligned} m_i \frac{d^2\vec{r}_i{}'}{dt'^2} &= -G \sum_{j\neq i} m_i m_j \frac{\vec{r}_i{}' - \vec{r}_j{}'}{|\vec{r}_i{}' - \vec{r}_j{}'|^3} \\ \rightsquigarrow \mathbf{R}\, m_i \frac{d^2\vec{r}_i}{dt^2} &= -\mathbf{R}\, G \sum_{j\neq i} m_i m_j \frac{\vec{r}_i - \vec{r}_j}{|\vec{r}_i - \vec{r}_j|^3} \\ \rightsquigarrow m_i \frac{d^2\vec{r}_i}{dt^2} &= -G \sum_{j\neq i} m_i m_j \frac{\vec{r}_i - \vec{r}_j}{|\vec{r}_i - \vec{r}_j|^3} \end{aligned} \tag{1.28}$$

Question: Is the electromagnetic force boost invariant?

2 Variational methods

Classical mechanics is usually formulated in terms of Newton's laws in the form of second order differential equations as in Eq. 1.27. There is an alternative, where we can formulate the laws of classical mechanics by means of the so called *variational principle*. It is based on a scalar quantity, called the action. Before we present this formulation, let us warm up with a few physics and mathematics problems in using variation calculus techniques.

2.1 The brachistochrone ($\beta\rho\alpha\chi\iota\sigma\tau os\ \chi\rho o\nu os$ = shortest time) problem

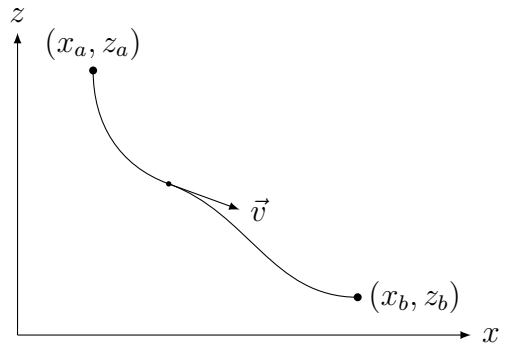


Figure 1: The brachistochrone problem: find which curve connecting (x_a, z_a) and (x_b, z_b) yields the fastest descent.

Consider a point-like object of mass m which is let to slide without friction on a vertical plane from a point (x_a, z_a) to a point (x_b, z_b) constrained on a curve $x = x(z)$, where z is the height and x is the horizontal position of the object. We can compute the velocity of the object at all points $(x(z), z)$ of its trajectory by using energy conservation. Assuming that the initial velocity is zero, we have:

$$mgz_a = mgz + \frac{1}{2}mv^2 \rightsquigarrow v = \sqrt{2g(z_a - z)}. \tag{2.1}$$

The elementary distance segment on the curve $x = x(z)$ is given by the

Pythagorean theorem as:

$$ds = \sqrt{dx^2 + dz^2} = dz\sqrt{1 + \left(\frac{dx}{dz}\right)^2} \tag{2.2}$$

and the magnitude of the velocity is

$$v = \frac{ds}{dt} \rightsquigarrow dt = \frac{ds}{v} = dz\sqrt{\frac{1 + \left(\frac{dx}{dz}\right)^2}{2g(z_a - z)}} \tag{2.3}$$

The total time for the transition is:

$$T = \int dt = \int_{z_a}^{z_b} dz\sqrt{\frac{1 + \left(\frac{dx}{dz}\right)^2}{2g(z_a - z)}} \tag{2.4}$$

We are seeking the function $x(z)$ which minimizes the time T.

2.2 Euler-Lagrange equations

The problem that we have just described above belongs to a more general class of variational calculus problems. We seek to find which function $y(t) = y_s(t)$ among all good-behaved functions $y(t)$ yields an extreme value (minimum or maximum) for the integral

$$I[y] = \int_{t_a}^{t_b} dt F\left(y(t), y'(t)\right) \tag{2.5}$$

given that the values of the function at the initial and final points in the integration domain $y(t_a) = y_a, y(t_b) = y_b$ are known.

Assume that we have found the curve $y_s(t)$ with

$$y_s(t_a) = y_a, y_s(t_b) = y_b \tag{2.6}$$

for which $I[y_s]$ is a minimum of $I[y]$ [1]. Consider now a small deformation around $y_s(t)$

$$y(t) = y_s(t) + \delta y_s(t)\,, \quad \delta y_s(t) = \epsilon\eta(t), \quad \epsilon \to 0\,, \tag{2.7}$$

[1] Notice that $I[y]$ is an integral function of a function. Such integrals are usually called *functionals*

where $\eta(t)$ is a smooth function which vanishes at the boundaries

$$\eta(t_a) = \eta(t_b) = 0, \tag{2.8}$$

in order to satisfy that $y(t_a) = y_a, y(t_b) = y_b$. Since $I[y_s]$ is an extremum, we must have that our small deformation (ignoring $\mathcal{O}(\epsilon^2)$ terms) does not change its value,

$$\begin{aligned} I[y_s] &= I[y_s + \delta y_s] \\ &= \int_{t_a}^{t_b} dt\, F\left(y_s(t) + \delta y_s(t), y_s'(t) + \delta y_s'(t)\right) . \end{aligned} \tag{2.9}$$

We observe that the operations of varying a function and taking its derivative with respect to its argument commute:

$$\delta y_s'(t) = \frac{d}{dt}\delta y_s(t) = \epsilon\eta'(t) \tag{2.10}$$

Thus we can Taylor expand the arguments of F, and write

$$\begin{aligned} I[y_s] &= \int_{t_a}^{t_b} dt\, F\left(y_s(t), y_s'(t)\right) + \int_{t_a}^{t_b} dt\, \left[\frac{\partial F}{\partial y_s}\epsilon\eta(t) + \frac{\partial F}{\partial y_s'}\epsilon\eta'(t)\right] + \mathcal{O}(\epsilon^2) \\ &= I[y_s] + \epsilon\int_{t_a}^{t_b} dt\, \left[\frac{\partial F}{\partial y_s}\eta(t) + \frac{\partial F}{\partial y_s'}\eta'(t)\right] + \mathcal{O}(\epsilon^2) \end{aligned} \tag{2.11}$$

This yields that

$$\begin{aligned} 0 &= \int_{t_a}^{t_b} dt\, \left[\frac{\partial F}{\partial y_s}\eta(t) + \frac{\partial F}{\partial y_s'}\eta'(t)\right] \\ &= \int_{t_a}^{t_b} dt\, \left[\frac{\partial F}{\partial y_s} - \frac{d}{dt}\frac{\partial F}{\partial y_s'}\right]\eta(t) + \int_{t_a}^{t_b} dt\, \frac{d}{dt}\left[\frac{\partial F}{\partial y_s'}\eta(t)\right] \\ &= \int_{t_a}^{t_b} dt\, \left[\frac{\partial F}{\partial y_s} - \frac{d}{dt}\frac{\partial F}{\partial y_s'}\right]\eta(t) + \frac{\partial F}{\partial y_s'}\eta(t)\Big|_{t_a}^{t_b} dt \end{aligned} \tag{2.12}$$

The last term is zero due to $\eta(t)$ vanishing on the boundaries. Thus, the integral

$$\int_{t_a}^{t_b} dt\, \left[\frac{\partial F}{\partial y_s} - \frac{d}{dt}\frac{\partial F}{\partial y_s'}\right]\eta(t) = 0 \tag{2.13}$$

vanishes for every function $\eta(t)$. In order for this to happen, we must have:

$$\frac{\partial F}{\partial y_s} - \frac{d}{dt}\frac{\partial F}{\partial y_s'} = 0. \tag{2.14}$$

This is the Euler-Lagrange equation.

A completely analogous derivation of Euler-Lagrange differential equations can be made for minimising multidimensional integrals which are functionals of many functions. Given a vector-valued function, $\vec{y} : \mathbb{R}^n \to \mathbb{R}^m$ which maps a vector $\vec{x} \in \mathbb{R}^n$ onto a vector $\vec{y}(\vec{x}) \in \mathbb{R}^m$, that is, given m functions, $y_1, \ldots, y_m$, each depending on n variables, $x_1, \ldots, x_n$, we may consider the integral,

$$I[y_1, \ldots, y_m] = \int dx_1 \ldots dx_n F\left[y_i(\vec{x}), \frac{\partial y_i(\vec{x})}{\partial x_j}\right]. \tag{2.15}$$

Requiring that the above integral takes a minimum or maximum value under small variations of the functions

$$y_i(\vec{x}) \to y_i(\vec{x}) + \delta y_i(\vec{x}) = y_i(\vec{x}) + \epsilon\eta_i(\vec{x})\,, \quad i = 1, \ldots, m\,,$$

the integral value should be unchanged:

$$\begin{aligned} 0 &= \delta I[y_1, \ldots, y_m] \\ &= \delta \int dx_1 \ldots dx_n F\left[y_i(\vec{x}), \frac{\partial y_i(\vec{x})}{\partial x_j}\right] \\ &= \int dx_1 \ldots dx_n \left[\sum_{i=1}^{m} \frac{\partial F}{\partial y_i}\delta y_i(\vec{x}) + \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{\partial F}{\partial(\partial_j y_i)}\delta\partial_j y_i(\vec{x})\right], \end{aligned} \tag{2.16}$$

where we have introduced the shorthand notation,

$$\partial_j \equiv \frac{\partial}{\partial x_j}.$$

We use again that the operations of varying a function and taking its derivative with respect to its arguments commute:

$$\delta\partial_j y_i(\vec{x}) = \partial_j \delta y_i(\vec{x}) \tag{2.17}$$

and after performing integration by parts, we have

$$\begin{aligned} 0 &= \epsilon \int dx_1 \dots dx_n \sum_{i=1}^{m} \eta_i \left[\frac{\partial F}{\partial y_i} - \sum_{j=1}^{n} \partial_j \frac{\partial F}{\partial(\partial_j y_i)} \right] \\ &\quad + \epsilon \sum_{ij} \int dx_1 \dots dx_m \partial_j \left[\eta_i \frac{\partial F}{\partial(\partial_j y_i)} \right]. \end{aligned} \tag{2.18}$$

The second term is a total derivative in the x_j integration and it vanishes by requiring that the variations $y_i(\vec{x})$ vanish at the boundaries of the integration. Since the above identity must be valid for an arbitrary choice of functions $\eta_i(\vec{x})$ we arrive at the general Euler-Lagrange equations:

$$\frac{\partial F}{\partial y_i} - \sum_{j=1}^{n} \partial_j \frac{\partial F}{\partial(\partial_j y_i)} = 0\,, \qquad i = 1, \dots, m\,. \tag{2.19}$$

A particularly useful case is the one of a vector-valued function of one variable, $\vec{y} : \mathbb{R} \to \mathbb{R}^m$ which maps a variable $x \in \mathbb{R}$ onto a vector $\vec{y}(\vec{x}) \in \mathbb{R}^m$, that is, the case treated above with $n = 1$. Then the general Euler-Lagrange equations become

$$\frac{\partial F}{\partial y_i} - \frac{d}{dx} \frac{\partial F}{\partial y_i'} = 0\,, \qquad i = 1, \dots, m\,. \tag{2.20}$$

2.3 Propagation of light

Equipped with the Euler-Lagrange equations (2.20) we consider the propagation of light. Fermat formulated the homonymous principle, stating that light travels from point to point choosing the path which yields the fastest time. For such a transition, the total time is

$$T = \int dt = \int \frac{ds}{v(\vec{x})}, \tag{2.21}$$

where ds is the distance element and $v(\vec{x})$ is the speed of light at the point $\vec{x}$. This is usually normalised to the speed of light in the vacuum c,

$$v(\vec{x}) = \frac{c}{n(\vec{x})}, \tag{2.22}$$

where $n(\vec{x})$ is the refraction index. Fermat's principle requires that

$$\begin{aligned} T &= \frac{1}{c}\int ds\, n(\vec{x}) = \frac{1}{c}\int \sqrt{dx^2+dy^2+dz^2}\, n(\vec{x}) \\ &= \frac{1}{c}\int dx \sqrt{1+\left(\frac{dy}{dx}\right)^2+\left(\frac{dz}{dx}\right)^2}\, n(\vec{x}) \end{aligned} \tag{2.23}$$

is a minimum.

Let us assume that the index of refraction is a constant,

$$n(\vec{x}) = n = \text{const.}$$

Then

$$T = \frac{n}{c}\int dx \sqrt{1+\left(\frac{dy}{dx}\right)^2+\left(\frac{dz}{dx}\right)^2}. \tag{2.24}$$

Formally, we can write it as,

$$T[y,z] = \int dx F\left(y'(x), z'(x)\right) \tag{2.25}$$

with

$$F\left(y'(x), z'(x)\right) = \frac{n}{c}\sqrt{1+y'^2(x)+z'^2(x)}\,. \tag{2.26}$$

Note that F does not depend explicitly on y and z.

The Euler-Lagrange equations read:

$$\begin{aligned} \frac{\partial F}{\partial y} - \frac{d}{dx}\frac{\partial F}{\partial y'} &= 0 \\ \frac{\partial F}{\partial z} - \frac{d}{dx}\frac{\partial F}{\partial z'} &= 0\,. \end{aligned} \tag{2.27}$$

Then computing the derivatives,

$$\begin{aligned} \frac{\partial F}{\partial y} &= 0\,, & \frac{\partial F}{\partial y'} &= \frac{n}{c}\frac{y'}{\sqrt{1+y'^2+z'^2}} \\ \frac{\partial F}{\partial z} &= 0\,, & \frac{\partial F}{\partial z'} &= \frac{n}{c}\frac{z'}{\sqrt{1+y'^2+z'^2}} \end{aligned} \tag{2.28}$$

the Euler-Lagrange equations become,

$$\begin{aligned} \frac{d}{dx}\frac{y'}{\sqrt{1+y'^2+z'^2}} &= 0 \\ \frac{d}{dx}\frac{z'}{\sqrt{1+y'^2+z'^2}} &= 0\,, \end{aligned} \tag{2.29}$$

which can be trivially integrated,

$$\begin{aligned} \frac{y'}{\sqrt{1+y'^2+z'^2}} &= c_1 \\ \frac{z'}{\sqrt{1+y'^2+z'^2}} &= c_2\,, \end{aligned} \tag{2.30}$$

where c_1 and c_2 are constants. The solution of these equations is a straight line.

Hint: In order to see it, solve first the simpler case of a light ray in the $x-y$ plane. Then the Euler-Lagrange equation is

$$\frac{d}{dx}\frac{y'}{\sqrt{1+y'^2}} = 0 \tag{2.31}$$

which is integrated to

$$\frac{y'}{\sqrt{1+y'^2}} = c \tag{2.32}$$

where c is a constant. This can be solved in y',

$$y' = \pm\frac{c}{\sqrt{1-c^2}} = a \tag{2.33}$$

where a is a constant. Integrating it, we get the equation of a straight line

$$y(x) = ax + b\,. \tag{2.34}$$

Likewise, the coupled system (2.30) can be solved to find that

$$\begin{aligned} y(x) &= a_1 x + b_1 \\ z(x) &= a_2 x + b_2\,. \end{aligned} \tag{2.35}$$

which are the parametric equations of a straight line.

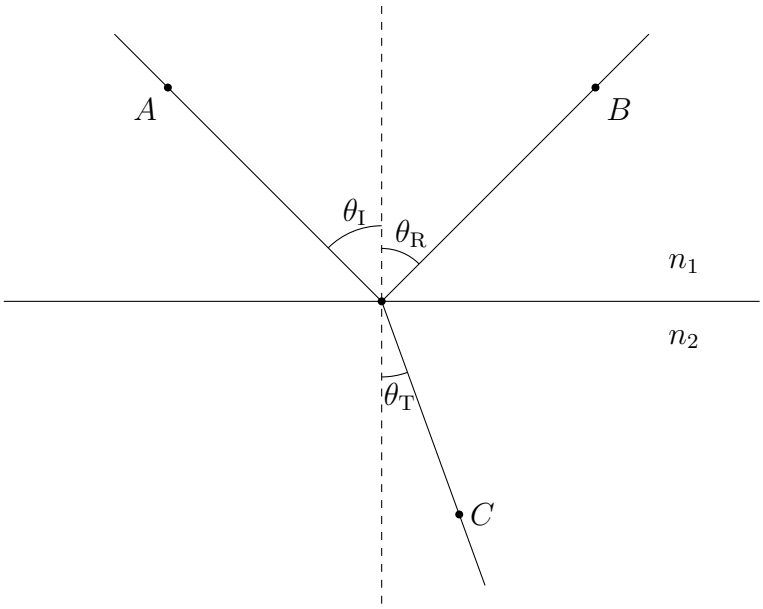


Figure 2: The transmission and reflection of light through a boundary separating media with different refraction indices

The principle of fastest time yields the correct laws for the reflection and refraction of light of geometrical optics, stating that the incidence and reflection angles are the same:

$$\theta_I = \theta_R \tag{2.36}$$

and that the transmitted light from a medium with refraction index n_1 to a medium with a refraction index n_2 is determined from Snell's law:

$$n_1 \sin\theta_I = n_2 \sin\theta_T. \tag{2.37}$$

Indeed, consider a light ray which passes from the point $A(x_a, y_a)$ in Fig. 2 and arrives at a point $B(x_b, y_b)$ after it gets reflected on a boundary separating two different media at the point $O(0, 0)$. The time required for this transition is

$$T_{A\to B} = \frac{\sqrt{x_a^2 + y_a^2}}{v_1} + \frac{\sqrt{x_b^2 + y_b^2}}{v_1}, \tag{2.38}$$

where $v_1 = \frac{c}{n_1}$ is the speed of light in the medium at the side of the boundary where the incident and reflected rays lie. If $L = x_b - x_a$ is the horizontal

distance with the two points, then we write:

$$cT_{A\to B} = n_1\sqrt{x_a^2+y_a^2} + n_1\sqrt{(L+x_a)^2+y_b^2} \tag{2.39}$$

Requiring the above to be a minimum, we find that the horizontal position x_a must satisfy

$$\begin{aligned} & \frac{\partial T_{A\to B}}{\partial x_a} = 0 \\ \rightsquigarrow \quad & \frac{x_a}{\sqrt{x_a^2+y_a^2}} + \frac{(L+x_a)}{\sqrt{(L+x_a)^2+y_b^2}} = 0 \\ \rightsquigarrow \quad & -\sin\theta_I + \sin\theta_R = 0\,, \end{aligned} \tag{2.40}$$

which yields that the incident and reflected light angles are the same,

$$\theta_I = \theta_R. \tag{2.41}$$

Repeating the same steps for a transition from the point A inside the medium with a refraction index n_1 to a point C inside the medium with refraction index n_2,

$$\begin{aligned} & \frac{\partial T_{A\to C}}{\partial x_a} = 0 \\ \rightsquigarrow \quad & n_1\frac{x_a}{\sqrt{x_a^2+y_a^2}} + n_2\frac{(L+x_a)}{\sqrt{(L+x_a)^2+y_c^2}} = 0 \\ \rightsquigarrow \quad & -n_1\sin\theta_I + n_2\sin\theta_T = 0\,, \end{aligned} \tag{2.42}$$

we find Snell's law,

$$n_1\sin\theta_I = n_2\sin\theta_T. \tag{2.43}$$

Interesting phenomena occur when the index of refraction is not uniform. For example, the atmospheric air density changes with the temperature. The non-uniformity of the density is responsible for for interesting phenomena, such as mirage images in deserts or on the surface of a hot road in the summer and the twinkling of stars. In such media, the trajectory of the light is not a straight line.

2.4 Solution of the brachistochrone problem

We can return now to the problem of subsection 2.1. The functional T of Eq. 2.4 can be written as

$$T[x] = \int_{z_a}^{z_b} dz F(x'(z)) \tag{2.44}$$

with

$$F = \sqrt{\frac{1 + x'^2(z)}{2g(z_a - z)}} \tag{2.45}$$

Note that F does not depend explicitly on x.

Requiring that the functional (2.44) be a minimum, we can write the Euler-Lagrange equation 2.14

$$\frac{\partial F}{\partial x} - \frac{d}{dz}\frac{\partial F}{\partial x'} = 0. \tag{2.46}$$

Performing the differentiations,

$$\frac{\partial F}{\partial x} = 0\,, \qquad \frac{\partial F}{\partial x'} = \frac{x'}{\sqrt{2g(z_a - z)(1 + x'^2(z))}}\,, \tag{2.47}$$

we obtain the second order differential equation:

$$\frac{d}{dz}\frac{x'}{\sqrt{(z_a - z)(1 + x'^2)}} = 0 \tag{2.48}$$

which, after a trivial first integration, gives:

$$\frac{x'}{\sqrt{(z_a - z)(1 + x'^2)}} = -C, \tag{2.49}$$

with C a constant to be fixed by our boundary conditions. Solving for x', we obtain:

$$x'^2 = \frac{C^2(z_a - z)}{1 - C^2(z_a - z)} \tag{2.50}$$

For the rhs to be a positive definite variable, we must have either

$$z \geq z_a \text{ and } C^2(z_a - z) \geq 1,$$

or

$$z \leq z_a \text{ and } C^2(z_a - z) \leq 1.$$

The first possibility is not physically allowed, since it is not satisfied by the final point (x_b, z_b) which lies lower than (x_a, z_a).

Let's now change variables:

$$C^2(z_a - z) = \sin^2\phi, \quad \phi \in \left[0, \frac{\pi}{2}\right] \tag{2.51}$$

with

$$\frac{dz}{d\phi} = -\frac{1}{C^2}\sin(2\phi). \tag{2.52}$$

The differential equation for $x(\phi)$ becomes:

$$\frac{dx}{d\phi} = 2\frac{C}{|C|^3}\sin^2\phi \tag{2.53}$$

It is now straightforward to perform the integrations over ϕ and compute $x = x(\phi), z = z(\phi)$. We can choose that $\phi = 0$ corresponds to the first point (x_a, z_a) of the curve. This gives the solutions:

$$x(\phi) = x_a + \frac{C}{2|C|^3}\left[2\phi - \sin(2\phi)\right] \tag{2.54}$$

$$z(\phi) = z_a - \frac{1}{2C^2}\left[1 - \cos(2\phi)\right] \tag{2.55}$$

Setting $\phi = \theta/2$ with $\theta \in [0, \pi]$, we have

$$x(\theta) = x_a + \frac{C}{2|C|^3}\left[\theta - \sin\theta\right] \tag{2.56}$$

$$z(\theta) = z_a - \frac{1}{2C^2}\left[1 - \cos\theta\right] \tag{2.57}$$

Choosing a positive value for C we obtain solutions where the curve lies to the right of the starting point (as in Figure 1) while for $C < 0$ we obtain solutions where the curve lies to the left of the starting point. For $C > 0$ we have

$$x(\theta) = x_a + \frac{1}{2C^2}\left[\theta - \sin\theta\right] \tag{2.58}$$

$$z(\theta) = z_a - \frac{1}{2C^2}\left[1 - \cos\theta\right] \tag{2.59}$$

These are the equations of a cycloid. To determine the constant C^2 we require that the final point (x_b, z_b) belongs to the curve, i.e. there exists a value $\theta = \theta_b$ such that

$$x_b = x_a + \frac{1}{2C^2}\left[\theta_b - \sin\theta_b\right] \tag{2.60}$$

$$z_b = z_a - \frac{1}{2C^2}\left[1 - \cos\theta_b\right] \tag{2.61}$$

For any specific value of the pairs (x_a, z_a) and (x_b, z_b) one can use eqs. (2.60) and (2.61) to determine numerically θ_b and C, and from this the radius of the cycle generating the required cycloid curve,

$$r = \frac{1}{2C^2} \tag{2.62}$$

Exercise 2.1. Determine C and r for $(x_a, z_a) = (0, 1)$ and $(x_b, z_b) = (2, 0)$.

You can find a nice demonstration of the brachistochrone problem in this video.

3 Hamilton's principle of least action

After the observation that the behaviour of light can be understood by means of a "least time principle" and variational techniques, the question arises whether variational methods can also be used more generally for the study of physical systems.

The main result of variational methods is the Euler-Lagrange differential equations. The laws of classical physics, Newton's laws and Maxwell equations of electrodynamics were first discovered in the form of differential equations. The question we are posing here is a reverse-engineering problem. If we know the Euler-Lagrange equations (physics laws in differential equations form) can we find the integral functional from which they arise? Hamilton's principle of least action states that this is possible.

Consider an isolated system of many particles $i = 1 \dots N$ for which we would like to know their trajectories, i.e their positions at all times. While it is common to use Cartesian coordinates, $\vec{r}_i(t)$, it is often useful to use other variables such as angles, radii, or even more exotic variables to describe the trajectory of a particle. We call these generic variables which describe a trajectory as *generalised coordinates*.

Assume now that we know at an initial time t_i the values of a set of **independent** generalised coordinates

$$q_1(t_i), \dots, q_m(t_i)$$

as well as the initial values of their time derivarives (generalised velocities)

$$\dot{q}_1\ (t_i), \dots, \dot{q}_m\ (t_i).$$

Hamilton's principle of least action states that the generalised coordinates and generalised velocities of the system are determined at a later time t_f by requiring that *action*:

$$S \equiv \int_{t_i}^{t_f} dt\, L\left[\{q_j(t), \dot{q}_j\ (t)\}\right] \tag{3.1}$$

is minimum:

$$\delta S = 0. \tag{3.2}$$

The integrand L of the action is called the *Lagrangian*. The Euler-Lagrange equations derived from the principle of least action for physical systems are

called *equations of motion* and are given:

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0. \tag{3.3}$$

We note that the Lagrangian of a system is not uniquely defined. Consider two Lagrangians L, L' which differ by a total derivative:

$$L' = L + \frac{df}{dt}, \tag{3.4}$$

where f is an arbitrary function. The corresponding actions

$$S' = \int_{t_i}^{t_f} dt\, L' = \int_{t_i}^{t_f} dt\, L + \int_{t_i}^{t_f} dt\, \frac{df}{dt} = S + f(t_f) - f(t_i) \tag{3.5}$$

differ by a constant which vanishes upon taking a variation of the generalised coordinates:

$$\delta S' = \delta S + \delta\left(f(t_f) - f(t_i)\right) = \delta S. \tag{3.6}$$

Therefore, Lagrangians which differ by a total derivative yield the same equations of motion.

3.1 The Lagrangian of a free particle in an inertial frame

We will now find Lagrangians which govern known physical systems, starting from the simplest case. This is a free isolated particle (does not interact with other particles) which is observed in an inertial frame. The equation of motion of the particle is given by

$$\vec{a} = 0, \tag{3.7}$$

where the acceleration of the free particle is

$$\vec{a} \equiv \ddot{\vec{r}} = \left(\ddot{r}_1, \ddot{r}_2, \ddot{r}_3\right). \tag{3.8}$$

The Lagrangian:

$$L = \frac{1}{2}m\vec{v}^2 = \frac{1}{2}m\sum_{i=1}^{3}\dot{r}_i^2, \tag{3.9}$$

yields the correct equation of motion:

$$0 = \frac{d}{dt}\frac{\partial L}{\partial \dot{r}_i} - \frac{\partial L}{\partial r_i} = m\ddot{r}_i \rightsquigarrow \ddot{r}_i = 0. \tag{3.10}$$

Could we have guessed a different Lagrangian, which results to the acceleration being zero for a freely moving particle? For example, a Lagrangian:

$$L = (\vec{v}^2)^n \tag{3.11}$$

yields the equations of motion

$$\left(\dot{r}_i^2\right)^{n-1}\ddot{r}_i = 0. \tag{3.12}$$

For $n \neq 1$, in a rest frame where the particle is moving at a certain moment with a velocity $\dot{\vec{r}} \neq 0$, we find that the acceleration must be zero. Nevertheless, such equations of motion do not conform with Galileo's principle of relativity which requires that under, for example, a boost

$$\vec{r} \to \vec{r}' = \vec{r} + \vec{v}t, \tag{3.13}$$

Eq. 3.12 must maintain the same form, i.e.

$$\left(\dot{r}_i'^2\right)^{n-1}\ddot{r}_i' = 0. \tag{3.14}$$

It does not, except for $n = 1$! Instead, one finds a frame dependent form:

$$\left(\dot{r}_i^2\right)^{n-1}\ddot{r}_i = 0 \overset{r_i \to r_i'}{\rightsquigarrow} \left((\dot{r}_i' - v_i)^2\right)^{n-1}\ddot{r}_i' = 0. \tag{3.15}$$

A comment is appropriate on the constant factor $\frac{m}{2}$ in the Lagrangian of the free particle of Eq. 3.9. Obviously, this factor is immaterial in deriving the equations of motion. However, we will soon extend the Lagrangian to account for interactions with other particles. When this happens, we will associate m with the physical mass of the particle. The principle of least action tells us that the physical mass is a positive quantity $m > 0$. Indeed, the general solution of the equations of motion is:

$$\dddot{\vec{r}} = 0 \rightsquigarrow \vec{r}(t) = \vec{A} + \vec{B}t. \tag{3.16}$$

We can fix the constants $\overrightarrow{A}, \overrightarrow{B}$ for a given transition from a space-time point $(t_i, \vec{r}_i)$ to a space-time point $(t_f, \vec{r}_f)$. Then, the physical trajectory is given by

$$\vec{r}_{\text{phys}}(t) = \vec{r}_i + \frac{t - t_i}{t_f - t_i} \left(\vec{r}_f - \vec{r}_i\right). \tag{3.17}$$

The value of the action for the physical trajectory is

$$S\left[\vec{r}_{\text{phys}}(t)\right] = \int_{t_i}^{t_f} dt L[\vec{r}_{\text{phys}}(t)] = \frac{1}{2} m \frac{\left(\vec{r}_f - \vec{r}_i\right)^2}{t_f - t_i}. \tag{3.18}$$

According to the principle of least action, any other trajectory than the physical that we can think of to join the space-time points $(t_i, \vec{r}_i)$ and $(t_f, \vec{r}_f)$ should yield a value for the action which is greater. Consider, for example a non-physical trajectory:

$$\vec{r}_{\text{non-phys}}(t) = \vec{r}_i + \left(\frac{t - t_i}{t_f - t_i}\right)^2 \left(\vec{r}_f - \vec{r}_i\right). \tag{3.19}$$

It must be

$$S\left[\vec{r}_{\text{non-phys}}(t)\right] - S\left[\vec{r}_{\text{phys}}(t)\right] > 0. \tag{3.20}$$

An explicit calculation shows that

$$S\left[\vec{r}_{\text{non-phys}}(t)\right] = \int_{t_i}^{t_f} dt L[\vec{r}_{\text{non-phys}}(t)] = \frac{2}{3} m \frac{\left(\vec{r}_f - \vec{r}_i\right)^2}{t_f - t_i} = \frac{4}{3} S\left[\vec{r}_{\text{phys}}(t)\right]. \tag{3.21}$$

Eq. 3.20 leads to the conclusion that the value of the action for the physical trajectory must be positive:

$$S\left[\vec{r}_{\text{phys}}(t)\right] = \frac{1}{2} m \frac{\left(\vec{r}_f - \vec{r}_i\right)^2}{t_f - t_i} > 0 \tag{3.22}$$

and, since $t_f > t_i$, we must have that the mass is positive $m > 0$.

3.2 Lagrangian of a particle in a homogeneous force field

Consider a particle subjected to a constant force $\overrightarrow{F}$. According to Newton's law, the equation of motion is

$$m\ddot{\vec{r}} - \overrightarrow{F} = 0, \tag{3.23}$$

or, in components,

$$m\,\ddot{r}_i - F_i = 0. \tag{3.24}$$

As for the earlier case of the free particle, the first term of the above equation can be written as

$$m\,\ddot{r}_i = \frac{d}{dt}\frac{\partial T}{\partial\,\dot{r}_i}, \tag{3.25}$$

where

$$T = \frac{1}{2} m\,\dot{r}^2, \tag{3.26}$$

the kinetic energy of the particle. We can write the second term as a derivative:

$$F_i = \frac{\partial(F_i r_i)}{\partial r_i} = \frac{\partial}{\partial r_i}\sum_{j=1}^{3} F_j r_j = -\frac{\partial(-\vec{F}\cdot\vec{r})}{\partial r_i} \tag{3.27}$$

The dot product $-\vec{F}\cdot\vec{r}$ is the work done against the force to bring the particle from the origin of our reference frame to the position $\vec{r}$, or in other words the potential energy of the particle:

$$U(\vec{r}) = -\vec{F}\cdot\vec{r}. \tag{3.28}$$

Therefore, Newton's law for a constant force takes the form:

$$\frac{d}{dt}\frac{\partial T}{\partial\,\dot{r}_i} - \frac{\partial(-U)}{\partial r_i} = 0. \tag{3.29}$$

Given that the potential energy U does not depend on velocities $\dot{r}_i$ and that the kinetic energy T does not depend on positions, we can rewrite the above equation as:

$$\frac{d}{dt}\frac{\partial(T-U)}{\partial\,\dot{r}_i} - \frac{\partial(T-U)}{\partial r_i} = 0. \tag{3.30}$$

This has taken the form of Euler-Lagrange equations,

$$\frac{d}{dt}\frac{\partial L}{\partial\,\dot{r}_i} - \frac{\partial L}{\partial r_i} = 0. \tag{3.31}$$

where the Lagrangian

$$L = T - U \tag{3.32}$$

is the difference of the kinetic and potential energy of the particle. While we have derived this result for a constant force $\vec{F}$, we shall see that it is more general.

3.3 Lagrangian of a system of particles with conserved forces

Consider an isolated system of N-particles, i.e. particles which interact only with each other. We will postulate that the Lagrangian of the system is the difference of the kinetic energy of the particles and the potential energy of the system:

$$L = \sum_a \frac{1}{2} m_a \dot{\vec{r}}_a^2 - U\left(\vec{r}_1, \ldots, \vec{r}_n\right). \tag{3.33}$$

The Euler-Lagrange equations for this Lagrangian read:

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{r}_{a,i}} = \frac{\partial L}{\partial r_{a,i}}, \tag{3.34}$$

where $r_{a,i}$ denotes the $i-th$ component of the vector $\vec{r}_a$ denoting the position of the $a-$the particle. Performing the differentiations we obtain:

$$m_a \ddot{r}_{a,i} = -\frac{\partial U}{\partial r_{a,i}} \tag{3.35}$$

or, in vector notation,

$$m\ddot{\vec{r}}_a = -\vec{\nabla}_a U \tag{3.36}$$

with $\vec{\nabla}_a \equiv (\partial/\partial r_{a,1}, \partial/\partial r_{a,2}, \partial/\partial r_{a,3})$. This is the known equation of motion according to Newton's law for a conserved force, i.e. a force which can be written as the gradient of a potential:

$$\vec{F}_a = -\vec{\nabla}_a U. \tag{3.37}$$

As an example consider an isolated system of N particles with masses m_a which interact among themselves gravitationally. The Lagrangian of the system is

$$L = T - U \tag{3.38}$$

where

$$T = \sum_{a=1}^{N} \frac{1}{2} m_a \dot{\vec{r}}_a \tag{3.39}$$

and

$$U = -\frac{1}{2} \sum_a \sum_{b \neq a} G \frac{m_a m_b}{|\vec{r}_a - \vec{r}_b|} \tag{3.40}$$

The Euler-Lagrange equations are:

$$m_a \ddot{\vec{r}}_a = -\overrightarrow{\nabla}_a U, \tag{3.41}$$

where the subscript in $\overrightarrow{\nabla}_a$ denotes that the differentiations are made with respect to the coordinates of the a-th particle.

Exercise 3.1. To carry out the differentiations in the rhs, we first prove that:

$$\overrightarrow{\nabla}_a \frac{1}{|\vec{r}_a - \vec{r}_b|} = -\frac{\vec{r}_a - \vec{r}_b}{|\vec{r}_a - \vec{r}_b|^3}. \tag{3.42}$$

We then obtain the equations of motion:

$$m_a \ddot{\vec{r}}_a = -\sum_{b \neq a} G m_a m_b \frac{\vec{r}_a - \vec{r}_b}{|\vec{r}_a - \vec{r}_b|^3}, \tag{3.43}$$

in agreement with Newton's law of gravitation.

3.4 Lagrangian for a charge inside an electromagnetic field

We now turn our attention to the force experienced by an electric charge which is in the vicinity of other electric charges. We can sum up the effects of all other charges into two vectors:

- $\overrightarrow{E}(t, \vec{r})$, the electric field and
- $\overrightarrow{B}(t, \vec{r})$, the magnetic field.

The force is then given by:

$$m\ddot{\vec{r}} = q\left(\overrightarrow{E} + \dot{\vec{r}} \times \overrightarrow{B}\right), \tag{3.44}$$

where m is the mass of the charge q. The electric and magnetic fields are determined from the **equations of Maxwell**:

$$\overrightarrow{\nabla} \cdot \overrightarrow{E} = \frac{\rho}{\epsilon_0}, \tag{3.45}$$

$$\overrightarrow{\nabla} \times \overrightarrow{E} = -\frac{\partial \overrightarrow{B}}{\partial t}, \tag{3.46}$$

$$\overrightarrow{\nabla} \cdot \overrightarrow{B} = 0, \tag{3.47}$$

$$\overrightarrow{\nabla} \times \overrightarrow{B} = \frac{\vec{j}}{c^2 \epsilon_0} + \frac{1}{c^2} \frac{\partial \overrightarrow{E}}{\partial t.} \tag{3.48}$$

where $\rho(\vec{r})$ is the electric charge density and $\vec{j}$ is the electric current density. ϵ_0 is a constant, the so-called vacuum permittivity, and has the value

$$\epsilon_0 = 8.854187817\ldots 10^{12}\frac{A\cdot s}{\text{Volt}\cdot m}. \tag{3.49}$$

c is the speed of light

$$c = 2.99792458\ldots 10^{8}\frac{m}{s}. \tag{3.50}$$

We will not attempt here to find explicit solutions of Maxwell equations for the $\overrightarrow{E}$ and $\overrightarrow{B}$ fields, but we will assume that such solutions exist and are known to us. Can we obtain the equation of motion Eq. 3.44 from Hamilton's variational principle? Unlike the example that we have seen so far, the electormagnetic force depends not only on the position of the charged particle but also its velocity. Nevetheless, we will be able to find a Lagrangian which gives the correct equation of motion.

To achieve that, we introduce first the scalar and vector potentials, $\phi(t,\vec{r})$ and $\overrightarrow{A}(t,\vec{r})$ respectively, defined as:

$$\overrightarrow{E} = -\frac{\partial\overrightarrow{A}}{\partial t} - \overrightarrow{\nabla}\phi \tag{3.51}$$

$$\overrightarrow{B} = \overrightarrow{\nabla}\times\overrightarrow{A} \tag{3.52}$$

You can easily verify that the second and third of Maxwell equations are automatically satisfied if we substitute the electric and magnetic field with the scalar and vector potentials. Consider now the Lagrangian:

$$L = \frac{1}{2}m\dot{\vec{r}}^2 - q\phi + q\dot{\vec{r}}\cdot\overrightarrow{A}. \tag{3.53}$$

Euler-Lagrange equations give the form

$$m\,\ddot{r}_i + q\frac{dA_i}{dt} = -q\partial_i\phi + q\sum_j \dot{r}_j\,\partial_i A_j, \tag{3.54}$$

where we use the shorthand notation

$$\partial_i \equiv \frac{\partial}{\partial r_i}.$$

The total time derivative of the vector potential is:

$$\frac{dA_i}{dt} = \frac{\partial A_i}{\partial t} + \sum_j \dot{r}_j \, (\partial_j A_i) \tag{3.55}$$

Then, Euler-Lagrange equations take the form

$$m \, \ddot{r}_i \;\; = \;\; q \left[-\frac{\partial A_i}{\partial t} - \partial_i \phi \right] + q \sum_j \left[\dot{r}_j \, \partial_i A_j - \dot{r}_j \, (\partial_j A_i) \right]. \tag{3.56}$$

In the first bracket, we recognize the i-th component of the electric field:

$$-\frac{\partial A_i}{\partial t} - \partial_i \phi = E_i \tag{3.57}$$

The second bracket is not as obvious to decipher, but we can prove it to be the i-th component of the cross-product of the velocity and the magnetic field. Indeed

$$\begin{aligned} (\dot{\vec{r}} \times \vec{B})_i &= \left[\dot{\vec{r}} \times \left(\vec{\nabla} \times \vec{A} \right) \right]_i \\ &= \sum_{jklm} \epsilon_{ijk} \, \dot{r}_j \, (\epsilon_{klm} \partial_l A_m) \\ &= \sum_{jlm} (\delta_{il}\delta_{jm} - \delta_{jl}\delta_{im}) \, \dot{r}_j \, \partial_l A_m \\ &= \sum_j \left[\dot{r}_j \, \partial_i A_j - \dot{r}_j \, (\partial_j A_i) \right]. \end{aligned} \tag{3.58}$$

Thus, we have proven that the Euler-Lagrange equations of the Lagrangian of Eq. 3.53 give the known Lorentz equation for the force of a charge inside an electromagnetic field:

$$m \, \ddot{r}_i = q E_i + q \left(\dot{\vec{r}} \times \vec{B} \right)_i \tag{3.59}$$

4 Dynamics of constrained systems

In the last section, we showed the equivalency of traditional Newtonian mechanics and Hamilton's variational principle for a variety of forces acting on systems of particles. In this section, we will show that the principle of least action can be applied to dynamical systems for which we do not have a priori knowledge of all forces acting on the particles but for some of them we only know their effect in limiting the allowed motion of the particles. For example, particles which make up a ball are constrained to be kept together with electromagnetic forces. It is difficult to account for these microscopic electromagnetic interactions in our Lagrangian when our aim is simply to describe the motion of the ball inside the gravitational field. However, in a variety of problems with constraints an explicit description of the forces responsible for the constraints can be avoided and we can implement directly their effect.

4.1 Constraints

The configuration of a system of N particles P_j is given by the positions $\vec{r}_j$ and the velocities $\dot{\vec{r}}_j$ of the N particles. If all possible configurations are allowed, the system is *free*. If there are limitations to the possible configurations, the motion of the system is constrained. The geometric or kinematic restrictions to the positions $\vec{r}_j$ of the particles of the system are called *constraints*.

We will study a class of constraints which limit the positions and/or the velocities of the particles, through equations of the form

$$f(\vec{r}_1, \ldots, \vec{r}_N, \dot{\vec{r}}_1, \ldots, \dot{\vec{r}}_N, t) = 0\,. \tag{4.1}$$

We call these constraints *differential*. If a differential constraint can be cast in a form that it does not depend explicitly on the velocities, it is called holonomic,

$$f(\vec{r}_1, \ldots, \vec{r}_N, t) = 0\,, \tag{4.2}$$

thus it will limit only the positions of the particles. If, in particular, a holonomic constraint (Eq. 4.2) does not depend explicitly on time,

$$\frac{\partial f}{\partial t} = 0\,, \tag{4.3}$$

it is called *stationary*. Constraints which are not holonomic are called *non-holonomic*.

Examples of holonomic constraints are:

1. A rigid body may be thought as a collection of N particles whose reciprocal distances L_{ij} define geometric stationary constraints like in eq. (4.10),

$$(\overrightarrow{r_i} - \overrightarrow{r_j})^2 - L_{ij}^2 = 0\,, \qquad i,j = 1,\ldots,N\,. \tag{4.4}$$

   This is a stationary holonomic constraint. The rigid body may be in motion, therefore the coordinates $\vec{r}_i = \vec{r}_i(t)$, $\vec{r}_j = \vec{r}_j(t)$ change with time. However the time-dependence of Eq. 4.10 is implicit (through the coordinates) and not explicit; thus we classify the constraint as stationary.

2. A particle is constrained to move on a fixed surface,

$$f(\vec{r}) = f(x,y,z) = 0\,. \tag{4.5}$$

   This is also a stationary holonomic constraint.

3. A particle is constrained to move on a surface which is itself in motion,

$$f(\vec{r},t) = 0\,. \tag{4.6}$$

   This is a holonomic constraint, but it depends explicitly on time. Therefore, the constraint is not stationary.

4. An ideal moving fluid (i.e. without viscosity) may be thought as a collection of N particles whose reciprocal distances $L_{ij}(t)$ change with time and which define holonomic constraints which are not stationary:

$$(\overrightarrow{r_i} - \overrightarrow{r_j})^2 - L_{ij}^2(t) = 0\,, \qquad i,j = 1,\ldots,N\,. \tag{4.7}$$

Differential constraints may appear, at a first sight, to depend on velocities. We may then rush to conclude that they are not holonomic. However, sometimes the apparent velocity dependence can be eliminated upon integration and we can, after all, cast the constraint in the form of Eq. 4.2 which is manifestly holonomic. To convince you that such a possibility exists, it suffices to observe that we can always produce a constraint with an apparent

velocity dependence starting from a holonomic constraint. Let us start from a constraint of the form of Eq. 4.2 and take a total time derivative:

$$f = 0 \quad \longrightarrow \quad \frac{df}{dt} = 0 \quad \longrightarrow \quad \sum_{j=1}^{N} \vec{\nabla}_j \cdot \dot{\vec{r}}_j + \frac{\partial f}{\partial t} = 0 \,, \tag{4.8}$$

which in Cartesian coordinates $\vec{r}_j \equiv (x_j, y_j, z_j)$ is written as,

$$\sum_{j=1}^{N} \left(\frac{\partial f}{\partial x_j} \dot{x}_j + \frac{\partial f}{\partial y_j} \dot{y}_j + \frac{\partial f}{\partial z_j} \dot{z}_j \right) + \frac{\partial f}{\partial t} = 0 \,. \tag{4.9}$$

In the form of Eq. 4.9, the constraint has an apparent velocity dependence. As a concrete example, consider a pendulum which moves on the $x - z$ plane, and it is attached to a string of length R. This yields a holonomic stationary constraint which is given by the equation,

$$\vec{r}^{\,2} - R^2 = x^2 + z^2 - R^2 = 0 \,. \tag{4.10}$$

which states that the distance of the pendulum from the center is fixed. We can make the constraint to have an apparent velocity dependence. Differentiating Eq. (4.10) with respect to time, we obtain

$$\frac{d}{dt}(x^2 + z^2 - R^2) = 2x\dot{x} + 2z\dot{z} = 0 \,, \quad \longrightarrow \quad \vec{r} \cdot \dot{\vec{r}} = 0 \,, \tag{4.11}$$

In this form, we read the constraint to state that the motion of the pendulum is orthogonal to the string. Had we been given the constraint in this second form, we would not be able to classify it as holonomic. Of course, once we realize that the differential constraint of Eq. 4.11 is a total differential, it can be integrated back making the form of the constraint manifestly holonomic.

Differential constraints which are total time differentials (Eq. 4.8) can be integrated and we obtain

$$f(\vec{r}_1, \ldots, \vec{r}_N, t) = c \,, \tag{4.12}$$

with c an arbitrary constant. Differential constraints of the form of Eq. 4.8 are said to be *integrable* and they are holonomic.

We shall consider only the simplest class of differential non-integrable constraints: the ones which depend linearly on the velocities $\dot{\vec{r}}_j$ of the N particles,

$$\sum_{j=1}^{N} \vec{l_j} \cdot \dot{\vec{r}}_j + D = 0 \,, \tag{4.13}$$

where the vectors $\vec{l_j}$ and the scalar D are functions of the positions $\vec{r}_j$ and of time t, $\vec{l_j} = \vec{l_j}(\vec{r}_1, \ldots, \vec{r}_N, t)$ and $D = D(\vec{r}_1, \ldots, \vec{r}_N, t)$. In Cartesian coordinates, eq. (4.13) reads,

$$\sum_{j=1}^{N} (A_j x_j + B_j y_j + C_j z_j) + D = 0\,. \tag{4.14}$$

As an example, consider an ice skate, which we represent by means of two particles connected by a rod of length L, as in Figure 3. The rod moves on the $x-y$ plane (the ice skating rink). Any point on the rod, and in particular the middle point, moves in the direction of the rod. So, in addition to the fixed length of the rod, which yields the constraint,

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 = L^2\,, \tag{4.15}$$

we must require that the position vector of the rod and the velocity vector of the middle point have the same direction, say at an angle α,

$$\tan\alpha = \frac{y_2 - y_1}{x_2 - x_1} = \frac{\dot{y}_1 + \dot{y}_2}{\dot{x}_1 + \dot{x}_2}\,. \tag{4.16}$$

Eq. 4.16 yields the differential constraint,

$$\frac{\dot{x}_1 + \dot{x}_2}{x_1 - x_2} = \frac{\dot{y}_1 + \dot{y}_2}{y_1 - y_2}\,, \tag{4.17}$$

which is non-integrable. Thus, the system defined by the constraints (4.15) and (4.17) is non-holonomic. Likewise, any rigid body whose velocity points in the direction of motion (like a car which is moving or parking without sliding) defines a non-holonomic system.

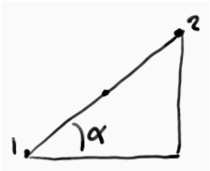


Figure 3: Ice skate represented as a rod of length L.

4.2 Possible and virtual displacements

Let us take a system of N particles P_j, with $j = 1, \ldots, N$, subject to d manifestly holonomic constraints,

$$f_i(\vec{r}_1, \ldots, \vec{r}_N, t) = 0\,, \qquad i = 1, \ldots, d\,, \tag{4.18}$$

and to g linear differential constraints,

$$\sum_{j=1}^{N} \vec{l}_{kj} \cdot \dot{\vec{r}}_j + D_k = 0\,, \qquad k = 1, \ldots, g\,. \tag{4.19}$$

By taking the total time derivative of the manifestly holonomic contraints, like in Eq. 4.8,

$$\sum_{j=1}^{N} \vec{\nabla}_j f_i \cdot \dot{\vec{r}}_j + \frac{\partial f_i}{\partial t} = 0\,, \qquad i = 1, \ldots, d\,, \tag{4.20}$$

we replace them by linear differential constraints as well. We see that the constraints of Eqs 4.19-4.20 limit the allowed velocities $\dot{\vec{r}}_j$ of the particles. The velocities $\dot{\vec{r}}_j$ which are allowed by the constraints Eqs 4.19-4.20 are called *possible*. At a given time t and positions $\vec{r}_j$ of the particles P_j, there are infinitely many sets of possible velocities, since we have not yet considered the other forces (not responsible for constraints) which are acting on the system. Once we do so, only one set of possible particle velocities is actually realised in the motion of the system.

Multiplying eqs. (4.19) and (4.20) by the time interval dt,

$$\begin{aligned} &\sum_{j=1}^{N} \vec{\nabla}_j f_i \cdot d\vec{r}_j + \frac{\partial f_i}{\partial t} dt = 0\,, \qquad i = 1, \ldots, d\,, \\ &\sum_{j=1}^{N} \vec{l}_{kj} \cdot d\vec{r}_j + D_k dt = 0\,, \qquad k = 1, \ldots, g\,, \end{aligned} \tag{4.21}$$

we write the linear differential constraints in terms of the *possible displacements*, $d\vec{r}_j = \dot{\vec{r}}_j dt$.

For two possible displacements,

$$d\vec{r}_j = \dot{\vec{r}}_j dt\,, \qquad d\vec{r'}_j = \dot{\vec{r'}}_j dt\,, \tag{4.22}$$

satisfying the constraints of Eq. 4.21, the difference

$$\delta\vec{r}_j = d\vec{r}_j - d\vec{r}'_j\,, \tag{4.23}$$

fulfils the constraints

$$\begin{aligned}
&\sum_{j=1}^{N} \vec{\nabla}_j f_i \cdot \delta\vec{r}_j = 0\,, \qquad i = 1, \ldots, d\,,\\
&\sum_{j=1}^{N} \vec{l}_{kj} \cdot \delta\vec{r}_j = 0\,, \qquad k = 1, \ldots, g\,,
\end{aligned} \tag{4.24}$$

In Cartesian coordinates the above can be written explicitly as:

$$\begin{aligned}
&\sum_{j=1}^{N} \left(\frac{\partial f_i}{\partial x_j}\delta x_j + \frac{\partial f_i}{\partial y_j}\delta y_j + \frac{\partial f_i}{\partial z_j}\delta z_j \right) = 0\,, \qquad i = 1, \ldots, d\,,\\
&\sum_{j=1}^{N} \left(A_{kj}x_j + B_{kj}y_j + C_{kj}z_j \right) = 0\,, \qquad k = 1, \ldots, g\,.
\end{aligned} \tag{4.25}$$

The displacements of Eq. 4.23 are called *virtual*. We can think of virtual displacements as the ones which take a possible configuration of the system at a time t to another one infinitely close to it, at the same time t. For stationary constraints, which have a vanishing time derivative, the set of virtual displacements coincides with the one of the possible displacements.

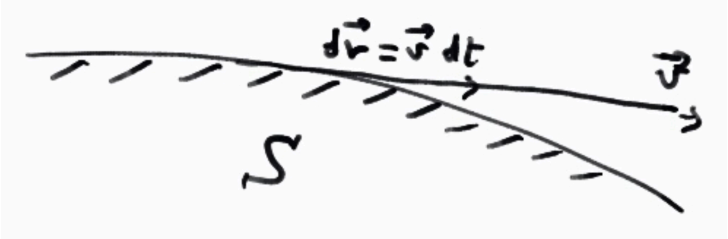


Figure 4: Particle in motion over a fixed surface S.

For an example of possible and virtual displacements consider a particle P which is in motion over a fixed surface S, as in Figure 4. The constraint is $f(\vec{r}) = 0$. Any vector $\vec{v}$ tangent to S is a possible velocity and the differential

elements $d\vec{r} = \vec{v}\, dt$ are possible displacements. The difference $\delta\vec{r} = d\vec{r} - d\overrightarrow{r'}$ of two tangent vectors is also a tangent vector. In this case, the sets of virtual and possible displacements coincide, as we expect since the constraint is stationary.

Consider now a particle P which is in motion on a moving surface S with velocity $\vec{u}$, as in Figure 5. We denote with w the possible velocities of the particle and we measure them with respect to the moving surface. Then the velocity of the particle is $\vec{v} = \overrightarrow{w} + \vec{u}$, where $\overrightarrow{w}$ is tangent to the surface S. Note that in general the velocity $\vec{v}$ is not tangent to S. Likewise, the possible displacement,

$$d\vec{r} = \vec{v}\, dt = \overrightarrow{w}\, dt + \vec{u}\, dt\,, \tag{4.26}$$

is a vector which is not necessarily tangent to S. Take, now, another possible displacement,

$$d\overrightarrow{r'} = \overrightarrow{w'}\, dt + \vec{u}\, dt\,, \tag{4.27}$$

and form a virtual the difference,

$$\delta\vec{r} = d\vec{r} - d\overrightarrow{r'} = (\overrightarrow{w} - \overrightarrow{w'})\, dt\,, \tag{4.28}$$

is a virtual displacement, and a vector tangent to S. Note that virtual and possible displacements do not coincide, since, as we know from eq. (4.6), the constraint of a moving surface is not stationary.

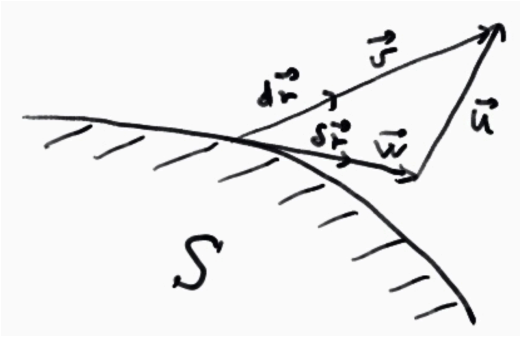


Figure 5: Particle in motion over a moving surface S.

4.3 Smooth constraints

From eq. (4.25), we know that for a system of N particles P_j, with $j = 1, \ldots, N$, we have $3N$ virtual displacements constrained by d geometric constraints (4.18) and g differential constraints (4.19). Thus there are $3N-d-g$ independent virtual displacements. We can say that our system of N particles has $3N - d - g$ *degrees of freedom*. The constraints, imply some restrictions in the acceleration of the particles. In fact, taking the total time derivative of eq. (4.8) for d constraints, we obtain

$$\begin{aligned}
&\sum_{j=1}^{N} \vec{\nabla}_j f_i \cdot \ddot{\vec{r}}_j + \sum_{j=1}^{N} \left(\frac{d}{dt} \vec{\nabla}_j f_i \right) \cdot \dot{\vec{r}}_j + \frac{d}{dt} \frac{\partial f_i}{\partial t} = 0 \,, \qquad i = 1, \ldots, d \,, \\
&\sum_{j=1}^{N} \vec{l}_{kj} \cdot \ddot{\vec{r}}_j + \sum_{j=1}^{N} \frac{d\vec{l}_{kj}}{dt} \cdot \dot{\vec{r}}_j + \frac{dD_k}{dt} = 0 \,, \qquad k = 1, \ldots, g \,.
\end{aligned} \tag{4.29}$$

Let us imagine for a moment the system without constraints. On the system are applied forces,

$$\vec{F}_j = \vec{F}_j(\vec{r}_i, \dot{\vec{r}}_i, t) \,, \qquad j = 1, \ldots, N \,, \tag{4.30}$$

which are known functions of the positions $\vec{r}_i$ and the velocities $\dot{\vec{r}}_i$ of the particles. Without constraints, the forces would induce accelerations:

$$\vec{F}_j = m_j \ddot{\vec{r}}_j \qquad j = 1, \ldots, N \,, \tag{4.31}$$

where m_j are the masses of the particles. However, if the system is constrained, the accelerations $\ddot{\vec{r}}_j$ are also due to additional forces $\vec{R}_j$, called *reaction forces*, which are responsible for the constraints. The accelerations consistent with Eq. 4.29 should satisfy:

$$m_j \ddot{\vec{r}}_j = \vec{F}_j + \vec{R}_j \,, \qquad j = 1, \ldots, N \,. \tag{4.32}$$

The general problem of the dynamics of a constrained system is as follows: *given the forces $\vec{F}_j$ and the initial positions $\vec{r}_{0,i}$ and velocities $\dot{\vec{r}}_{0,i}$ of the particles, we need to determine the trajectories of the particles and the reaction forces $\vec{R}_j$ of the constraints:*

$$\vec{r}_i(t) \,, \ \vec{R}_j \,, \qquad j = 1, \ldots, N \,, \tag{4.33}$$

Often (if $d+g<3N$), the known equations of motion and constraints

$$\begin{aligned} m_j\ddot{\vec{r}}_j &= \overrightarrow{F}_j+\overrightarrow{R}_j\,, \qquad & j=1,\ldots,N\,,\\ f_i(\vec{r}_j,t) &= 0\,, & i=1,\ldots,d\,,\\ \sum_{j=1}^{N}\vec{l}_{kj}\cdot\dot{\vec{r}}_j+D_k &= 0\,, & k=1,\ldots,g\,. \end{aligned} \tag{4.34}$$

do not suffice to determine all unknowns, leaving $n=3N-d-g$ degrees of freedom undetermined. In order to determine the motion of the system, we need additional $n=3N-d-g$ independent relations among the variables (4.33). We will find the necessary additional equations by examining the properties of the constraints and making an assumption for the reaction forces which generate them.

There is a large class of constraints, called *smooth*, for which the work of the reaction forces vanishes over the virtual displacements,

$$\sum_{j=1}^{N}\overrightarrow{R}_j\cdot\delta\vec{r}_j=0\,, \tag{4.35}$$

which in Cartesian coordinates is

$$\sum_{j=1}^{N}\left(R_{j,x}\delta x_j+R_{j,y}\delta y_j+R_{j,z}\delta z_j\right)=0\,. \tag{4.36}$$

Consider the following examples:

Example 4.1. Consider a particle P which moves on a fixed smooth surface S, as in Figure 6. As we have seen, the virtual and possible displacements coincide in that case; they are vectors tangent to S. Because S is smooth (there are no frictions), the reaction force R is orthogonal to S. Then $\overrightarrow{R}\cdot d\vec{r}=0$ and $\overrightarrow{R}\cdot\delta\vec{r}=0$. Therefore, the constraint is smooth.

Example 4.2. We now consider a pendulum, for which we have cast the constraint, Eq. 4.11, in the differential form $\vec{r}\cdot\dot{\vec{r}}=0$. Because the virtual and possible displacements coincide, it is also true that $\vec{r}\cdot\delta\vec{r}=0$, where the direction of the string $\hat{r}$ identifies also the direction of the reaction force $\overrightarrow{R}$. Thus, we have that $\overrightarrow{R}\cdot\delta\vec{r}=0$ and the constraint is therefore smooth.

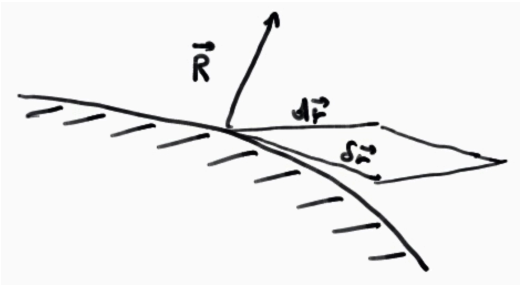


Figure 6: Reaction R on a fixed surface S.

Example 4.3. A point P moves on a smooth surface S, which moves with velocity $\vec{u}$, as in Figure 7. Because S is smooth, the reaction force R is orthogonal to S. As we have already seen, in this example the virtual displacement $\delta\vec{r}$ is tangent to S, while the possible displacement $d\vec{r}$ is generally not. Then $\vec{R} \cdot \delta\vec{r} = 0$, while $\vec{R} \cdot d\vec{r} \neq 0$. Albeit not stationary, this constraint is smooth.

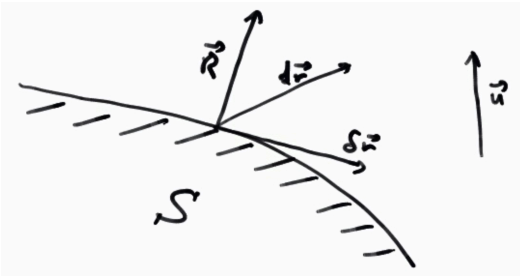


Figure 7: Reaction R on a moving surface S.

From Eq.(4.24) for the virtual displacements, we know that among the $3N$ virtual displacements there are $n = 3N - d - g$ independent ones, which define an n-dimensional vector space $\mathbb{R}^n$. We can write any displacement as a function of the n independent ones,

$$\delta\vec{r}_j = \sum_{k=1}^{n} \delta\alpha_k \vec{e}_{k,j}\,, \qquad j = 1, \ldots, N\,, \tag{4.37}$$

where $\delta\alpha_k$ are the independent displacements and $\vec{e}_{k,j}$ are unit vectors. Substituting it into Eq. (4.35) we obtain that

$$\sum_{j=1}^{N} \overrightarrow{R}_j \cdot \sum_{k=1}^{n} \delta\alpha_k \vec{e}_{k,j} = \sum_{k=1}^{n} \delta\alpha_k \left(\sum_{j=1}^{N} \overrightarrow{R}_j \cdot \vec{e}_{k,j} \right) = 0 \,, \tag{4.38}$$

which must be true for any virtual displacements $\delta\alpha_k$ which are independent. This requirement gives n additional equations,

$$\sum_{j=1}^{N} \overrightarrow{R}_j \vec{e}_{k,j} = 0 \,, \qquad k = 1, \ldots, n \,, \tag{4.39}$$

which are the missing n conditions which were needed in order to determine completely the system.

4.4 The general equation of the dynamics

In the previous section, we have established the equations of motion of a constrained system,

$$m_j \ddot{\vec{r}}_j = \overrightarrow{F}_j + \overrightarrow{R}_j \,, \qquad j = 1, \ldots, N \,, \tag{4.40}$$

with m_j the masses of the particles, $\overrightarrow{F}_j$ the forces applied on the system and $\overrightarrow{R}_j$ the reactions due to the constraints. In addition, we require that the constraints are smooth (4.35),

$$\sum_{j=1}^{N} \overrightarrow{R}_j \cdot \delta\vec{r}_j = 0 \,, \tag{4.41}$$

Substituting the reactions $\overrightarrow{R}_j$ as obtained from eq. (4.40) into eq. (4.41), we obtain

$$\sum_{j=1}^{N} \left(\overrightarrow{F}_j - m_j \ddot{\vec{r}}_j \right) \cdot \delta\vec{r}_j = 0 \,, \tag{4.42}$$

which is the *general equation of the dynamics*, which states that *the sum of the work of the effective forces $\overrightarrow{F}_j$ and the inertial forces $m_j \ddot{\vec{r}}_j$ vanishes for any virtual displacement.*

4.5 Independent generalised coordinates

In what follows, we will limit ourselves to a system of N particles, with d constraints,

$$f_i(\vec{r}_j, t) = 0\,, \qquad\qquad i = 1, \ldots, d\,. \tag{4.43}$$

written in a manifestly holonomic form.

We know from eq.(4.24) that among the $3N$ virtual displacements there are $n = 3N - d$ independent ones. Thus we can express the d equations (4.43) as functions of n independent coordinates, q_i with $i = 1, \ldots, n$, and of the time t. Also the positions of the particles can be taken as functions of the n independent coordinates,

$$\vec{r}_j = \vec{r}_j(q_1(t), \ldots, q_n(t), t)\,, \qquad\qquad j = 1, \ldots, N\,, \tag{4.44}$$

In the specific case of stationary contraints (4.3), the positions of the particles $\vec{r}_j$ do not depend explicitly on the time t but only through $q_i(t)$.

The essential information about the degrees of freedom being reduced from $3N$ to n is encoded into the n independent coordinates of eq. (4.44), thus if we substitute eq. (4.44) into eq. (4.43), these become identities. The n independent coordinates q_i are called *independent generalised coordinates*. Consider the following examples:

Example 4.4. A pendulum is moving on the $x - z$ plane and it is attached to a string of length R. the motion of the pendulum spans an angle ϕ: the pendulum has only one degree of freedom. The stationary holonomic constraint is given by the equation,

$$x^2 + z^2 - R^2 = 0\,. \tag{4.45}$$

If we take ϕ as the independent coordinate, then Eqs. 4.44 become

$$\begin{aligned} x &= R\cos\phi\,, \\ z &= R\sin\phi\,. \end{aligned} \tag{4.46}$$

If we substitute the Eqs. 4.46 into Eq. 4.45, we obtain the identity,

$$\cos^2\phi + \sin^2\phi - 1 = 0\,. \tag{4.47}$$

Example 4.5. A particle is constrained to move on a sphere of radius R: it can move with two degrees of freedom. As independent coordinates, we can

take the longitude ϕ and the latitude θ. The holonomic stationary constraint is given by the equation,

$$x^2 + y^2 + z^2 - R^2 = 0\,. \tag{4.48}$$

In terms of the ϕ and θ angles, Eqs. 4.44 become

$$\begin{aligned} x &= R\cos\phi\cos\theta\,, \\ y &= R\sin\phi\cos\theta\,, \\ z &= R\sin\theta\,. \end{aligned} \tag{4.49}$$

If we substitute the Eqs. (4.50) into the constraint of Eq. (4.48), we obtain the identity,

$$(\cos^2\phi + \sin^2\phi)\cos^2\theta + \sin^2\theta - 1 = 0\,. \tag{4.50}$$

4.6 Euler-Lagrange equations for systems with smooth constraints and potential forces

We may now express the $3N$ virtual displacements as a function of the $n = 3N - d$ independent coordinates Eq. 4.44,

$$\delta\vec{r}_j = \sum_{i=1}^{n} \frac{\partial\vec{r}_j}{\partial q_i}\delta q_i\,, \qquad\qquad j = 1,\ldots,N\,, \tag{4.51}$$

and substitute it into the general equation of the dynamics (Eq. 4.42),

$$\sum_{i=1}^{n}\sum_{j=1}^{N}\left(\vec{F}_j - m_j\ddot{\vec{r}}_j\right)\cdot\frac{\partial\vec{r}_j}{\partial q_i}\delta q_i = 0\,, \tag{4.52}$$

where we have inverted the order of the sums. Because the coordinates q_i are independent, so are their virtual displacements δq_i. Thus, their coefficients must identically vanish, yielding:

$$\sum_{j=1}^{N}\left(\vec{F}_j - m_j\ddot{\vec{r}}_j\right)\cdot\frac{\partial\vec{r}_j}{\partial q_i} = 0\,, \qquad\qquad i = 1,\ldots,n\,. \tag{4.53}$$

The first term of Eq. 4.53,

$$\sum_{j=1}^{N}\vec{F}_j\cdot\frac{\partial\vec{r}_j}{\partial q_i} \equiv Q_i\,, \qquad\qquad i = 1,\ldots,n\,. \tag{4.54}$$

is called the *generalised force*. The generalised force has the dimension of a force only if the independent coordinates have the dimension of a length. Now let us analyse the second term of Eqs. 4.53,

$$\sum_{j=1}^{N} m_j \ddot{\vec{r}}_j \cdot \frac{\partial \vec{r}_j}{\partial q_i}\,, \qquad\qquad i = 1, \ldots, n\,, \tag{4.55}$$

which we rewrite as

$$\sum_{j=1}^{N} m_j \frac{d\dot{\vec{r}}_j}{dt} \cdot \frac{\partial \vec{r}_j}{\partial q_i} = \frac{d}{dt}\left(\sum_{j=1}^{N} m_j \dot{\vec{r}}_j \cdot \frac{\partial \vec{r}_j}{\partial q_i}\right) - \sum_{j=1}^{N} m_j \dot{\vec{r}}_j \cdot \frac{d}{dt}\frac{\partial \vec{r}_j}{\partial q_i}\,, \tag{4.56}$$

Next, we derive a couple of useful identities. We start from the velocities in terms of independent coordinates,

$$\dot{\vec{r}}_j = \sum_{i=1}^{n} \frac{\partial \vec{r}_j}{\partial q_i}\dot{q}_i + \frac{\partial \vec{r}_j}{\partial t}\,, \tag{4.57}$$

from which we immediately obtain that

$$\frac{\partial \dot{\vec{r}}_j}{\partial \dot{q}_i} = \frac{\partial \vec{r}_j}{\partial q_i}\,. \tag{4.58}$$

Then we derive eq. (4.57) with respect to the independent coordinates,

$$\frac{\partial \dot{\vec{r}}_j}{\partial q_i} = \sum_{k=1}^{n} \frac{\partial^2 \vec{r}_j}{\partial q_i \partial q_k}\dot{q}_k + \frac{\partial^2 \vec{r}_j}{\partial q_i \partial t} = \frac{d}{dt}\frac{\partial \vec{r}_j}{\partial q_i}\,. \tag{4.59}$$

We substitute the identities (4.58) and (4.59) into eq. (4.56) and we obtain,

$$\sum_{j=1}^{N} m_j \ddot{\vec{r}}_j \cdot \frac{\partial \vec{r}_j}{\partial q_i} = \frac{d}{dt}\left(\sum_{j=1}^{N} m_j \dot{\vec{r}}_j \cdot \frac{\partial \dot{\vec{r}}_j}{\partial \dot{q}_i}\right) - \sum_{j=1}^{N} m_j \dot{\vec{r}}_j \cdot \frac{\partial \dot{\vec{r}}_j}{\partial q_i}\,. \tag{4.60}$$

Considering that the kinetic energy of the holonomic system of N particles is,

$$T = \frac{1}{2}\sum_{j=1}^{N} m_j \dot{r}_j^2\,, \tag{4.61}$$

we can rewrite eq. (4.60) as

$$\sum_{j=1}^{N} m_j \ddot{\vec{r}}_j \cdot \frac{\partial \vec{r}_j}{\partial q_i} = \frac{d}{dt}\frac{\partial T}{\partial \dot{q}_i} - \frac{\partial T}{\partial q_i}\,, \qquad i = 1, \ldots, n\,. \tag{4.62}$$

Substituting it into eq. (4.53) and remembering the definition of generalised force (4.54), we obtain:

$$\frac{d}{dt}\frac{\partial T}{\partial \dot{q}_i} - \frac{\partial T}{\partial q_i} = Q_i\,, \qquad i = 1, \ldots, n\,. \tag{4.63}$$

These equations form a system of n second order differential equations in n unknown variables q_i. Therefore, the motion of the holonomic system with n degrees of freedom is determined fully if the values of q_i and $\dot{q}_i$ are given at an initial time t_0.

Note that the reaction forces $\vec{R}_j$ do not appear explicitly in Eqs. 4.63. They can be obtained by solving Eqs. 4.63 first, and then using Eq. 4.40.

Now, let us suppose that some forces can be derived from a potential energy $U = U(q_i, t)$,

$$\vec{F}_j = -\frac{\partial U}{\partial \vec{r}_j}\,, \tag{4.64}$$

such that the generalised forces become

$$Q_i = \sum_{j=1}^{N} \vec{F}_j \cdot \frac{\partial \vec{r}_j}{\partial q_i} = -\sum_{j=1}^{N} \frac{\partial U}{\partial \vec{r}_j} \cdot \frac{\partial \vec{r}_j}{\partial q_i} = -\frac{\partial U}{\partial q_i}\,, \qquad i = 1, \ldots, n\,, \tag{4.65}$$

and that in addition there are non-potential generalised forces,

$$\tilde{Q}_i = \tilde{Q}_i(q_k, \dot{q}_k, t)\,. \tag{4.66}$$

Eqs. 4.63) become

$$\frac{d}{dt}\frac{\partial T}{\partial \dot{q}_i} - \frac{\partial T}{\partial q_i} = -\frac{\partial U}{\partial q_i} + \tilde{Q}_i\,, \qquad i = 1, \ldots, n\,. \tag{4.67}$$

Recalling the defintion of the Lagrangian,

$$L = T - U\,, \tag{4.68}$$

we can cast Eq. 4.67 in the form:

$$\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{q}_i} - \frac{\partial \mathcal{L}}{\partial q_i} = \tilde{Q}_i \,, \qquad i = 1, \ldots, n \,. \tag{4.69}$$

In the absence of non-potential forces equations 4.67) give the Euler-Lagrange equations:

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0 \,, \qquad i = 1, \ldots, n \,. \tag{4.70}$$

We have therefore proved that Hamilton's principle of least action, which yields the same Euler-Lagrange equations, should also hold for constrained systems with smooth holonomic constraints and potential forces.

5 Application of Euler-Lagrange equations: Simple pendulum

We have seen that we can obtain the equations of motion for a system with holonomic constraints from a Lagrangian $\mathcal{L}$ by requiring that the action integral is ninimum. This gives rise to Euler-Lagrange equations (the equations of motion):

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \dot{q}_i} - \frac{\partial \mathcal{L}}{\partial q_i} = 0 \tag{5.1}$$

In the above, all generalised coordinates q_i are independent, and we have first used the holonomic constraints to eliminate additional variables which are not independent.

Eqs. (5.1) are of course equivalent to (and have also been derived from) Newton's law. It is instructive to take a simple physics problem and use it as a testbed to compare the two methods. Consider a simple pendulum: According to Newton's law,

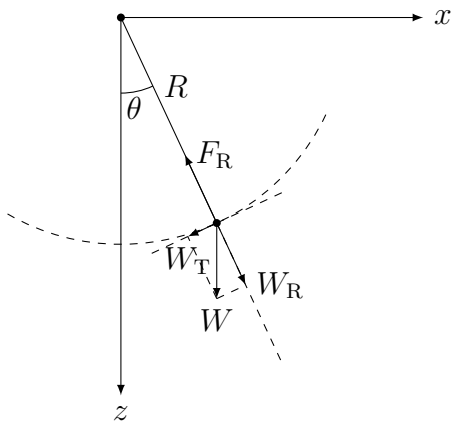


Figure 8: A simple pendulum consists of a particle of mass m which is attached via a (massless) rope of a certain length R to a fixed point.

$$m\ddot{\vec{r}} = \overrightarrow{F_R} + \overrightarrow{W}. \tag{5.2}$$

where $\overrightarrow{W}$ is the weight force and $\overrightarrow{F_R}$ is the reaction force from the rope. The acceleration is:

$$\ddot{\vec{r}} = \ddot{x}\hat{x} + \ddot{y}\hat{y} + \ddot{z}\hat{z} \tag{5.3}$$

We will need to change to polar coordinates, with the following change of

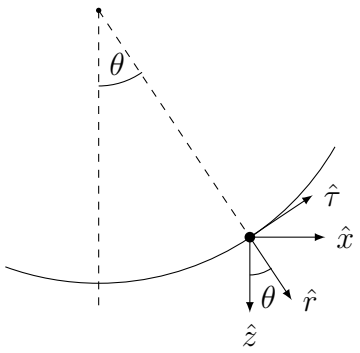


Figure 9: The unit vectors definig the Cartesian and polar coordinates.

basis for the unit vectors:

$$(\hat{x}, \hat{y}, \hat{z}) \longrightarrow (\hat{r}, \hat{y}, \hat{\tau}) \tag{5.4}$$

where $\hat{r}$ is the radial unit vector and $\hat{\tau}$ the tangential unit vector to the motion of the pendulum. We have:

$$\left.\begin{aligned} \hat{z} &= (\hat{z}\cdot\hat{r})\hat{r} + (\hat{z}\cdot\hat{\tau})\hat{\tau} \\ \hat{x} &= (\hat{x}\cdot\hat{r})\hat{r} + (\hat{x}\cdot\hat{z})\hat{z} \end{aligned}\right\} \Longrightarrow \begin{cases} \hat{z} = \cos\theta\,\hat{r} - \sin\theta\,\hat{\tau} \\ \hat{x} = \sin\theta\,\hat{r} + \cos\theta\,\hat{z} \end{cases} \tag{5.5}$$

or, in matrix form:

$$\begin{pmatrix} \hat{z} \\ \hat{x} \\ \hat{y} \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \hat{r} \\ \hat{\tau} \\ \hat{y} \end{pmatrix} \tag{5.6}$$

As expected, the change from Cartesian to polar coordinates is a rotation by an angle θ around the $\hat{y}$-axis. Expressing the acceleration in the polar basis and using that the length of the pendulum is fixed,

$$r = R$$

we have:

$$\ddot{\vec{r}} = \hat{r}(-R\dot{\theta}^2) + \hat{z}(R\ddot{\theta}). \tag{5.7}$$

The force acting on the pendulum is decomposed as:

$$\overrightarrow{F} = \overrightarrow{W} + \overrightarrow{F}_R = -(mg\sin\theta)\hat{z} + (mg\cos\theta - F_R)\hat{r}.$$

Newton's law gives then two equations: The equation of motion determining the evolution of the angle θ:

$$mR\ddot{\theta} = -mg\sin\theta \rightsquigarrow \boxed{\ddot{\theta} + \frac{g}{R}\sin\theta = 0} \tag{5.8}$$

and an equation which determines the reaction force:

$$-mR\dot{\theta}^2 = mg\cos\theta - F_R \rightsquigarrow \boxed{F_R = m(R\dot{\theta}^2 + g\cos\theta)} \tag{5.9}$$

Let us now use the Euler-Lagrange method. The Lagrangian is

$$L = T - V = \frac{1}{2}m\dot{\vec{r}}^2 - mg(z_0 - z).$$

We have:

$$\vec{r} = \hat{z}r\cos\theta + \hat{x}r\sin\theta + \hat{y}y$$

We can use the constraints

$$r = R = \text{const.}, y = \text{const.}.$$

Then,

$$\dot{\vec{r}} = R(\hat{z}\sin\theta + \hat{x}\cos\theta)\dot{\theta} \rightsquigarrow \dot{\vec{r}}^2 = R^2\dot{\theta}^2$$

Thus,

$$L = \frac{1}{2}mR^2\dot{\theta}^2 + mgR\cos\theta \underbrace{\qquad -mgz_0}_{\text{unimportant constant}}$$

The equation of motion is:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial\dot{\theta}} - \frac{\partial L}{\partial\theta} = 0 \rightsquigarrow mR^2\ddot{\theta} + mgR\sin\theta = 0 \rightsquigarrow \boxed{\ddot{\theta} + \frac{g}{R}\sin\theta = 0}$$

which is the same result as with Newton's law (Eqs. 5.8).

Notice that in neither of the two methods we said anything about how to solve the equation of motion. In this sence, which of the two methods we use in order to derive (Eq 5.8) is a matter of taste. Nevertheless, we can make a few comparisons:

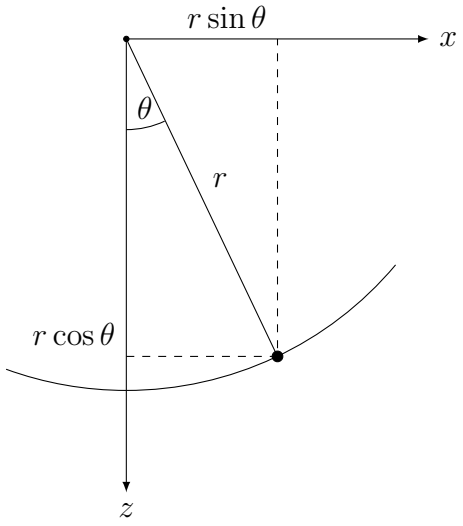


Figure 10: Visualization of the constraints.

- The Euler-Lagrange method requires scalar quantities. No need to perform vector rotations.
- The Euler-Lagrange does not make an explicit reference to the reaction forces. This is an advantage if we are not interested in them, but a disadvantage if we care about them. For example, an engineer building a pendulum would like to know how much the reaction force is, in order to use a sufficiently strong rope.

6 Minimization with Lagrange multipliers

The mathematical problem that we are facing in classical mechanics is one of minimization of a functional (the action). This problem becomes more complicated due to contstraints which make the variables of the functional (generalized coordinates) not independent. We have seen that for simple constraints (holonomic) it may be possible to eliminate the dependent variables. When this is not possible, we may consider the more sophisticated technique of Lagrange multipliers. We will recall the salient features of this method for the minimization of functions with constraints first and then we will see how to apply it for functionals and the action specifically.

6.1 Minimization of a multivariate function with Lagrange multipliers

Consider a function

$$u = f(x_1, \ldots, x_N)$$

where $x_1, \ldots, x_N$ are all independent variables. We are interested in finding the extrema (minima or maxima) of the function. These are given by the condition that the total differential of the function vanishes:

$$\mathrm{d}u = 0 \Rightarrow \sum_{i=1}^{N} \frac{\partial f}{\partial x_i} \mathrm{d}x_i = 0.$$

Since we have taken the variables x_i to be independent, the coefficients of $\mathrm{d}x_i$ must all vanish. Thus, we have that an extremum occurs if

$$\frac{\partial f}{\partial x_i} = 0 \quad \forall i = 1, \ldots, N.$$

Let us now assume that the variables x_i are not all independent due to, for example, a constraint:

$$\phi(x_1, \ldots, x_N) = 0.$$

The extrema of the function are given in this case too by the condition

$$\mathrm{d}u = 0,$$

but unlike earlier one of the variables, let's say x_N, is not independent any more. A straightforward way to find the minimum would be to solve the constraint

$$\phi(x_1, \ldots, x_N) = 0 \rightsquigarrow x_N = x_N(x_1, \ldots, x_{N-1})$$

and eliminate this variable from the function:

$$u = f(x_1, \ldots, x_N(x_1, \ldots, x_{N-1})) = g(x_1, \ldots, x_{N-1})$$

Then, we can minimize $g(x_1, \ldots, x_{N-1})$ as before, since the remaining variables are all independent. However, there is an alternative way where we do not need to solve the constraint explicitly. We start from

$$\begin{aligned} \mathrm{d}u = 0 \rightsquigarrow & \sum_{i=1}^{N-1} \left(\frac{\partial f}{\partial x_i} + \frac{\partial f}{\partial x_N} \frac{\partial x_N}{\partial x_i} \right) \mathrm{d}x_i = 0 \\ \rightsquigarrow & \sum_{i=1}^{N-1} \frac{\partial f}{\partial x_i} \mathrm{d}x_i + \frac{\partial f}{\partial x_N} \sum_{i=1}^{N-1} \frac{\partial x_N}{\partial x_i} \mathrm{d}x_i = 0. \end{aligned} \tag{6.1}$$

The sum in the last term is:

$$\sum_{i=1}^{N-1} \frac{\partial x_N}{\partial x_i} \mathrm{d}x_i = \mathrm{d}x_N.$$

We therefore find again the same condition

$$\mathrm{d}u = 0 \rightsquigarrow \frac{\partial f}{\partial x_i} \mathrm{d}x_i + \cdots + \frac{\partial f}{\partial x_{N-1}} \mathrm{d}x_{N-1} + \frac{\partial f}{\partial x_N} \mathrm{d}x_N = 0, \tag{6.2}$$

although now the $\mathrm{d}x_i$'s are not all independent and we cannot demand that their coefficients vanish independently. Consider now the constraint:

$$\begin{aligned} \phi(x_1, \ldots, x_N) = 0 \rightsquigarrow & \ \mathrm{d}\phi(x_1, \ldots, x_N) = 0 \\ \Rightarrow & \frac{\partial \phi}{\partial x_1} \mathrm{d}x_1 + \cdots + \frac{\partial \phi}{\partial x_N} \mathrm{d}x_N = 0 \end{aligned} \tag{6.3}$$

We can multiply (6.3) with a constant λ, the so called *Lagrange multiplier*, and subtract it from (6.2):

$$\boxed{\sum_{i=1}^{N} \left[\frac{\partial f}{\partial x_i} - \lambda \frac{\partial f}{\partial x_i} \right] \mathrm{d}x_i = 0}$$

It can be proven (Analysis course) that we can use the extra parameter λ in order to make all terms in the brackets to vanish,

$$\frac{\partial f}{\partial x_i} - \lambda \frac{\partial \phi}{\partial x_i} = 0 \quad i = 1, \ldots, N.$$

Solving the system of equations above will determine the minimum.

Therefore, the procedure to find the extremum in the presence of a constraint is identical to the one without contraints, replacing the original function

$$f(x_1, \ldots, x_N)$$

with a function which includes the constraint

$$f_\lambda(x_1, \ldots, x_N) \equiv f(x_1, \ldots, x_N) - \lambda \phi(x_1, \ldots, x_N).$$

Let's see how this works in a couple of examples.

Example 6.1. Find the minimum distance of a straight line from the origin. We will solve this problem in two ways:

1. Eliminating dependent variables.
   The square of the distance of a point (x, y) from the origin is:

   $$f(x, y) \equiv d(x, y)^2 = x^2 + y^2.$$

   The constraint demands that the point belongs to a line

   $$y = ax + b \quad a, b \in \mathbb{R}.$$

   Eliminating the y-variable, we have:

   $$f(x, y) = f(x, y(x)) = x^2 + (ax + b)^2 \equiv g(x).$$

   We now have one independent variable x. We can minimize, by demanding:

   $$\frac{\partial g(x)}{\partial x} = 0 \rightsquigarrow 2x + 2a(ax + b) = 0 \Rightarrow x = \frac{-ab}{1 + a^2}$$

   Then

   $$y = a\left(\frac{-ab}{1 + a^2}\right) + b = \frac{b}{1 + a^2}.$$

   The mininum distance of the line to the origin is therefore at the point

   $$(x, y) = \left(\frac{-ab}{1 + a^2}, \frac{b}{1 + a^2}\right)$$

2. Using Lagrange multipliers.
   We first define the function

$$f_\lambda(x,y) = x^2 + y^2 - \lambda(y - ax - b)$$

   which contains the square of the distance of the pont (x,y) as well as a Lagrange multiplier with the constraint that the point must belong to a line $y - ax - b = 0$. We require that the partial derivatives of this function with respect to all variables must vanish:

$$\left.\begin{array}{l} \frac{\partial}{\partial x} f_\lambda(x,y) = 0 \\ \frac{\partial}{\partial y} f_\lambda(x,y) = 0 \end{array}\right\} \Rightarrow \left.\begin{array}{r} 2x + \lambda a = 0 \\ 2y - \lambda = 0 \end{array}\right\} \Rightarrow \begin{array}{l} x = -\frac{\lambda}{2} a \\ y = \frac{\lambda}{2} \end{array}$$

   We can now use the constraint:

$$y - ax - b = 0 \Rightarrow \frac{\lambda}{2}(1 + a^2) = b \Rightarrow \lambda = \frac{2b}{1 + a^2}.$$

   Thus the minimum is at

$$(x,y) = \left(\frac{-a}{1 + a^2} b, \frac{b}{1 + a^2}\right)$$

   as expected.

You may wonder what is the advantage of Lagrange multipliers since in our first example it was trivial to solve the problem by eliminating the dependent variable. However, it occurs in other problems to be easier (or the only possibility) to solve for a Lagrange multiplier instead. We demonstrate this in the following more complicated example.

Example 6.2. Find the minimum and maximum distance from the origin on an ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

We introduce a Lagrange mutiplier for the constraint and minimize the function

$$f_\lambda(x,y) = x^2 + y^2 - \lambda\left(\frac{x^2}{a^2} + \frac{y^2}{b^2} - 1\right)$$

by demanding that

$$\frac{\partial f_\lambda}{\partial x} = 0 \rightsquigarrow x\left(1 - \frac{\lambda}{a^2}\right) = 0$$
$$\frac{\partial f_\lambda}{\partial y} = 0 \rightsquigarrow y\left(1 - \frac{\lambda}{b^2}\right) = 0$$

and

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

The solutions of the three equation are:

$$x = 0, \lambda = b^2, y = \pm b$$
$$\text{or}$$
$$y = 0, \lambda = a^2, x = \pm a$$

The extrema are then

$$(x, y) = \{(0, \pm b), (\pm a, 0)\}.$$

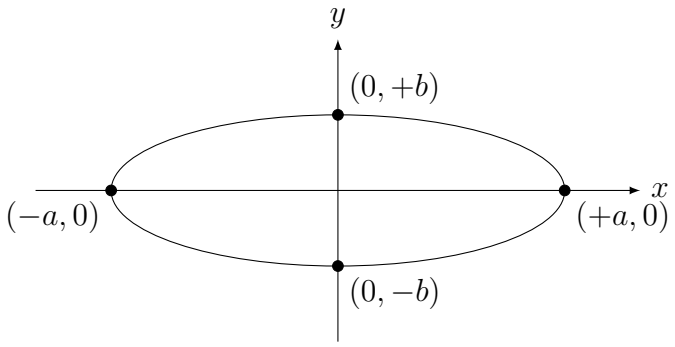


Figure 11: Ellipse with semimajor a and semiminor b.

6.2 Minimizing the action with Lagrange multipliers

We have seen that the principle of minimum action (Hamilton's principle) can be applied to physics problems with constraints. It is natural to use the

method of Lagrange multipliers in order to impose these constraints. We can achieve this as follows. Consider a system with $\{q_i\}$ independent coordinates described by the action

$$S = \int \mathrm{d}t\, L[q_i, \dot{q}_i].$$

Minimizing the action gives

$$\delta S = 0 \rightsquigarrow \int \mathrm{d}t \sum_i \delta q_i \left[\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} \right] = 0.$$

Since δq_i's are all independent of each other the coefficients must vanish, yielding the familiar Euler-Lagrange equations

$$\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0.$$

In the presence of a constraint

$$f(q_i, \dot{q}_i) = 0$$

the δq_i's are not independent independent anymore. From the constraint, we deduce that the following integral vanishes

$$\int \mathrm{d}t\, \lambda f(q_i, \dot{q}_i) = 0$$

and thus

$$\delta S_\lambda = 0 \qquad \text{with} \qquad S_\lambda = \int \mathrm{d}t\, L_\lambda(q_i, \dot{q}_i)$$

and

$$L_\lambda(q_i, \dot{q}_i) = \underbrace{L(q_i, \dot{q}_i)}_{\substack{\text{original} \\ \text{Lagrangian}}} - \underbrace{\lambda}_{\substack{\uparrow \\ \text{Lagrange} \\ \text{multiplier}}} \underbrace{f(q_i, \dot{q}_i)}_{\text{constraint}}$$

gives us

$$\delta S_\lambda = 0 \rightsquigarrow \int \mathrm{d}t \sum_i \delta q_i \left[\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L_\lambda}{\partial \dot{q}_i} - \frac{\partial L_\lambda}{\partial q_i} \right] = 0.$$

The presence of the λ parameter allows us to take all the coefficients of δq_i to vanish

$$\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L_\lambda}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0$$

besides the fact that the δq_i's are not independent.

Example 6.3. Let us solve the simple pendulum problem once again by using Lagrange multipliers this time. The Lagrangian with a Lagrange multiplier for the constraint is

$$L_\lambda = T - V - \lambda(r - \underbrace{R}_{\text{constant}})$$

with

$$T = \frac{1}{2}m\dot{\vec{r}}^2 = \frac{1}{2}m\dot{r}^2 + \frac{1}{2}mr^2\dot{\theta}^2 \qquad \text{and} \qquad V = mgr\cos\theta.$$

The corresponding Euler-Lagrange equations are:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L_\lambda}{\partial \dot{r}} - \frac{\partial L_\lambda}{\partial r} = 0 \rightsquigarrow m\ddot{r} + mr\dot{\theta}^2 + mg\cos\theta - \lambda = 0$$
$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L_\lambda}{\partial \dot{\theta}} - \frac{\partial L_\lambda}{\partial \theta} = 0 \rightsquigarrow mr^2\ddot{\theta} + mg\sin\theta = 0.$$

Using in addition the constraint

$$\boxed{r = R = \text{const.}}$$

we have

$$\boxed{\ddot{\theta} + \frac{g}{L}\sin\theta = 0} \tag{6.4}$$

$$\boxed{mg\cos\theta + mR\dot{\theta}^2 = \lambda} \tag{6.5}$$

(6.4) is the equation of motion for the pendulum.
(6.5) is very interesting. In the l.h.s. we recognize the reaction force F_R through the rope. We see then that the Lagrange multiplier has some physical meaning in terms of the reaction force. This is not an accident. On one hand, we have:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L_\lambda}{\partial \dot{q}_i} - \frac{\partial L_\lambda}{\partial q_i} = 0$$
$$\rightsquigarrow \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = \lambda\left(\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial f}{\partial \dot{q}_i} - \frac{\partial f}{\partial q_i}\right).$$

On the other hand, we have proven in general that

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = \underbrace{\tilde{Q}_i}_{\substack{\uparrow \\ \text{generalized} \\ \text{forces}}} = \sum_{j=1}^{N} \underbrace{\vec{F}_{R,j}}_{\substack{\uparrow \\ \text{non-potential} \\ \text{forces}}} \cdot \frac{\partial \vec{r}_j}{\partial q_i}.$$

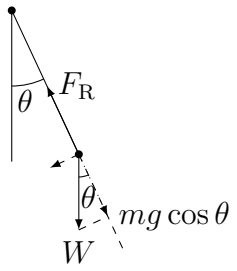


Figure 12: The simple pendulum with reaction force F_R.

Thus,

$$\lambda\left(\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial f}{\partial \dot{q}_i}-\frac{\partial f}{\partial q_i}\right)=\tilde{Q}_i.$$

In our case $f=r-R$ and the l.h.s. is

$$\lambda\left(\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial (r-R)}{\partial \dot{r}}-\frac{\partial (r-R)}{\partial r}\right)=-\lambda=\tilde{Q}_r.$$

7 Conservation Laws

Observing the time evolution of many physical systems, we often find that certain quantities remain invariant at all times. A classical example is the energy of mechanical systems without friction. The Lagrangian formalism will allow us to gain a deeper understanding for the conservation of such quantities. We will realize that conservation and symmetry are interconnected.

7.1 Conservation of energy

Let us take a look at systems with a Lagrangian

$$L[q_i(t), \dot{q}_i(t), t] = L[q_i(t), \dot{q}_i(t)].$$

The lack of explicit time dependence of the Lagrangian means that there is no special time. Whenever we let the system to evolve, it will be so in exactly the same way irrespective of what is the starting time t_0.

The total time derivative of the Lagrangian is:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}L &= \sum_i \left[\frac{\partial L}{\partial q_i}\frac{\partial q_i}{\partial t} + \frac{\partial L}{\partial \dot{q}_i}\frac{\partial \dot{q}_i}{\partial t} + \underbrace{\frac{\partial L}{\partial t}}_{=0}\right] \\ &= \sum_i \left[\frac{\partial L}{\partial q_i}\dot{q}_i + \frac{\partial L}{\partial \dot{q}_i}\ddot{q}_i\right].\end{aligned}$$

Let us now recall Euler-Lagrange equations

$$\frac{\partial L}{\partial q_i} = \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_i}.$$

Then we have:

$$\begin{aligned}\frac{\mathrm{d}L}{\mathrm{d}t} &= \sum_i \left[\left(\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_i}\right)\dot{q}_i + \frac{\partial L}{\partial \dot{q}_i}\ddot{q}_i\right] \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\sum_i \frac{\partial L}{\partial \dot{q}_i}\dot{q}_i\end{aligned}$$

This implies that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\sum_i \frac{\partial L}{\partial \dot{q}_i}\dot{q}_i - L\right] = 0$$

and therefore

$$\boxed{\sum_i \frac{\partial L}{\partial \dot{q}_i}\dot{q}_i - L \equiv E = \text{const.}}. \tag{7.1}$$

We have found a conserved quantity, which we call energy. The reason for its invariance is the fact that the behaviour of the system does not change if we let it evolve at a different starting time.

We will now justify why the expression (7.1) is the familiar energy. Consider a particle with potential energy $U(\vec{r})$. The Lagrangian is

$$\begin{aligned} L &= T - U(\vec{r}) \\ &= \frac{1}{2}m\dot{\vec{r}}^2 - U(\vec{r}). \end{aligned}$$

According to (7.1), the energy is

$$\begin{aligned} E &= \frac{\partial L}{\partial \dot{x}}\dot{x} + \frac{\partial L}{\partial \dot{y}}\dot{y} + \frac{\partial L}{\partial \dot{z}}\dot{z} - L \\ &= m\dot{x}^2 + m\dot{y}^2 + m\dot{z}^2 - \frac{1}{2}m\dot{\vec{r}}^2 + U(\vec{r}) \\ &= m\dot{\vec{r}}^2 - \frac{1}{2}m\dot{\vec{r}}^2 + U(\vec{r}) \\ &= \frac{1}{2}\dot{\vec{r}}^2 + U(\vec{r}) = \underbrace{T}_{\text{kinetic}} + \underbrace{U}_{\text{potential}}. \end{aligned}$$

This is indeed the sum od the kinetic and potential energy.

7.2 Conservation of momentum

Let us now trace the origin of the conservation of momentum. We will find that momentum is conserved for physical systems for which we can shift the position of all of their constituents without changing their physical behavior. For example, if a physical system can be studied in an experiment in Zürich but also in an independent experiment in New York, where the two experiments are anticipated to find the same results then the total momentum of the system is conserved. Formally, we assume that performing a translation for the positions of all particles

$$\vec{r}_i \to \vec{r}_i + \vec{\epsilon}$$

should leave the equations of motion invariant. Let us see how the Lagrangian changes if $\vec{\epsilon}$ is taken to be infinitesimal.

$$\begin{aligned}\Delta L &= L[\vec{r}_i + \vec{\epsilon}, \dot{\vec{r}}_i + \dot{\vec{\epsilon}}, t] - L[\vec{r}_i, \dot{\vec{r}}_i, t] \\ &= \sum_{i=1}^{N} \left[\frac{\partial L}{\partial r_{i,x}} \epsilon_x + \frac{\partial L}{\partial r_{i,y}} \epsilon_y + \frac{\partial L}{\partial r_{i,z}} \epsilon_z \right] \\ &= \sum_{i=1}^{N} \left(\overrightarrow{\nabla}_i L \cdot \vec{\epsilon} \right).\end{aligned}$$

The equations of motion are invariant by this displacement if the change in the Lagrangian is zero:

$$\Delta L = 0 \rightsquigarrow \boxed{\vec{\epsilon} \cdot \sum_{i=1}^{N} \overrightarrow{\nabla}_i L = 0}.$$

The direction of $\vec{\epsilon}$ is arbitrary. Choosing for example

$$\vec{\epsilon} = (1,0,0) \qquad \text{or} \qquad \vec{\epsilon} = (0,1,0) \qquad \text{or} \qquad \vec{\epsilon} = (0,0,1),$$

we obtain

$$\sum_{i=1}^{N} \frac{\partial L}{\partial x_i} = 0 \qquad \text{and} \qquad \sum_{i=1}^{N} \frac{\partial L}{\partial y_i} = 0 \qquad \text{and} \qquad \sum_{i=1}^{N} \frac{\partial L}{\partial x_i} = 0 \qquad (7.2)$$

where $(x_i, y_i, z_i) \equiv \vec{r}_i$ the position vector of the i-th particle. For the Lagrangian of the form $L = T - U$ Eq. (7.2) gives

$$\overrightarrow{\nabla}_i U = 0.$$

However, $F_i = -\overrightarrow{\nabla} U$ is the force acting on the i-th particle. Thus we conclude that

$$\boxed{\sum_{i=1}^{N} F_i = 0},$$

where N is the number of particles in the system. In the special case of two only particles, $N = 2$, we have that

$$\overrightarrow{F}_1 = -\overrightarrow{F}_2$$

which is Newton's "action and reaction" law.

Let us now recall the Euler-Lagrange equations:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{x}_i} &= \frac{\partial L}{\partial x_i}\\ \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{y}_i} &= \frac{\partial L}{\partial y_i}\\ \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{z}_i} &= \frac{\partial L}{\partial z_i}.\end{aligned}$$

Then, we conclude that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_i \frac{\partial L}{\partial \dot{x}_i}\right) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_i \frac{\partial L}{\partial \dot{y}_i}\right) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_i \frac{\partial L}{\partial \dot{z}_i}\right) = 0.$$

Therefore the quantities

$$\begin{aligned}P_x &\equiv \sum_i \frac{\partial L}{\partial \dot{x}_i} = \text{const.}\\ P_y &\equiv \sum_i \frac{\partial L}{\partial \dot{y}_i} = \text{const.}\\ P_z &\equiv \sum_i \frac{\partial L}{\partial \dot{z}_i} = \text{const.}.\end{aligned}$$

are constant and the vector

$$\vec{P} = (P_x, P_y, P_z)$$

is conserved. We can associate $\vec{P}$ with the total momentum of the system.

Exercise 7.1. For a Lagrangian

$$L = \sum_i \frac{1}{2} m_i \dot{\vec{r}}_i^{\,2} - V(\vec{r}_1, \ldots, \vec{r}_n) \tag{7.3}$$

verify that

$$\frac{\partial L}{\partial \dot{x}_i} = m_i \dot{x}_i, \quad \frac{\partial L}{\partial \dot{y}_i} = m_i \dot{y}_i, \quad \frac{\partial L}{\partial \dot{z}_i} = m_i \dot{z}_i. \tag{7.4}$$

We have therefore found that

$$\vec{P} = \sum_{i=1}^{N} m_i \dot{\vec{r}}_i = \sum_{i=1}^{N} \vec{P}_i$$

is indeed the total momentum of all particles in the system and that it is conserved.

7.3 Angular momentum

Consider a system of particles with masses m_i which are at positions $\vec{r}_i(t)$ at a time t. Now assume that we can rotate all vectors $\vec{r}_i$ by an angle $\delta\phi$ around a certain axis $\hat{n}$. Let us define the vector

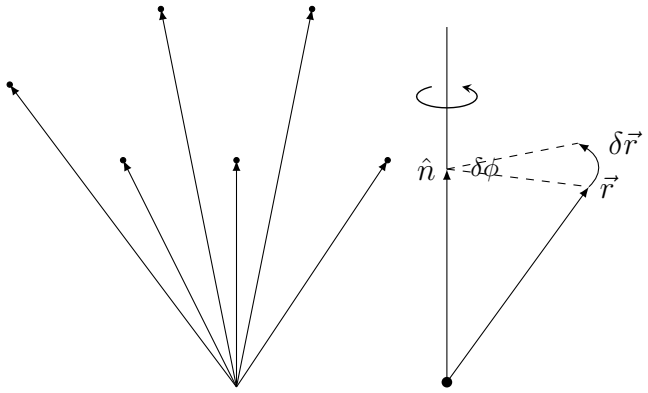


Figure 13: Rotation of all position vectors $\vec{r}_i$ in a system of particles around an axis $\hat{n}$ by an angle $\delta\phi$.

$$\delta\vec{\phi} = \hat{n}\delta\phi.$$

A position vector $\vec{r}$ after a rotation becomes

$$\vec{r} \to \vec{r} + \delta\vec{r} + \mathcal{O}(\delta\phi^2)$$

with $\delta\vec{r} = \delta\vec{\phi} \times \vec{r}$. For illustration, we can verify this for a rotation around the $\hat{z}$-axis:

$$\begin{aligned}
\vec{r} \to \vec{r}' &= \mathbf{R} \cdot \vec{r} \\
&= \begin{pmatrix} \cos\delta\phi & -\sin\delta\phi & 0 \\ \sin\delta\phi & \cos\delta\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \\
&\approx \begin{pmatrix} 1 & -\delta\phi & 0 \\ \delta\phi & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \mathcal{O}(\delta\phi^2) \\
&= \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \delta\phi \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \\
&= \vec{r} + \delta\phi \begin{pmatrix} -y \\ x \\ 0 \end{pmatrix}.
\end{aligned}$$

Consider

$$\begin{aligned}
(\hat{n} \times \vec{r})_i &= \sum_{j,k} \epsilon_{ijk} \hat{n}_j r_k \qquad \hat{n} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \\
&= \epsilon_{i3k} r_k
\end{aligned}$$

Thus

$$\begin{aligned}
(\hat{n} \times \vec{r})_1 &= \epsilon_{132} r_2 = -1y = -y \\
(\hat{n} \times \vec{r})_2 &= \epsilon_{231} r_1 = +1x = x \\
(\hat{n} \times \vec{r})_3 &= \epsilon_{133} r_k = 0
\end{aligned}$$

and

$$\begin{aligned}
\vec{r} \to \vec{r}' &= \vec{r} + \delta\vec{r} \\
&= \vec{r} + \delta\vec{\phi} \times \vec{r} + \mathcal{O}(\delta\phi^2).
\end{aligned}$$

Now we can find what are the conserved quantities associated with rotation symmetry by repeating the same steps as for space translations in the

previous section. Under a rotation transformation the Lagrangian changes by

$$L \to L' = L + \delta L$$

with

$$\begin{aligned}
\delta L &\equiv L[\vec{r} + \delta\vec{\phi} \times \vec{r}, \dot{\vec{r}} + \delta\vec{\phi} \times \dot{\vec{r}}] - L[\vec{r}, \dot{\vec{r}}] \\
&= \sum_i \left[\frac{\partial L}{\partial x_i}(\delta r_i)_x + \frac{\partial L}{\partial y_i}(\delta r_i)_y + \frac{\partial L}{\partial z_i}(\delta r_i)_z + \frac{\partial L}{\partial \dot{x}_i}(\delta \dot{r}_i)_x + \frac{\partial L}{\partial \dot{y}_i}(\delta \dot{r}_i)_y + \frac{\partial L}{\partial \dot{z}_i}(\delta \dot{r}_i)_z \right] \\
&= \sum_i \left[\dot{P}_{x,i}(\delta r_i)_x + \dot{P}_{y,i}(\delta r_i)_y + \dot{P}_{z,i}(\delta r_i)_z + P_{x,i}(\delta \dot{r}_i)_x + P_{y,i}(\delta \dot{r}_i)_y + P_{z,i}(\delta \dot{r}_i)_z \right] \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\delta\vec{\phi} \cdot \left[\sum_i \vec{r}_i \times \overrightarrow{P}_i \right] \\
&= \delta\vec{\phi} \cdot \frac{\mathrm{d}}{\mathrm{d}t} \sum_i \vec{r}_i \times \overrightarrow{P}_i.
\end{aligned}$$

If this change of the Lagrangian is zero, then the rotation has no effect on the physical system. This implies that

$$\delta L = 0 \rightsquigarrow \delta\vec{\phi} \cdot \frac{\mathrm{d}}{\mathrm{d}t} \sum_i \vec{r}_i \times \overrightarrow{P}_i = 0.$$

Since $\delta\vec{\phi}$ is an arbitrary vector (we can choose the rotation axis at wish) we must have:

$$\frac{\mathrm{d}}{\mathrm{d}t} \sum_i \vec{r}_i \times \overrightarrow{P}_i = 0 \Rightarrow \boxed{\overrightarrow{M} = \sum_i \vec{r}_i \times \overrightarrow{P}_i = \text{const.}}.$$

We have therefore found that the total angular momentum is conserved.

7.4 Summary of conservation laws & symmetry

We have realized that symmetry plays an important role in conservation laws. In the following table we summarize out findings, for typical conserved quantities and their symmetry origin.

Symmetry	*Conserved quantity*
Time translation	Energy: $E = T + U$
Space translation	Momentum: $\overrightarrow{P} = \sum_i \overrightarrow{\nabla}_i L = m_i \dot{\vec{r}}_i$
Rotations	Angular momentum: $\overrightarrow{M} = \sum_i \vec{r}_i \times \overrightarrow{P_i}$

These symmetries are profound and we expect them to hold for many systems.

7.5 Conservation of generalized momenta

In analogy to "Cartesian momenta" if the Lagrangian is expressed in terms of generalized coordinates, we can define the generalized mometa as

$$P_i = \frac{\partial L}{\partial \dot{q}_i}$$

and generalized force

$$\dot{P}_i = \frac{\partial L}{\partial q_i}.$$

For a Lagrangian which depends only on $\dot{q}_i$ but not on q_i, we have

$$\tilde{Q}_i \equiv \dot{P}_i \equiv \frac{\partial L}{\partial q_i} = 0$$

(the generalized force vanishes). Then, the generalized momentum P_i is conserved:

$$0 = \frac{\partial L}{\partial q_i} = \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_i} = \frac{\mathrm{d}}{\mathrm{d}t}\tilde{Q}_i \rightsquigarrow \boxed{P_i = \text{const.}}.$$

7.6 Noether's Theorem

If the equations of motion are invariant under symmetry transformations, then there are quantities which are conserved. This is a general statement known as *Noether's theorem*.

Assume that an infinitesimal transformation of the generalized coordinates

$$q_i(t) \to q_i(t) + \delta q_i(t)$$

is a symmetry transformation, i.e. it let's the equations of motion invariant. The change that this transformation induces to the Lagrangian is

$$\Delta L = \sum_i \left[\frac{\partial L}{\partial q_i} \Delta q_i + \frac{\partial L}{\partial \dot{q}_i} \Delta \dot{q}_i \right].$$

Using the Euler-Lagrange equations we substitute

$$\frac{\partial L}{\partial q_i} = \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L}{\partial \dot{q}_i}.$$

Then

$$\begin{aligned} \Delta L &= \sum_i \left[\left(\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L}{\partial \dot{q}_i} \right) \Delta q_i + \frac{\partial L}{\partial \dot{q}_i} \frac{\mathrm{d}}{\mathrm{d}t} (\Delta q_i) \right] \\ &= \frac{\mathrm{d}}{\mathrm{d}t} \sum_i \frac{\partial L}{\partial \dot{q}_i} \Delta q_i \\ \Rightarrow \boxed{\Delta L = \frac{\mathrm{d}}{\mathrm{d}t} \sum_i \frac{\partial L}{\partial \dot{q}_i} \Delta q_i}. \end{aligned}$$

If $\Delta L = 0$, then the Lagrangian is invariant and the action is also invariant, leading to the same equations of motion. But we should be more careful! More generally, the action is invariant if the Lagrangian changes under the symmetry transformation by a total time derivative:

$$\Delta L = \frac{\mathrm{d}F}{\mathrm{d}t} \rightsquigarrow \frac{\mathrm{d}}{\mathrm{d}t} \sum_i \frac{\partial L}{\partial \dot{q}_i} \Delta q_i = \frac{\mathrm{d}F}{\mathrm{d}t} \rightsquigarrow \frac{\mathrm{d}}{\mathrm{d}t} \left(\sum_i \frac{\partial L}{\partial \dot{q}_i} \Delta q_i - F \right) = 0.$$

Then we have that the quantity

$$\boxed{J = \sum_i \frac{\partial L}{\partial q_i} \Delta q_i - F}$$

is a conserved quantity.

Example 7.1. Consider a particle in a homogeneous gravitational field:

$$L = \frac{1}{2} m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - mgz.$$

Since the field along the $z-$axis, we expect that translations

$$z \to z + a$$

to be a symmetry. With this transformation, the Lagrangian changes by

$$L \to L - mga \Rightarrow \Delta L = -mga = -a\frac{\mathrm{d}}{\mathrm{d}t}(mgt).$$

We have been able to cast the change as a total derivative. Then, according to Noether's theorem,

$$\begin{aligned}
\frac{\partial L}{\partial \dot{z}}\Delta z &= \Delta L \\
\frac{\partial L}{\partial \dot{z}}\Delta z &= -a\frac{\mathrm{d}}{\mathrm{d}t}(mgt) \\
\Rightarrow \frac{\mathrm{d}}{\mathrm{d}t}(m\dot{z}a) &= -a\frac{\mathrm{d}}{\mathrm{d}t}(mgt) \\
\Rightarrow \frac{\mathrm{d}}{\mathrm{d}t}(\dot{z} + gt) &= 0 \\
\Rightarrow \boxed{\dot{z} + gt = \text{const.}}&.
\end{aligned}$$

We have then found that there is a conserved quantity. We can verify that this is correct directly from the equation of motion:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{z}} - \frac{\partial L}{\partial z} = 0 \Rightarrow m\ddot{z} + mg = 0 \overset{\int \mathrm{d}t}{\Rightarrow} \boxed{\dot{z} + gt = \text{const.}}.$$

For the other possible translations along the x and y directions induce no change in the Lagrangian:

$$x \to x + a \Rightarrow \Delta L = 0.$$

These are symmetry transformations. According to Noether's theorem:

$$\frac{\partial L}{\partial \dot{x}}\Delta x = \text{const.} \Rightarrow m\dot{x}a = \text{const.} \Rightarrow \dot{x} = \text{const.}.$$

The equations of motion give

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial x} - \frac{\partial L}{\partial \dot{x}} = 0 \rightsquigarrow m\ddot{x} = 0 \rightsquigarrow \boxed{\dot{x} = \text{const.}}.$$

and $y \to y + a \rightsquigarrow \dot{y} = \text{const.}$, confirming the result of Noether's theorem.

7.7 Noether's theorem & time symmetries[2]

In the action integral,

$$S = \int_{t_i}^{t_f} dt L[q_i(t), \dot{q}_i(t)] \tag{7.5}$$

time is an integration variable, while the generalised coordinates q_i and their corresponding velocities $\dot{q}_i$ are unintegrated parameters. In Eq. 7.6 we examined a symmetry transformation which acts on the coordinates (integral parameters). However, as we have seen in Sec 7.1 a time symmetry (time translation), which affects the integration variable t of the action integral, leads to a conservation law. Our proof of Noether's theorem in the previous section does not cover this case. We will see here that Noether's theorem can be extended for symmetries of the integration variable as well. This proof will be particularly useful in the future, when we study continuum systems and fields. For convenience, in this Section we will use Einstein's summation convention: $A_i B_i \equiv \sum_i A_i B_i$.

Consider a symmetry transformation

$$t \to t' = t'(t) = t + \delta t$$

which leaves the action invariant:

$$\begin{aligned}
\int_{t_i}^{t_f} \mathrm{d}t L[q_i(t), \dot{q}_i(t)] &= \int_{t_i'}^{t_f'} \mathrm{d}t' \, L[q_i'(t'), \dot{q}_i'(t')] \\
&= \int_{t_i}^{t_f} \mathrm{d}t \left(\frac{\mathrm{d}t'}{\mathrm{d}t} \right) L[q'(t'), \dot{q}_i'(t')] \\
&= \int_{t_i}^{t_f} \mathrm{d}t \left[1 + \frac{\mathrm{d}}{\mathrm{d}t} \delta t \right] L[q_i(t) + \Delta q_i'(t'), \dot{q}_i(t') + \Delta \dot{q}_i(t')].
\end{aligned} \tag{7.6}$$

We have assumed that the coordinates also change by a certain amount when we perform this transformation:

$$q_i(t) \to q_i'(t') = q_i(t) + \Delta q_i'(t').$$

We can now write

$$L[q(t) + \Delta q_i'(t'), \dot{q}_i(t) + \Delta \dot{q}_i'(t')] = L[q_i(t), \dot{q}_i(t)] + \frac{\partial L}{\partial q_i} \Delta q_i'(t') + \frac{\partial L}{\partial \dot{q}_i} \Delta \dot{q}_i'(t').$$

[2]This section can be ommitted in a first reading. It is not part of an exam

Then, (7.6) gives that

$$\int_{t_i}^{t_f} \mathrm{d}t \left[L\frac{\mathrm{d}\delta t}{\mathrm{d}t} + \frac{\partial L}{\partial q_i}\Delta q'(t') + \frac{\partial L}{\partial \dot{q}_i}\Delta \dot{q}_i'(t') + \frac{\partial L}{\partial t}\delta t \right] = 0. \tag{7.7}$$

We decompose the change in the coordinates as follows:

$$\begin{aligned}
\Delta q_i'(t') &\equiv q_i'(t') - q_i(t) \\
&= \underbrace{q_i'(t') - q_i(t')}_{\delta_* q_i} + q_i(t') - q_i(t) \\
&= \delta_* q_i + q_i(t) + \frac{\mathrm{d}}{\mathrm{d}t}q_i(t)\delta t - q_i(t) \\
&= \delta_* q_i + \dot{q}_i \delta t \\
\rightsquigarrow \Delta q_i'(t') &= q_i'(t') - q_i(t') + \dot{q}_i\delta t
\end{aligned}$$

where we separate the change due to changing the argument of the coordinate function from one time instant t to another t' and an instantaneous change of the function itself q_i to q_i' at the same time t'. Similarly,

$$\begin{aligned}
\Delta \dot{q}_i'(t') &= \dot{q}_i'(t') - \dot{q}_i(t') + \dot{q}_i(t') - \dot{q}_i(t) \\
\rightsquigarrow \Delta \dot{q}_i' &= \frac{\mathrm{d}}{\mathrm{d}t}(\delta_* q_i) + \ddot{q}_i \delta t.
\end{aligned}$$

Substituting into (7.7), we obtain:

$$\begin{aligned}
0 &= \int_{t_i}^{t_f} \mathrm{d}t \left[L\frac{\mathrm{d}}{\mathrm{d}t}\delta t + \frac{\partial L}{\partial q_i}(\delta_* q_i + \dot{q}\delta t) + \frac{\partial L}{\partial \dot{q}_i}\left(\frac{\mathrm{d}}{\mathrm{d}t}\delta_* q_i + \ddot{q}_i \delta t \right) \right] \\
&= \int_{t_i}^{t_f} \mathrm{d}t \left\{ \left[L\frac{\mathrm{d}}{\mathrm{d}t}\delta t + \left(\frac{\partial L}{\partial q_i}\dot{q}_i + \frac{\partial L}{\partial \dot{q}_i}\ddot{q} \right)\delta t \right] + \left(\frac{\partial L}{\partial q_i}\delta_* q_i + \frac{\partial L}{\partial \dot{q}_i}\frac{\mathrm{d}}{\mathrm{d}t}\delta_* q_i \right) \right\}.
\end{aligned}$$

The term in the square bracket is:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[(\delta t) L[q_i, \dot{q}_i] \right] = L\frac{\mathrm{d}}{\mathrm{d}t}(\delta t) + \delta t \left(\frac{\partial L}{\partial q_i}\dot{q}_i + \frac{\partial L}{\partial \dot{q}_i}\ddot{q}_i \right).$$

Thus

$$\begin{aligned}
0 &= \int \mathrm{d}t \left[\frac{\mathrm{d}}{\mathrm{d}t}(L\delta t) + \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_i}\delta_* q_i + \frac{\partial L}{\partial \dot{q}_i}\frac{\mathrm{d}}{\mathrm{d}t}\delta_* q_i \right] \\
&= \int \mathrm{d}t \frac{\mathrm{d}}{\mathrm{d}t}\left[L\delta t + \frac{\partial L}{\partial \dot{q}_i}\delta_* q_i \right].
\end{aligned}$$

The above should hold for arbitrary boundaries. Thus we must have that the integrand must vanish:

$$\Rightarrow L\delta t + \frac{\partial L}{\partial \dot{q}_i}\delta_* q_i = \text{const.}.$$

We can now write

$$\delta_* q_i = \Delta q_i - \dot{q}_i \delta t \qquad \text{with} \qquad \Delta q_i \equiv q_i'(t') - q_i(t).$$

We then have

$$\boxed{\left(L - \frac{\partial L}{\partial \dot{q}_i}\dot{q}_i\right)\delta t + \frac{\partial L}{\partial \dot{q}_i}\Delta q_i = \text{const.}}.$$

For the special case (but also very common) of $\Delta q = 0$, we recover the result of Section 7.1 that the energy is conserved:

$$\left(L - \frac{\partial L}{\partial \dot{q}_i}\dot{q}_i\right) = E = \text{constant.} \tag{7.8}$$

7.8 Continuous Symmetry Transformations

As we have seen, symmetry plays an important role for the conservation of physical quantities. In this section, we will review some of the mathematical properties of continuous symmetry transformations.

Consider a set of generalized coordinates

$$\begin{pmatrix} q_1 \\ \vdots \\ q_N \end{pmatrix}$$

describing a system with a Lagrangian

$$L[q_i(t), \dot{q}_i(t), t].$$

Assume that the system is symmetric under a symmetry transformation

$$q_i \to q_i' = T_{ij}(\theta_a) q_j \quad \text{where } a = 1, \ldots, N \quad \text{and} \quad i, j = 1, \ldots, M.$$

We take the parameters θ_a to be continuous. For example, if a system is described by two Cartesian coordinates on a plane

$$\begin{pmatrix} x \\ y \end{pmatrix}$$

and is symmetric under rotations

$$\begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} x \\ y \end{pmatrix}' = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

the angle θ is a the continuous parameter which parameterizes the transformations. The symmetry transformations form a *Lie group* and the product

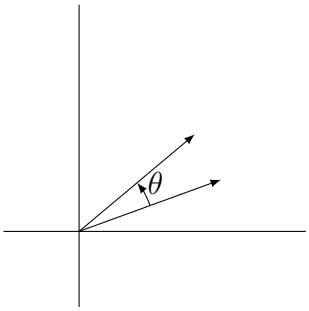


Figure 14: Rotation symmetry.

of two such transformations is also a symmetry transformation belonging to the group:

$$T(\theta_1^a)\, T(\theta_2^a) = T(\theta_3^a) \quad \text{where} \quad \theta_3^a = f^a(\theta_1^a, \theta_2^a). \tag{7.9}$$

We arrange so that the unit element of the symmetry group corresponds to all values of the parameters being zero:

$$T(\theta^a = 0) = 1. \tag{7.10}$$

Then,

$$\begin{aligned} T(f^a(0,\theta^a)) &= T(0)T(\theta^a) \\ &= 1\, T(\theta^a) = T(\theta^a) \\ \rightsquigarrow f^a(0,\theta^a) &= \theta^a. \end{aligned} \tag{7.11}$$

Similarly

$$f^a(\theta^a, 0) = \theta^a. \tag{7.12}$$

Let us now perform a Taylor expansion around $\theta^a = 0$:

$$\theta_1^a = f^a(\theta_1^a, \cancelto{0}{\theta_2^a}) = f^a(0,0) + \frac{\partial f^a}{\partial \theta_2^b}\theta_1^b + \cancel{\frac{\partial f^a}{\partial \theta_2^b}\theta_2^b} + \frac{1}{2}\frac{\partial^2 f^a}{\partial\theta_1^b \partial\theta_1^c}\theta_1^b\theta_1^c \cancelto{0}{} + \cancel{\frac{1}{2}\frac{\partial^2 f^a}{\partial\theta_2^b \partial\theta_2^c}\theta_2^b\theta_2^c} + \cancel{\frac{\partial^2 f^a}{\partial\theta_1^b \partial\theta_1^c}\theta_1^b\theta_2^c} + \dots$$

Impressing (7.11) and (7.12), we have

$$f^a(\theta_1^a, \theta_2^a) = \theta_1^a + \theta_2^a + \sum_{b,c} f_{bc}^a \theta_1^b \theta_2^c + \dots$$

For small θ^a parameters, we can expand:

$$T(\theta^a) = 1 + \mathrm{i}\theta^a t_a + \frac{1}{2}\sum_{b,c} \theta^b\theta^c t_{bc} + \dots$$

where 1, t_a, t_{bc} are matrices of the same dimensionality as $T(\theta^a)$. In addition

$$t_{bc} = t_{cb} \quad \text{(symmetric)}.$$

Making a Taylor expansion of (7.9), we have:

$$\left(1 + \mathrm{i}\theta_1^a t_a + \frac{1}{2}\theta_1^b\theta_2^c t_{bc}\right)\left(1 + \mathrm{i}\theta_2^d t_d + \frac{1}{2}\theta_1^d\theta_2^e t_{de}\right) = 1 + \mathrm{i}\left(\theta_1^a + \theta_2^a + f_{bc}^a\theta_1^b\theta_2^c + \dots\right)t_a + \frac{1}{2}\left(\theta_1^b + \theta_2^b + \dots\right)\left(\theta_1^c + \theta_2^c + \dots\right)t_{bc} + \dots$$

Matching the terms of the expansion, we have:

$$t_{bc} = -t_b t_c - \mathrm{i} f_{bc}^a t_a.$$

From the symmetry of $t_{bc} = t_{cb}$:

$$t_b t_c + \mathrm{i} f_{bc}^a t_a = t_c t_b + \mathrm{i} f_{bc}^a t_a$$
$$\rightsquigarrow t_b t_c - t_c t_b = \mathrm{i}(f_{cb}^a - f_{bc}^a)t_a$$

For reasons to become clear later, the matrices t_a are called the *generators* of the symmetry transformations. We also define the commutator of two such generators as:

$$[t_a, t_b] \equiv t_a t_b - t_b t_a.$$

The combinations

$$C^a_{bc} \equiv f^a_{cb} - f^a_{bc}$$

are called the *structure constants* and they are antisymmetric:

$$\boxed{C^a_{bc} = -C^a_{cb}}.$$

The equation

$$[t_a, t_b] = \mathrm{i} C^a_{bc} t_a \tag{7.13}$$

defines the *Lie algebra* of the group. Notice that

$$[t_a, [t_b, t_c]] + [t_b, [t_c, t_a]] + [t_c, [t_a, t_b]] = 0. \tag{7.14}$$

This is the so called *Jacobi identity* and it is satisfied by the generators. This gives an identity for the structure constants. Using the Lie algebra, we have:

$$\begin{aligned} &[t_a, \mathrm{i} f^d_{bc} t_d] + [t_b, \mathrm{i} f^d_{ca} t_d] + [t_c, \mathrm{i} f^d_{ab} t_d] = 0 \\ \rightsquigarrow &\boxed{C^e_{ad} C^d_{bc} + C^e_{bd} C^d_{ca} + C^e_{cd} C^d_{ab} = 0}. \end{aligned}$$

Let us now define:

$$\tilde{t}^b_{ac} = \mathrm{i} C^c_{ab}.$$

Then the above equation becomes:

$$[\tilde{t}^a, \tilde{t}^b] = \mathrm{i} C^d_{ab} \tilde{t}^d, \quad \text{i.e.,}$$

the structure constant furnish another *representation* of the Lie algebra. We call this the adjoint representation.

It is now time to justify why we call the t_a matrices generators. Using the product rule of Eq. 7.9 an infinite amount of times, we can write for a symmetry transformation with parameters θ^a

$$\begin{aligned} T(\theta^a) &= \lim_{N\to\infty} T\left(\frac{\theta^a}{N}\right)^N \quad \text{(product rule)} \\ &= \lim_{N\to\infty} \left(1 + \mathrm{i}\frac{\theta^a}{N} t_a\right)^N \\ &= \exp(\mathrm{i}\theta^a t_a). \end{aligned}$$

In other words, the generators t_a, together with the product rule are sufficient to produce any ẗẗlarge" transformation from infinitesimal ones by means of a simple exponentiation:

$$T(\theta^a) = \exp(\mathrm{i}\theta^a t_a). \tag{7.15}$$

Example 7.2. For rotations on a plane, we have:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \underbrace{\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}}_{T(\theta)}.$$

For a small θ angle:

$$\begin{aligned} T(\theta) &\approx \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \theta \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} + \mathcal{O}(\theta^2) \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \mathrm{i}\theta \begin{pmatrix} 0 & \mathrm{i} \\ -\mathrm{i} & 0 \end{pmatrix} + \mathcal{O}(\theta^2). \end{aligned}$$

The generator is

$$t \equiv \begin{pmatrix} 0 & \mathrm{i} \\ -\mathrm{i} & 0 \end{pmatrix}$$

(we have only one). Notice that

$$t^2 = \begin{pmatrix} 0 & \mathrm{i} \\ -\mathrm{i} & 0 \end{pmatrix} \begin{pmatrix} 0 & \mathrm{i} \\ -\mathrm{i} & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1.$$

Thus,

$$\begin{aligned} \exp(\mathrm{i}\theta t) &= \sum_{n=0}^{\infty} \underbrace{\mathrm{i}^n \theta^n t^n}_{n!} \\ &= \sum_{n=0}^{\infty} \frac{\mathrm{i}^{2n}\theta^{2n}(t^2)^n}{(2n)!} + \sum_{n=0}^{\infty} \frac{\mathrm{i}^{2n+1}\theta^{2n+1}t^{2n+1}}{(2n+1)!} \\ &= \sum_{n=0}^{\infty} \frac{(i\theta)^{2n}}{(2n)!} + \mathrm{i}\sum_{n=0}^{\infty} \frac{(\mathrm{i}\theta)^{2n+1}}{(2n+1)!}. \end{aligned}$$

But

$$\begin{aligned} \mathrm{e}^{i\theta} &= \cos\theta + \mathrm{i}\sin\theta \\ &= \sum_{n=0}^{\infty} \frac{(\mathrm{i}\theta)^{2n}}{(2n)!} + \sum_{n=0}^{\infty} \frac{(\mathrm{i}\theta)^{2n+1}}{(2n+1)!}. \end{aligned}$$

Thus:

$$\begin{aligned} \exp(\mathrm{i}\theta t) &= \cos\theta\, 1 + \mathrm{i}t\sin\theta \\ &= \begin{pmatrix} \cos\theta & 0 \\ 0 & \cos\theta \end{pmatrix} + \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \sin\theta \\ &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \checkmark. \end{aligned}$$

Exercise 7.2. Find the generators and structure constants for rotations in 3 dimensions (SO(3) group).

8 Motion in one dimension

The knowledge of conservation laws such as energy conservation can be helpful for solving the equations of motion of a system and determining the trajectories of its particles. We can demonstrate this in the simplest possible example, an one-particle system which is specified by one only generalized coordinate.

Let's consider a Lagrangian of the form

$$L = \frac{1}{2}a(q)\dot{q}^2 - V(q)$$

with q a generalized coordinate. This Lagrangian does not depend explicitly on time. Thus, the energy of the system is conserved:

$$E = \frac{\partial L}{\partial \dot{q}}\dot{q} - L = \frac{1}{2}a(q)\dot{q}^2 + V(q) = \text{const.}.$$

It now becomes easier to solve the equations of motion, without even having written the Euler-Lagrange equation down. From energy conservation, we have

$$\left(\frac{\mathrm{d}q}{\mathrm{d}t}\right)^2 = \frac{2(E - V(q))}{a(q)}$$

$$\rightsquigarrow \int \mathrm{d}t = \int \mathrm{d}q \frac{a(q)^{\frac{1}{2}}}{\sqrt{2(E - V(q))}}$$

$$\Rightarrow \boxed{t = \int \mathrm{d}q \sqrt{\frac{a(q)}{2(E - V(q))}} + \text{const.}}.$$

For a particle moving in one dimension

$$q = x, \quad a(q) = m, \qquad L = \frac{1}{2}m\dot{x}^2 - V(x)$$

we have

$$t - t_0 = \sqrt{\frac{m}{2}} \int \frac{\mathrm{d}x}{\sqrt{(E - V(x))}}.$$

The kinetic energy of the particle must always be a positive definite quantity

$$\frac{1}{2}m\dot{x}^2 = E - V(x) \geq 0.$$

This condition may limit the allowed interval in which the particle can move. The solution of the inequality determines the range of the motion. If the motion is bounded by two points, as it happens for a potential as illustrated in the figure,the motion is oscillatory. Then the period of the oscillation is

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

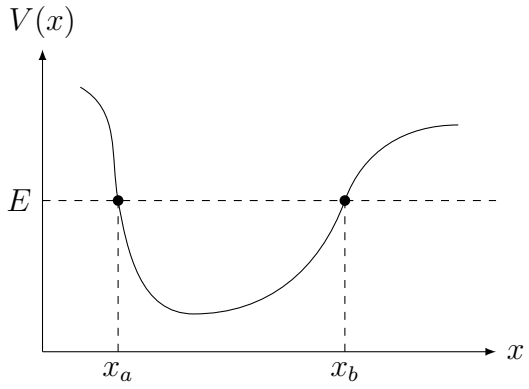


Figure 15: Potential $V(x)$ with $E = V(x_a) = V(x_b)$.

UNEDITED

9 Two-body problem

In this chapter we will discuss the interaction of two particles with masses m_1 and m_2, assuming that the force which exerts on each other depends only on their relative distance. The Lagrangian which describes such a system is

$$L = \frac{1}{2}m_1\dot{\vec{r}}_1^2 + \frac{1}{2}m_2\dot{\vec{r}}_2^2 - V(|\vec{r}_1 - \vec{r}_2|).$$

We need to find the trajectories $\vec{r}_1 = \vec{r}_1(t)$ and $\vec{r}_2 = \vec{r}_2(t)$ of the two particles. We shall see that the symmetry of the problem simplifies this task.
A first simplification we can make is to use a clever frame of reference. This is the, so called, *center of mass* frame in which the total momentum of all particles in a system is zero. To determine this frame, we first find the average position of the masses:

$$\overrightarrow{R} = \frac{\sum_i m_i\vec{r}_i}{\sum_i m_i} = \frac{m_1\vec{r}_1 + m_2\vec{r}_2}{m_1 + m_2}. \tag{9.1}$$

We also define the distance vector of the two particles. Assuming that $m_2 > m_1$, we define conventionally:

$$\vec{r} = \vec{r}_1 - \vec{r}_2. \tag{9.2}$$

We can express the position vectors of the two particles in terms of $\vec{r}$ and $\overrightarrow{R}$. After a little algebra, we find:

$$\vec{r}_1 = \overrightarrow{R} + \frac{m_2}{m_1 + m_2}\vec{r} \tag{9.3}$$

$$\vec{r}_2 = \overrightarrow{R} - \frac{m_1}{m_1 + m_2}\vec{r}. \tag{9.4}$$

In terms of the $(\vec{r}, \overrightarrow{R})$ variables, the Lagrangian becomes:

$$L = \frac{1}{2}(m_1 + m_2)\dot{\overrightarrow{R}}^2 + \frac{1}{2}\mu\dot{\vec{r}}^2 - V(r), \tag{9.5}$$

where $r \equiv |\vec{r}|$ and μ is the *reduced mass*

$$\mu = \frac{m_1 m_2}{m_1 + m_2}. \tag{9.6}$$

The nomenclature is more transparent if we cast (9.6) in the form

$$\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2} \tag{9.7}$$

from which we can easily see that

$$\frac{1}{\mu} > \frac{1}{m_1}, \frac{1}{m_2} \rightsquigarrow \mu < m_1, m_2. \tag{9.8}$$

As the name suggests, the reduced mass is smaller than any of the masses of the two particles. In (16), we plot the vectors $\vec{r}$, $\overrightarrow{R}$, $\vec{r}_1$ and $\vec{r}_2$. An interesting

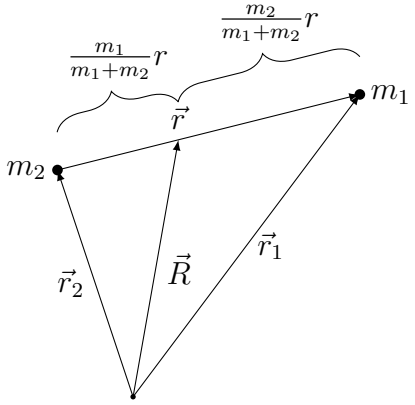


Figure 16: The position vectors of the two particles, their relative distance and the position of the *average* of the mass (center of mass).

configuration is when the mass m_2 is much larger than m_1. This corresponds, for example, to a typical *star-planet* system like the Sun and the Earth. For $m_2 \gg m_1 \rightsquigarrow \frac{m_1}{m_2} \to 0$, we have

$$\vec{r}_2 \approx \overrightarrow{R} \quad \vec{r}_1 \approx \overrightarrow{R} + \vec{r} \tag{9.9}$$

while the reduced mass is:

$$\mu \approx m_1 \tag{9.10}$$

which is the mass of the small object.

Let us now observe the structure of the Lagrangian of (9.8). We notice that the coordinates of the center of mass vector $\vec{R} \equiv (R_x, R_y, R_z)$ do not appear in the Lagrangian, but only the velocities $\dot{\vec{R}} = (\dot{R}_x, \dot{R}_y, \dot{R}_z)$ do. Therefore, the velocity of the center of mass must be constant:

$$\dot{\vec{R}} = \text{const.}. \tag{9.11}$$

Indeed, we recall that

$$\frac{\partial L}{\partial R_x} = 0 \rightsquigarrow \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{R}_x} = 0 \rightsquigarrow \frac{\mathrm{d}}{\mathrm{d}t}[(m_1+m_2)\dot{R}_x] = 0 \rightsquigarrow \dot{R}_x = \text{const.}. \quad \text{(similarly for } \dot{R}_y,\ \dot{R}_z).$$

We can now exercise our freedom to choose our frame of reference such as the center of mass is stationary:

$$\dot{\vec{R}} = 0. \tag{9.12}$$

In this frame, the Lagrangian takes the form:

$$L = \frac{1}{2}\mu\dot{\vec{r}}^2 - V(r). \tag{9.13}$$

We have turned the two-body problem into a dual problem of a single particle with "mass" equal to the reduced mass and a "position" the original distance vector of the two particles.
The potential in (9.13) does not depend on the direction of $\vec{r}$ but only on its magnitude. Therefore, we expect the Lagrangian to be symmetric under rotations:

$$\vec{r} \to \vec{r}' = \delta\vec{\phi} \times \vec{r}. \tag{9.14}$$

Then, the angular-momentum

$$\overrightarrow{M} = \vec{r} \times \overrightarrow{P} \tag{9.15}$$

with

$$\overrightarrow{P} = \left(\frac{\partial L}{\partial \dot{x}}, \frac{\partial L}{\partial \dot{y}}, \frac{\partial L}{\partial \dot{z}}\right) = \mu(\dot{x}, \dot{y}, \dot{z}) = \mu\dot{\vec{r}} \tag{9.16}$$

is conserved:

$$\frac{\mathrm{d}}{\mathrm{d}t}\overrightarrow{M} = 0 \rightsquigarrow \vec{r} \times \overrightarrow{P} = \mu(\vec{r} \times \dot{\vec{r}}) = \text{const.}. \tag{9.17}$$

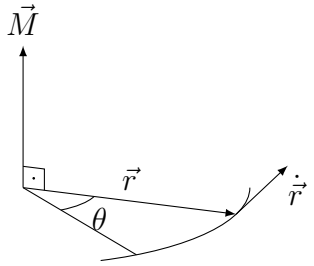


Figure 17: Directions.

The position $\vec{r}$ and the velocity $\dot{\vec{r}}$ are always perpendicular to the angular momentum. Since the latter is constant, pointing always to the same direction, the vector $\vec{r}$ lies always on the same plane.
We can therefore use polar coordinates, with the $\hat{z}$-axis in the direction of the angular momentum:

$$\overrightarrow{M} = M\hat{z}.$$

The Lagrangian takes then the form:

$$L = \frac{1}{2}\mu(\dot{r}^2 + r^2\dot{\theta}^2) - V(r). \tag{9.18}$$

The conservation of angular momentum is manifest in the Lagrangian of (9.18) since it does not depend explicitly on the angle θ, but only on the angular velocity $\dot{\theta}$. The corresponding Euler-Lagrange equation gives indeed:

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{\theta}} - \underbrace{\frac{\partial L}{\partial \theta}}_{=0} = \frac{\mathrm{d}}{\mathrm{d}t}(\mu r^2\dot{\theta}) \rightsquigarrow \boxed{\mu r^2\dot{\theta} = M = \text{const.}}. \tag{9.19}$$

As usual, the total energy is also conserved:

$$E = \frac{1}{2}\mu(\dot{r}^2 + r^2\dot{\theta}^2) + V(r) = \text{const.}. \tag{9.20}$$

Or, equivalently, eliminating the angle θ using (9.19) we write:

$$E = \frac{1}{2}\mu\dot{r}^2 + V(r) + \frac{M^2}{2\mu r^2}. \tag{9.21}$$

Exercise 9.1. Verify that by taking the time derivative of (9.20) or (9.21)

$$\frac{\mathrm{d}E}{\mathrm{d}t} = 0$$

you can obtain the second equation of motion

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{r}} - \frac{\partial L}{\partial r} = 0. \tag{9.22}$$

We can interpret (9.21) as the energy of a particle with mass μ moving in one only dimension and a potential

$$V_{\text{eff}}(r) = V(r) + \underbrace{\frac{M^2}{2\mu r^2}}_{\substack{\text{"centrifugal} \\ \text{energy"}}} . \tag{9.23}$$

Let us rewrite (9.21) in the form:

$$\frac{1}{2}\mu\dot{r}^2 = E - V_{\text{eff}}(r) = E - \frac{M^2}{2\mu r^2} - V(r). \tag{9.24}$$

The kinetic energy on the l.h.s. must be positive:

$$\frac{1}{2}\mu\dot{r}^2 \geq 0 \rightsquigarrow E - V_{\text{eff}}(r) \geq 0. \tag{9.25}$$

The motion is therefore restricted to distances for which the total energy exceeds the effective potential energy V_{eff}. The extrema of the motion are the solutions of the equation

$$E = V_{\text{eff}}(r). \tag{9.26}$$

This can yield different types of trajectories:

Example 9.1. In figure (18), the motion is restricted in a finite range of distances in between $[r_{\min}, r_{\max}]$.

Example 9.2. In figure (19), the energy is larger than $V_{\text{eff}}(r)$ up to infinite distances.

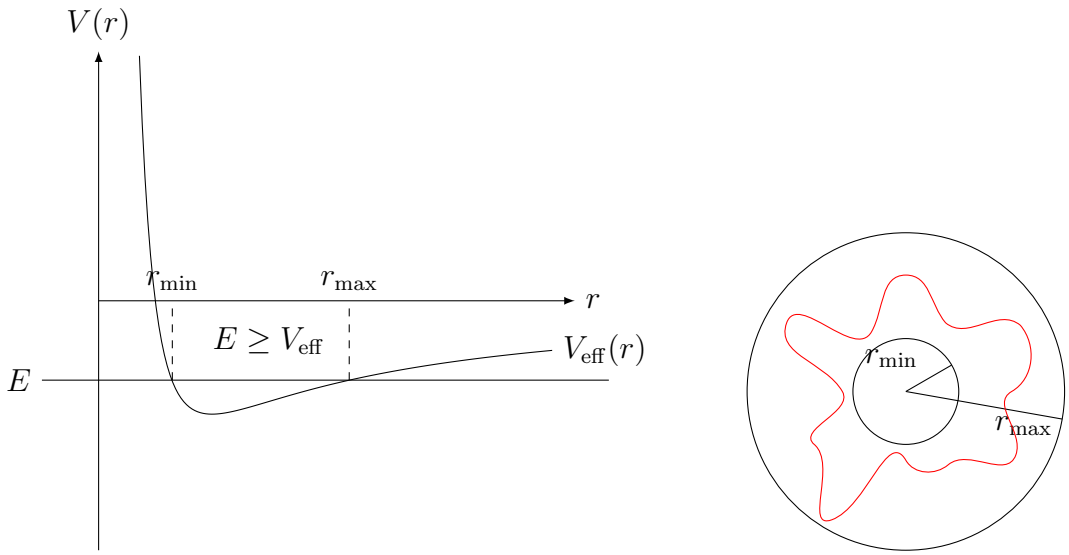


Figure 18: Trajectory 1.

Let us now look at the condition

$$\begin{aligned} E - V_{\text{eff}}(r) &\geq 0 \\ \rightsquigarrow E - V(r) - \frac{M^2}{2\mu r^2} &\geq 0 \\ \rightsquigarrow \boxed{2\mu E r^2 - 2\mu r^2 V(r) - M^2 \geq 0}&. \end{aligned}$$

Taking the limit $r \to 0$, we have:

$$-\lim_{r\to 0}[r^2 V(r)] \geq M^2 \tag{9.27}$$

which is a condition on the potential so that the distance r becomes zero. We clearly need an attractive potential ($V(r) < 0$) in order for (9.27) to have a chance to be realized. But even then, this is not guaranteed if the angular momentum is not zero. For example, for a gravity or an attractive electromagnetic potential

$$V(r) = -\frac{a}{r}, \quad a > 0 \tag{9.28}$$

(9.27) is not satisfied. This explains, for example, why the moon does not fall on the earth. We may rush to conclude that this is also the reason for

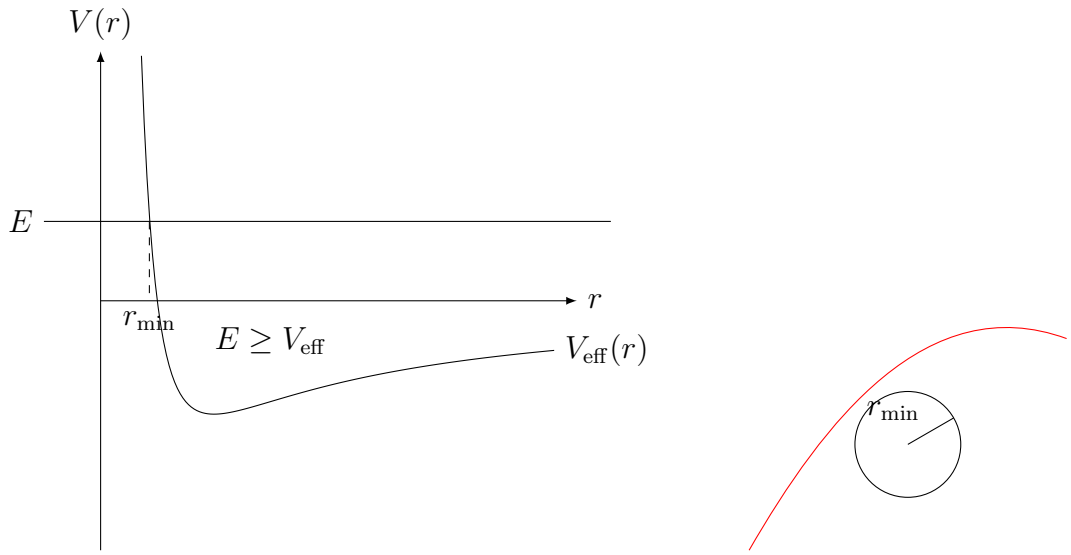


Figure 19: Trajectory 2.

the stability of atoms, explaining the orbits of electrons in analogy to planetary motions in a solar system. However, this analogy is not correct. In classical electrodynamics electrons in orbit around a nucleus must emit electromagnetic radiation thus decreasing continuously their energy and angular momentum. Eventually, electrons of classical physics should collapse on the nucleus. To explain the stability of atoms we will need quantum mechanics.

Angular momentum and energy conservation are sufficient to determine fully the trajectory $r(t)$ from

$$
\begin{aligned}
E &= \frac{1}{2}\mu\dot{r}^2 + V_{\text{eff}}(r) \\
\rightsquigarrow \frac{\mathrm{d}r}{\mathrm{d}t} &= \sqrt{\frac{2}{\mu}}\left(E - V_{\text{eff}}(r)\right)^{\frac{1}{2}} \\
\rightsquigarrow \mathrm{d}t &= \sqrt{\frac{\mu}{2}}\frac{\mathrm{d}r}{\sqrt{E - V_{\text{eff}}(r)}} \qquad (9.29) \\
\rightsquigarrow t &= \sqrt{\frac{\mu}{2}}\int \frac{\mathrm{d}r}{\sqrt{E - V_{\text{eff}}(r)}} + \text{const.}. \qquad (9.30)
\end{aligned}
$$

For the angle $\theta(t)$, we have

$$M = \mu r^2 \dot{\theta} \rightsquigarrow \int \mathrm{d}\theta = \int \mathrm{d}t \frac{M}{\mu r^2}.$$

Using (9.29) to trade $\mathrm{d}t$ for $\mathrm{d}r$, we have

$$\begin{aligned} \int \mathrm{d}\theta &= \sqrt{\frac{\mu}{2}} \int \frac{\mathrm{d}r}{\sqrt{E - V_{\text{eff}}(r)}} \frac{M}{\mu r^2} \\ \Rightarrow &\boxed{\theta = \frac{M}{\sqrt{2\mu}} \int \frac{\mathrm{d}r}{r^2} \frac{1}{\sqrt{E - V_{\text{eff}}}} + \text{const.}.} \end{aligned} \tag{9.31}$$

9.1 $\frac{1}{r}$ potential

Let us now consider a potential of the form

$$V(r) = -\frac{a}{r}, \quad a > 0. \tag{9.32}$$

This potential corresponds to the gravitational potential due to a massive object, or an attractive electrostatic potential due to a point charge. We will study first the case $a > 0$, but our results can be easily used for repulsive potentials with $a < 0$.

The effective potential which takes into account the centrifugal energy is:

$$V_{\text{eff}} = \frac{M^2}{2\mu r^2} - \frac{a}{r}. \tag{9.33}$$

We notice that the centrifugal energy is positive and behaves as $\frac{1}{r^2}$ while the potential energy $V(r)$ is negative and behaves as $\frac{1}{r}$. At small distances, $r \to 0$, the centrifugal energy dominates and V_{eff} tends asymptotically to plus infinity. In this region V_{eff} is attractive:

$$\lim_{r \to 0} -\frac{\partial V_{\text{eff}}}{\partial r} > 0.$$

At large distances, $r \to \infty$, the $\frac{1}{r}$ term is larger than the $\frac{1}{r^2}$ and the effective potential is dominated by $V(r)$. In this limit the force is attractive:

$$\lim_{r \to \infty} -\frac{\partial V_{\text{eff}}}{\partial r} < 0.$$

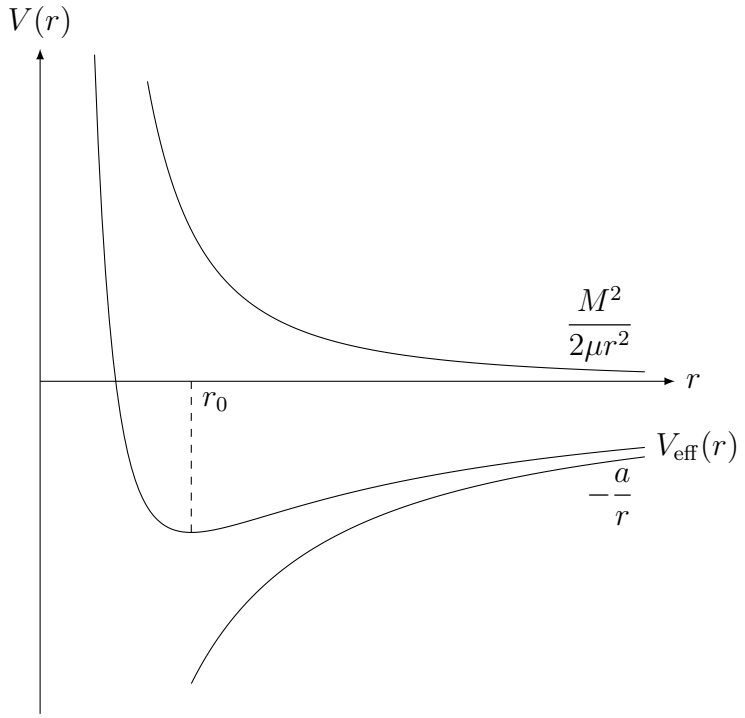


Figure 20: Effective potential.

In the intermediate region the potential develops a minimum:

$$\left.\frac{\partial V_{\text{eff}}}{\partial r}\right|_{r=r_0} \rightsquigarrow r_0 = \frac{M^2}{\mu a}. \tag{9.34}$$

The shape of the potential is shown in Figure (20). To calculate the trajectory, we need to calculate the integral of (9.31) for V_{eff} given by (9.33):

$$\theta(r) = \frac{1}{\sqrt{2\mu}} \int \frac{\mathrm{d}r}{r^2} \frac{1}{\sqrt{E - \frac{M^2}{2\mu r^2} + \frac{a}{r}}} + \text{const.}. \tag{9.35}$$

We can cast the above formula as

$$\theta(r) = \int \frac{\mathrm{d}r}{r^2} \frac{1}{\sqrt{-\left(\frac{1}{r} - \frac{1}{p}\right)^2 + \frac{\epsilon^2}{p^2}}} + \text{const.} \tag{9.36}$$

where

$$p \equiv \frac{M^2}{\mu a} \tag{9.37}$$

and

$$\epsilon \equiv \sqrt{1 + \frac{2EM^2}{\mu a^2}}. \tag{9.38}$$

With the change of variables,

$$\frac{1}{r} = \frac{1}{p} + \frac{e}{p}\cos\omega \tag{9.39}$$

the integral of (9.36) becomes

$$\theta = \int \mathrm{d}\omega + \text{const.}$$

which leads to

$$\theta = \omega + \theta_0. \tag{9.40}$$

Thus, from (9.39) we find the trajectory

$$\frac{1}{r} = \frac{1}{p} + \frac{e}{p}\cos(\theta - \theta_0)$$

or, equivalently:

$$\boxed{\frac{p}{r} = 1 + e\cos(\theta - \theta_0)}. \tag{9.41}$$

(9.41) describes the trajectory $r = r(\theta)$ for the distance of the two particles. As we can see, the trajectory depends on the value of the two parameters (e, p) which in term depend on the energy E and angular momentum M. We will find three types of trajectories, according to the following table. For the

$e > 1$	$E > 0$	hyperbola
$e = 1$	$E = 0$	parabola
$e < 1$	$E < 0$	ellipsis

minimum value of the energy,

$$E = -\frac{\mu a^2}{2M^2},$$

we have $\epsilon = 0$ and the trajectory is a perfect circle of radius p. Let us check that (9.41) describes indeed these types of trajectories. Substituting in (9.41)

$$\cos\theta = \frac{x}{r} \quad \text{and} \quad r = \sqrt{x^2+y^2}$$

we have:

$$\begin{aligned} p &= \sqrt{x^2+y^2} + ex \\ \rightsquigarrow x^2 + y^2 &= (p-ex)^2 \\ \rightsquigarrow x^2(1-e^2) + 2epx + y^2 &= p^2. \end{aligned} \tag{9.42}$$

For $e \neq 1$ we have:

$$\begin{aligned} x^2 + \frac{2epx}{1-e^2} + \frac{y^2}{1-e^2} &= \frac{p^2}{1-e^2} \\ \rightsquigarrow \left(x + \frac{ep}{1-e^2}\right)^2 + \frac{y^2}{1-e^2} &= \frac{p^2}{1-e^2} + \frac{e^2p^2}{(1-e^2)^2} \\ \Rightarrow \left[x + \frac{ep}{1-e^2}\right]^2 + \frac{y^2}{1-e^2} &= \frac{p^2}{(1-e^2)^2}. \end{aligned}$$

For $p \neq 0$, we can write:

$$\frac{\left[x + \frac{ep}{1-e^2}\right]^2}{\frac{p^2}{(1-e^2)^2}} + \frac{y^2}{\frac{p^2}{(1-e^2)^2}} = 1. \tag{9.43}$$

Let us now take $e < 1$. Then, (9.43) matches indeed the trajectory of an ellipse

$$\frac{(x-x_0)^2}{A^2} + \frac{(y-y_0)^2}{B^2} = 1$$

where

$$x_0 = -\frac{ep}{1-e^2}, \quad y_0 = 0 \tag{9.44}$$

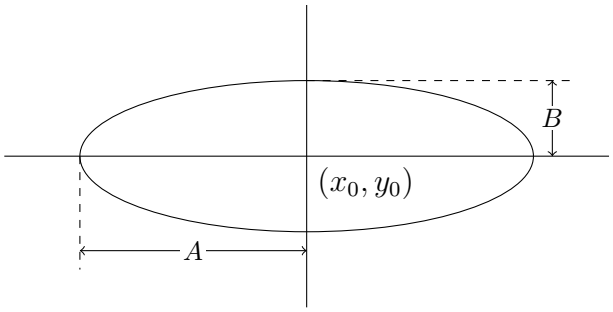


Figure 21: Ellipse with large semi-axis A and small semi-axis B.

and

$$A = \frac{p}{1-e^2} \tag{9.45}$$

$$B = \frac{p}{\sqrt{1-e^2}} = \sqrt{1-e^2}A \leq A \tag{9.46}$$

are the large and small semi-axis of the ellipse. We observe that

$$A = \frac{p}{1-e^2} = -\frac{a}{2E},$$

the large semi-axis of the ellipse depends only on the energy E, while the small semi-axis

$$B = A\sqrt{1-e^2} = A\sqrt{-\frac{2EM^2}{\mu a^2}} = \sqrt{-\frac{M^2}{2\mu E}}$$

depends on both energy and angular momentum.

Let us now return to (9.43) and consider the case $E > 0$, corresponding to $e > 1$. Then we can map (9.43) to the equation of a hyperbola,

$$\frac{(x-x_0)^2}{A^2} - \frac{(y-y_0)}{B^2} = 1 \tag{9.47}$$

with

$$x_0 = \frac{ep}{e^2-1}, \quad y_0 = 0$$

and

$$A = \frac{p}{e^2 - 1} \tag{9.48}$$

$$B = \frac{p}{\sqrt{e^2 - 1}} = \sqrt{e^2 - 1}A \tag{9.49}$$

We let the reader as an exercise to investigate the special cases of $\epsilon = 1$ and

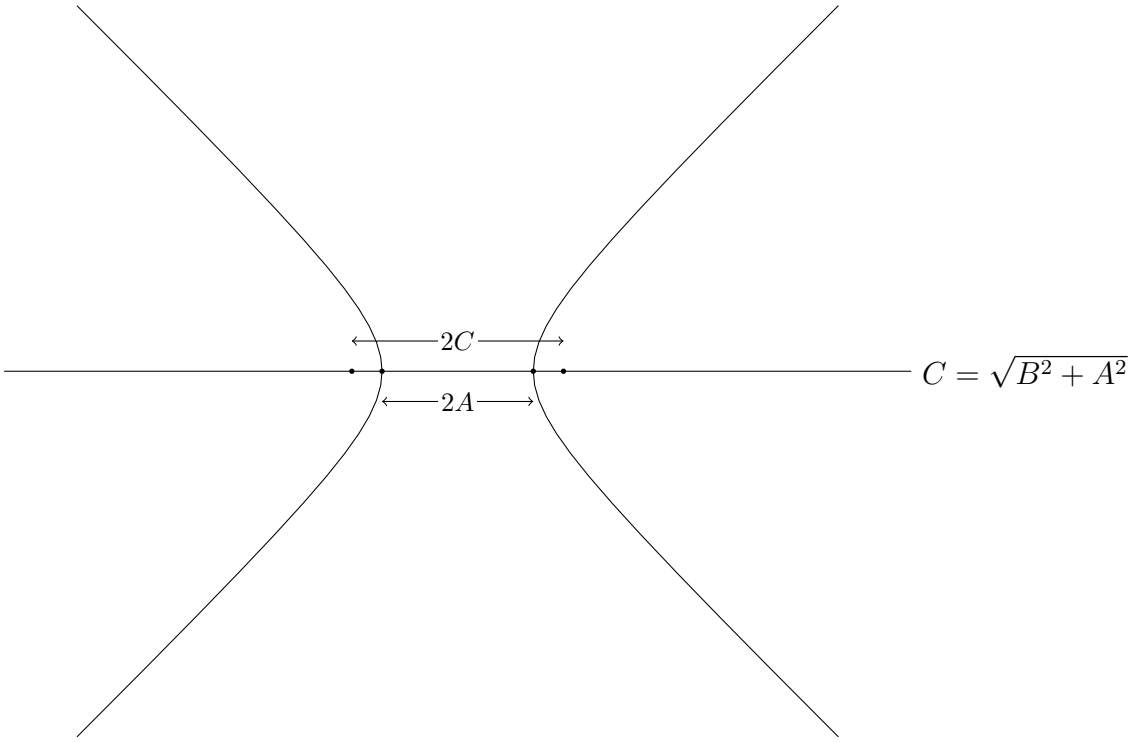


Figure 22: Hyperbola.

$E = 0$.

9.2 Laplace-Runge-Lenz vector

Consider the Lagrangian of a particle in a central $\frac{a}{r}$ potential:

$$L = \frac{1}{2}m\dot{\vec{r}}^2 + \frac{a}{r}. \tag{9.50}$$

The Euler-Lagrange equations yield Newton's law:

$$m\ddot{\vec{r}} = -a\frac{\vec{r}}{r^3}. \tag{9.51}$$

Now let's consider the cross product of the momentum and the angular momentum

$$\overrightarrow{P} \times \overrightarrow{M}.$$

Differentiating with time, we have:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}(\overrightarrow{P} \times \overrightarrow{M}) &= \frac{\mathrm{d}\overrightarrow{P}}{\mathrm{d}t} \times \overrightarrow{M} \\
&= -\frac{a}{r^3}(\vec{r} \times \overrightarrow{M}) \\
&= -\frac{a}{r^3}\left[\vec{r} \times (\vec{r} \times \overrightarrow{P})\right] \\
&= -\frac{ma}{r^3}\left[\vec{r} \times (\vec{r} \times \dot{\vec{r}})\right]. \qquad (9.52)
\end{aligned}$$

Exercise 9.2. Prove the identity:

$$\vec{r} \times (\vec{r} \times \dot{\vec{r}}) = \vec{r}(\vec{r} \cdot \dot{\vec{r}}) - r^2\dot{\vec{r}},$$

where

$$2r\dot{r} = \frac{\mathrm{d}r^2}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}(\vec{r} \cdot \vec{r}) = 2\vec{r}\dot{\vec{r}} \rightsquigarrow \vec{r}\dot{\vec{r}} = r\dot{r}.$$

Thus,

$$\vec{r} \times (\vec{r} \times \dot{\vec{r}}) = r^3\left(\frac{\vec{r}\dot{\vec{r}}}{r^2} - \vec{r}\frac{\dot{r}}{r^3}\right) = -r^3\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\vec{r}}{r}\right).$$

Returning to (9.52), we find:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\overrightarrow{P} \times \overrightarrow{M} &= ma\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\vec{r}}{r}\right) \\
\rightsquigarrow \frac{\mathrm{d}}{\mathrm{d}t}\left[\overrightarrow{P} \times \overrightarrow{M} - am\frac{\vec{r}}{r}\right] &= 0.
\end{aligned}$$

Thus, the vector

$$\overrightarrow{A} \equiv \overrightarrow{P} \times \overrightarrow{M} - \frac{\vec{r}}{r}ma = \text{const.} \tag{9.53}$$

is conserved. The vector is known as the *Laplace-Runge-Lenz vector*.

It is easy to verify that the LRL vector is vertical to the angular momentum

$$\vec{A} \cdot \overrightarrow{M} = 0 \tag{9.54}$$

and, thus, it must lay on the plane of the motion. We will see in late lectures that the LRL vector is conserved due to a special symmetry, owed to the special form $(\frac{1}{r})$ of the potential.

Exercise 9.3. Find the size and the orientation of the LRL vector for

(i) an elliptical orbit

(ii) a hyperbolic orbit.

It is important to note that any variations of the LRL vector indicate possible deviations from Newton's law of gravity. Thus, measurements of $\vec{A}(t)$ as a function of time are stringent tests of Newton's law. Such deviations are expected, for example, in Einstein's theory of General Relativity.

What is the symmetry giving use to the conservation law of LRL?

$$r_i \to r_i + \delta_k r_i$$

with

$$\delta_k r_i = \epsilon\mu \left(\dot{r}_i r_k - \frac{1}{2} r_i \dot{r}_k - \delta_{ik} \left(\frac{1}{2} \vec{r} \cdot \dot{\vec{r}} \right) \right) \tag{9.55}$$

where we change all three r_i's simultaneously. From (9.55), we obtain:

$$\delta_k \dot{r}_i = \frac{\epsilon\mu}{2} \left[\dot{r}_i \dot{r}_k - \delta_{ik} \dot{r}^2 - \frac{a}{\mu} \frac{r_i r_k}{r^3} + \delta_{ik} \frac{a}{\mu} \frac{1}{r} \right]. \tag{9.56}$$

The Lagrangian changes by:

$$\begin{aligned} \delta L &= \mu \dot{r} \cdot \delta \dot{r} - a \vec{r} \cdot \frac{\delta \vec{r}}{r^3} \\ \overset{(9.55),(9.56)}{\Rightarrow} \delta_k L &= \epsilon \frac{\mathrm{d}}{\mathrm{d}t} \left(\mu a \frac{r_k}{r} \right). \end{aligned}$$

Then, we can use Noether's theorem:

$$\begin{aligned} \frac{\partial L}{\partial_k r_i} \delta_k r_i - F &= \text{const.} \\ \Rightarrow \vec{A} = \vec{P} \times \overrightarrow{M} - \mu a \frac{\vec{r}}{r} &= \text{const.}. \end{aligned}$$

Exercise 9.4. Fill the gaps.

9.3 Scattering

Consider a particle which scatters against a fixed target at the center of our coordinate system. We will assume that the interaction of the particle and the target is described by a central potential $V(r)$. We will assume that the particle comes from very far away ($r \to \infty$) and scatters at the target. The

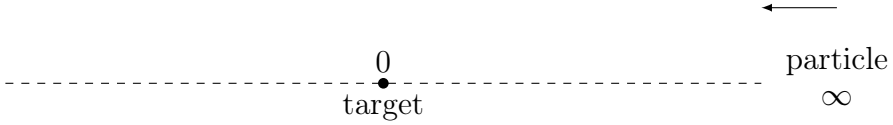


Figure 23: Particle coming from very far away and target at center.

trajectory of the particle is as in Figure 24. Therefore the deflection angle

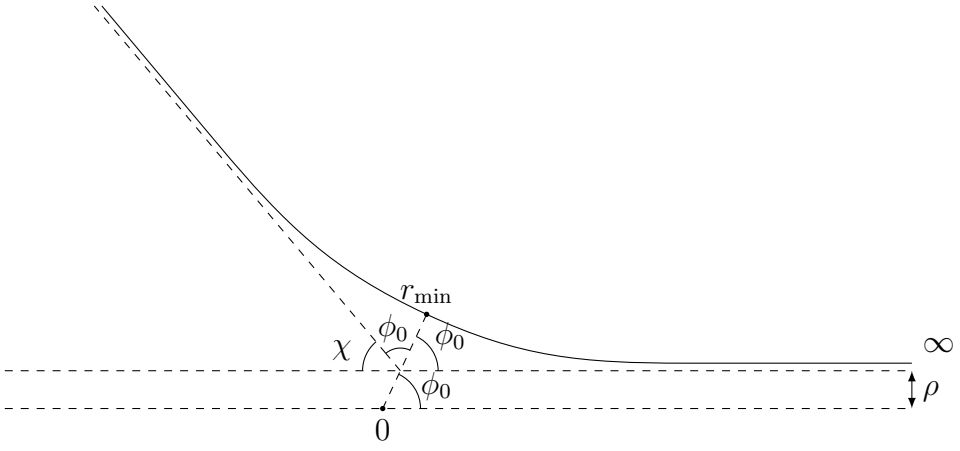


Figure 24: Trajectory of scattered particle.

χ of the particle is

$$\chi = |\pi - 2\phi_0|$$

where

$$\phi_0 = \int_{r_{\min}}^{\infty} \frac{\frac{M}{r^2}\,\mathrm{d}r}{\sqrt{2m[E - V(r)] - \frac{M^2}{r^2}}}. \tag{9.57}$$

To compute $r_{\min}$, we note that

$$\frac{1}{2}m\dot{r}^2 = E - V(r) - \frac{M^2}{2mr^2} \tag{9.58}$$

and $\dot{r} = 0$ for $r = r_{\min}$ leads to

$$E - V(r_{\min}) + \frac{M^2}{2mr_{\min}^2} = 0. \tag{9.59}$$

Let us now express the energy and the angular momentum in terms of the velocity of the particle at infinity, v_∞, and the *impact parameter* ρ. We have:

$$E = \frac{1}{2}mv_\infty^2 \tag{9.60}$$

$$M = m\rho v_\infty. \tag{9.61}$$

Thus:

$$\phi_0 = \int_{r_{\min}}^{\infty} \frac{\frac{\rho}{r^2}\,\mathrm{d}r}{\sqrt{1 - \frac{\rho^2}{r^2} - \frac{2V}{mv_\infty^2}}}. \tag{9.62}$$

Let us now consider a beam of particles with the same velocity at infinity which scatter on the target. The beam particles have different impact parameters. Thus they scatter at different angles. We denote with

$$\mathrm{d}N \equiv \frac{\#\text{ of scattered particles in between } \chi \text{ and } \chi + \mathrm{d}\chi}{\text{unit time}}. \tag{9.63}$$

This rate depends on the flux of particles in the incoming beam. We define the flux as

$$n \equiv \frac{\#\text{ of incoming particles}}{\text{time} \cdot \text{area}}.$$

We can measure experimentally the ratio

$$\mathrm{d}\sigma = \frac{\mathrm{d}N}{n}. \tag{9.64}$$

This ratio is called the *effective scattering cross-section*. It has units of are and it characterizes the physical properties of the scattering process and the potential. Let's assume that particles with impact factor $\rho(\chi)$ to $\rho(\chi) + \mathrm{d}\rho$ scatter in between χ and $\mathrm{d}\chi$. Then

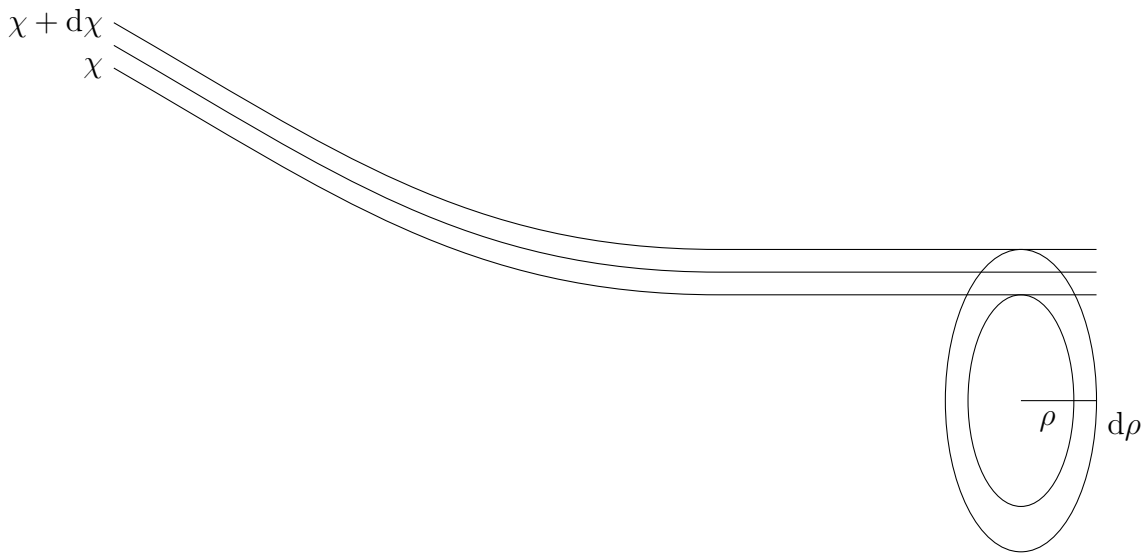


Figure 25: Particles with impact factor $\rho(\chi)$ to $\rho(\chi) + \mathrm{d}\rho$ scattering in between χ and $\mathrm{d}\chi$.

$$\mathrm{d}N = \underbrace{(2n\rho\,\mathrm{d}\rho)}_{\text{area}} \cdot \underbrace{n}_{\frac{\text{particles}}{\text{time}\cdot\text{area}}}$$

The cross-section is then

$$\mathrm{d}\sigma = \frac{\mathrm{d}N}{n} = 2\pi\rho\,\mathrm{d}\rho, \tag{9.65}$$

changing variables,

$$\begin{aligned}\mathrm{d}\sigma &= 2\pi\rho \left|\frac{\mathrm{d}\rho}{\mathrm{d}\chi}\right| \mathrm{d}\chi \\ &= \underbrace{(2\pi \sin\chi\,\mathrm{d}\chi)}_{\text{solid angle } \mathrm{d}\Omega} \frac{\rho}{\sin x} \left|\frac{\mathrm{d}\rho}{\mathrm{d}x}\right|.\end{aligned}$$

9.3.1 Rutherford's formula

For a $V(r) = \frac{a}{r}$ potential, we can compute

$$\phi = \cdots = \arccos \frac{\frac{a}{mv_\infty^2\rho}}{\sqrt{1 + \frac{a}{mv_\infty^2\rho}}}$$

Exercise 9.5. Verify this result.

$$\rightsquigarrow \rho^2 = \frac{a^2}{m^2 v_\infty^4} \cot^2 \frac{\chi}{2} \quad \left(\phi_0 = \frac{\pi - \chi}{2} \right).$$

Then $\mathrm{d}\rho^2 = 2\rho \, \mathrm{d}\rho$ is easy to compute. We then obtain

$$\mathrm{d}\sigma = \pi \frac{a}{(2mv_\infty^2)^2} \frac{\mathrm{d}\Omega}{\sin^4 \frac{\chi}{2}}.$$

10 Virial Theorem

Consider N particles interacting via Newton's law. If positions are bounded,

$$\langle T \rangle = -\frac{\langle V \rangle}{2}.$$

Consider

$$\begin{aligned} G &= \sum_i \vec{P}_i \cdot \vec{r}_i \\ \rightsquigarrow \frac{\mathrm{d}G}{\mathrm{d}t} &= \sum_i \frac{\mathrm{d}\vec{P}_i}{\mathrm{d}t} \cdot \vec{r}_i + \sum_i \vec{P}_i \cdot \frac{\mathrm{d}\vec{r}_i}{\mathrm{d}t} \\ &= \sum_i \vec{F}_i \cdot \vec{r}_i + 2T \\ \rightsquigarrow \frac{\mathrm{d}}{\mathrm{d}t} \sum_i \vec{P}_i \cdot \vec{r}_i &= 2T + \sum_i \vec{F}_i \cdot \vec{r}_i. \end{aligned}$$

The time average is

$$\langle F \rangle = \frac{1}{\tau} \int_0^\tau F(t)\, \mathrm{d}t.$$

Thus

$$\begin{aligned} \frac{1}{\tau} \int_0^\tau \mathrm{d}t \frac{\mathrm{d}}{\mathrm{d}t} \sum_i \vec{P}_i \cdot \vec{r}_i &= 2 \langle T \rangle + \left\langle \sum_i \vec{F}_i \cdot \vec{r}_i \right\rangle \\ \Rightarrow 2 \langle T \rangle + \left\langle \sum_i \vec{F}_i \cdot \vec{r}_i \right\rangle &= G(\tau) - G(0). \end{aligned}$$

Assume that the motions are bounded. Then $G(\tau)$ is bounded as well. If we then take $\tau \to \infty$, we have

$$\frac{G(\tau) - G(0)}{\tau} \overset{\tau \to \infty}{\longrightarrow} 0.$$

Then

$$\boxed{\langle T \rangle = -\frac{1}{2} \left\langle \sum_i \vec{F}_i \cdot \vec{r}_i \right\rangle}.$$

This is known as the *virial theorem*.

For a single particle,

$$\begin{aligned}\langle T\rangle &= -\frac{1}{2}\left\langle \sum_{i=1}^{1} \vec{F}_i \cdot \vec{r}_i \right\rangle \\ &= -\frac{1}{2}\left\langle (-\vec{\nabla}V)\cdot\vec{r}\right\rangle .\end{aligned}$$

For a central potential, we get

$$\langle T\rangle = \frac{1}{2}\left\langle \frac{\partial V}{\partial r} r \right\rangle .$$

If $V = ar^n$, we have $\dfrac{\partial V}{\partial r} = nar^n = nV$, thus:

$$\langle T\rangle = \frac{n}{2}\langle V\rangle .$$

For a gravity or electrostatic potential, we have:

$$n = -1 \rightsquigarrow \boxed{\langle T\rangle = -\frac{\langle V\rangle}{2}}.$$

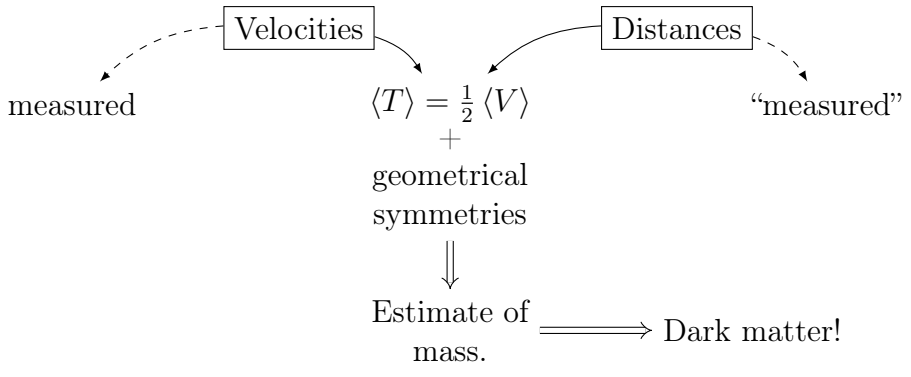
For many particles and a $\frac{1}{r}$ potential:

$$\begin{aligned}
-\sum_i (\vec{\nabla}_i V)\cdot\vec{r}_i &= -\sum_i \vec{r}_i \vec{\nabla}\frac{1}{2}\sum_{k\neq l}\frac{a}{|\vec{r}_k-\vec{r}_l|} \\
&= +2\frac{1}{2}\sum_i \vec{r}_i\vec{\nabla}\sum_{l\neq i}\frac{a(\vec{r}_i-\vec{r}_l)}{|\vec{r}_i-\vec{r}_l|^3} \\
&= +\sum_{l<i}\frac{a(\vec{r}_i-\vec{r}_l)\cdot\vec{r}_i}{|\vec{r}_i-\vec{r}_l|^3}+\sum_{i<l}\frac{a(\vec{r}_i-\vec{r}_l)\cdot\vec{r}_i}{|\vec{r}_i-\vec{r}_l|^3} \\
&= +\sum_{l<i}\frac{a(\vec{r}_i-\vec{r}_l)\cdot\vec{r}_i}{|\vec{r}_i-\vec{r}_l|^3}+\sum_{l<i}\frac{a(\vec{r}_l-\vec{r}_i)\cdot\vec{r}_l}{|\vec{r}_i-\vec{r}_l|^3} \\
&= +\sum_{l<i}\frac{a\left[(\vec{r}_i-\vec{r}_l)\cdot\vec{r}_i-(\vec{r}_i-\vec{r}_l)\cdot\vec{r}_l\right]}{|\vec{r}_i-\vec{r}_l|^3} \\
&= +\sum_{l<i}\frac{a(\vec{r}_i-\vec{r}_l)^2}{|\vec{r}_i-\vec{r}_l|^3} \\
&= +\sum_{l<i}\frac{a}{|\vec{r}_i-\vec{r}_l|} = V
\end{aligned}$$

Thus we have proven that

$$\langle T\rangle = -\frac{1}{2}\langle V\rangle .$$

Application in astrophysics



11 Oscillations

11.1 Small Oscillations

ROUGH DRAFT

Oscillations can occur in potentials with stable points of equilibrium. For systems with one degree of freedom, this means that for a potential $V(q)$ it exists a q_0 with

$$\frac{\partial V}{\partial q_0} > 0 \quad \text{and} \quad -\frac{\partial V}{\partial q_0} < 0.$$

This returns the particle to q_0. The Taylor expansion for this potential is

$$V(q) = V(q_0) + \cancelto{0}{V'(q_0)}(q - q_0) + \frac{1}{2}V''(q_0)(q - q_0)^2 + \mathcal{O}\left[(q - q_0)^3\right].$$

If we set $V(q_0) = 0$, which is the shifting value of the potential by a constant, we need to find

$$V(k) \approx \frac{1}{2}k(q - q_0)^2 \quad \text{with} \quad k > 0.$$

For a simple particle, we have

$$L \approx \frac{1}{2}m\dot{x}^2 - \frac{1}{2}kx^2$$

which gives us the following Euler-Lagrange equations:

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} = 0 \\ &\rightsquigarrow m\ddot{x} + kx = 0 \\ &\rightsquigarrow \ddot{x} + \frac{k}{m}x = 0 \\ &\boxed{\ddot{x} + \omega^2 x = 0} \quad \text{with} \quad \omega = \sqrt{\frac{k}{m}}. \end{aligned}$$

Then the solution is

$$x = c_1 \cos\omega t + c_2 \sin\omega t.$$

Indeed,

$$\begin{aligned} \dot{x} &= \omega\left[-c_1 \sin\omega t + c_2 \cos\omega t\right] \\ \ddot{x} &= \omega^2\left[-c_1 \cos\omega t - c_2 \sin\omega t\right] = -\omega^2 x \quad \checkmark. \end{aligned}$$

We can rewrite

$$x = \sqrt{c_1^2 + c_2^2}\left[\frac{c_1}{\sqrt{c_1^2 + c_2^2}}\cos\omega t + \frac{c_2}{\sqrt{c_1^2 + c_2^2}}\sin\omega t\right].$$

Set

$$\frac{c_1}{\sqrt{c_1^2 + c_2^2}} \equiv \cos\theta_0 \qquad -\frac{c_2}{\sqrt{c_1^2 + c_2^2}} \equiv \sin\theta_0 \qquad A \equiv \sqrt{c_1^2 + c_2^2}$$

to obtain

$$\begin{aligned} x &= A\left(\cos\theta_0\cos\omega t - \sin\theta_0\sin\omega t\right) \\ \rightsquigarrow &\boxed{x = A\cos(\omega t + \theta_0)}. \end{aligned}$$

The frequency depends on the characteristics of the system (m, k) and on

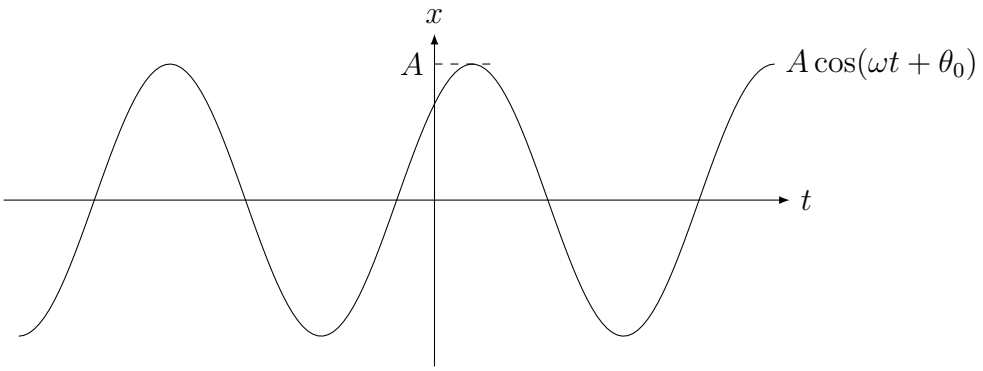


Figure 26: Harmonic oscillation with amplitude A.

initial conditions (for small oscillations only).

The energy of the system is

$$\begin{aligned} E &= \frac{1}{2}m\dot{x}^2 + \frac{1}{2}kx^2 \\ &= \frac{1}{2}m\left[-A\omega\sin(\omega t + \theta - 0)\right]^2 + \frac{m}{2}\omega^2\left[A\cos(\omega t + \theta_0)\right]^2 \\ \rightsquigarrow &\boxed{E = \frac{1}{2}m\omega^2 A^2}. \end{aligned}$$

Thus,

$$x = \sqrt{\frac{2E}{m\omega}} \cos(\omega t + \theta_0).$$

The amplitude A of the oscillation depends on the energy, as expected.

It is often convenient to "complexify" the amplitude:

$$A = |A| \mathrm{e}^{\mathrm{i}\theta_0}.$$

Then,

$$x = \mathrm{Re}\left(A \mathrm{e}^{\mathrm{i}\omega t}\right).$$

Indeed,

$$\begin{aligned} x &= \mathrm{Re}\left(|A| \mathrm{e}^{\mathrm{i}(\omega t + \theta_0)}\right) \\ &= |A| \cos(\omega t + \theta_0). \end{aligned}$$

11.2 Forced Oscillations

Exert an external time-dependent force on an oscillatory system, then

$$\begin{aligned} V &\approx \frac{1}{2}kx^2 + V_{\mathrm{ex}}(x,t) \\ &\approx \frac{1}{2}kx^2 + \left[\frac{\partial V_{\mathrm{ex}}}{\partial x}\right]_{x=0} + \underbrace{V_{\mathrm{ex}}(0,t)}_{\substack{\text{can be written} \\ \text{as a time derivative}}} . \end{aligned}$$

Thus,

$$\begin{aligned} V(x,t) &\approx \frac{1}{2}kx^2 - xF(t) \\ L &= \frac{1}{2}m\dot{x}^2 - \frac{1}{2}kx^2 + xF(t) \end{aligned} \tag{11.1}$$

which yields the following equation of motion:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} &= 0 \\ \rightsquigarrow m\ddot{x} &= -kx + F(t) \\ \rightsquigarrow \boxed{\ddot{x} + \omega^2 x = \frac{F(t)}{m}}. \end{aligned} \tag{11.2}$$

For the general solution, we set

$$\begin{aligned} \zeta &= \dot{x} + \mathrm{i}\omega x \\ \dot{\zeta} &= \ddot{x} + \mathrm{i}\omega\dot{x} \\ -\mathrm{i}\omega\zeta &= -\mathrm{i}\omega\dot{x} + \omega^2 x \end{aligned} \tag{11.3}$$

which gives us

$$\frac{\mathrm{d}}{\mathrm{d}t}\zeta - \mathrm{i}\omega\zeta = \frac{F(t)}{m}. \tag{11.4}$$

Set now $\zeta = \mathrm{e}^{\mathrm{i}\omega t}y$, then

$$\begin{aligned} \mathrm{e}^{\mathrm{i}\omega t}\dot{y} + \mathrm{i}\omega\zeta - \mathrm{i}\omega\zeta &= \frac{F(t)}{m} \\ \Rightarrow \dot{y} &= \frac{\mathrm{e}^{-\mathrm{i}\omega t}F(t)}{m} \\ \Rightarrow y &= \int \mathrm{d}t \mathrm{e}^{-\mathrm{i}\omega t}\frac{F(t)}{m} + y_0 \\ \Rightarrow &\boxed{\zeta = \mathrm{e}^{\mathrm{i}\omega t}\left[y_0 + \int \mathrm{d}t \mathrm{e}^{-\mathrm{i}\omega t}\frac{F(t)}{m}\right]}. \end{aligned} \tag{11.5}$$

The energy of the system is

$$\begin{aligned} E &= \frac{1}{2}m\dot{x}^2 + \frac{1}{2}kx^2 \\ &= \frac{1}{2}m(\dot{x}^2 + \omega^2 x^2) = \frac{1}{2}m\zeta^2 \\ &= \frac{1}{2}m(\dot{x} - \mathrm{i}\omega x)(\dot{x} + \mathrm{i}\omega x) \\ \Rightarrow &\boxed{E = \frac{1}{2}m\left|\zeta(t)\right|^2} \end{aligned} \tag{11.6}$$

The energy is not conserved. Assuming that we "pump" all energy in via $F(t)$, we have for the energy of the system after infinite time,

$$\begin{aligned} E &= \frac{1}{2}m\left|\zeta(+\infty)\right|^2 \\ &= \frac{1}{2}m\left|\int_{-\infty}^{\infty}\mathrm{d}t\,\frac{F(t)}{m}\mathrm{e}^{-\mathrm{i}\omega t} + y_0\right|^2. \end{aligned} \tag{11.7}$$

where the last integral can be recognized as the Fourier transform of the force.

Let us now work out explicitly the trajectory for

$$F(t) = f\cos(\gamma t + \beta). \tag{11.8}$$

Then

$$\begin{aligned}\zeta(t) &= \mathrm{e}^{+\mathrm{i}\omega t}\left[\int \mathrm{d}t\,\mathrm{e}^{-\mathrm{i}\omega t}\cos(\gamma t+\beta)\frac{f}{m} + y_0\right] \\ \Rightarrow \zeta(t) &= \mathrm{e}^{\mathrm{i}\omega t}y_0 + \frac{f\gamma}{m(\gamma^2-\omega^2)}\sin(\gamma t+\beta) - \mathrm{i}\frac{f\omega}{m(\gamma^2-\omega^2)}\cos(\gamma t+\beta)\end{aligned} \tag{11.9}$$

But

$$\begin{aligned}\zeta(t) &= \dot{x} + \mathrm{i}\omega x \\ \Rightarrow \mathrm{e}^{+\mathrm{i}\omega t}\zeta(t) &= \frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{e}^{\mathrm{i}\omega t}x) \\ \Rightarrow x &= \mathrm{e}^{-\mathrm{i}\omega t}\left(c + \int \mathrm{d}t\,\mathrm{e}^{+\mathrm{i}\omega t}\zeta(t)\right).\end{aligned} \tag{11.10}$$

Performing the integrations, we have

$$x(t) = -\frac{f/m}{\gamma^2-\omega^2}\cos(\gamma t+\beta) + A\mathrm{e}^{-\mathrm{i}\omega t}.$$

The real part of the solution is

$$\boxed{x(t) = |A|\cos(\omega t+\theta_0) + \frac{f/m}{\omega^2-\gamma^2}\cos(\gamma t+\beta)}. \tag{11.11}$$

The first term is the solution of the homogeneous equation and it is the same as when the external force is absent. The second term is new. It is also an oscillatory function of time, but with the frequency of the external force.

The amplitude of the oscillation increases as the frequency of the external force γ approaches the value of the characteristic frequency of the system ω. This phenomenon is called *resonance*. To calculate the motion at exactly $\omega = \gamma$, we write:

$$x(t) = |B|\cos(\omega t+\theta_0) + \frac{f}{m(\omega^2-\gamma^2)}\left[\cos(\gamma t+\beta) - \cos(\omega t + beta)\right] \tag{11.12}$$

At $\gamma \to \omega$, we have

$$x(t) = |B| \cos(\omega t + \theta_0) + \frac{f}{2m\omega} t \sin(\omega t + \beta) \tag{11.13}$$

Notice that the amplitude increases with time. The above formula is not valid if our Lagrangian of (11.1) is only valid for small x, since after some time the amplitude of the oscillation becomes large.

Let us now complexify (11.11):

$$\begin{aligned} x &= A_1 \mathrm{e}^{\mathrm{i}\omega t} + A_2 \mathrm{e}^{\mathrm{i}\gamma t} \\ &= \left[A_1 + A_2 \mathrm{e}^{\mathrm{i}(\gamma-\omega)t}\right] \mathrm{e}^{\mathrm{i}\omega t}. \end{aligned} \tag{11.14}$$

For $\gamma = \omega + \epsilon$, close the resonance with ϵ being small, we can write

$$\begin{aligned} x(t) &= \left[A_1 + A_2 \mathrm{e}^{\mathrm{i}\epsilon t}\right] \mathrm{e}^{\mathrm{i}\omega t} \\ &= C(t) \mathrm{e}^{\mathrm{i}\omega t} \end{aligned} \tag{11.15}$$

where the "amplitude" $C(t)$ varies very slowly in time. Setting

$$A_1 = |A_1| \mathrm{e}^{\mathrm{i}\theta_1} \quad \text{and} \quad A_2 = |A_2| \mathrm{e}^{\mathrm{i}\theta_2},$$

we obtain:

$$\begin{aligned} |C(t)|^2 &= |A_1|^2 + |A_2|^2 + 2|A_1||A_2| \cos(\theta_2 - \theta_1 + \epsilon t) \\ \Rightarrow (|A_1||A_2|)^2 &\leq |C(t)|^2 \leq (|A_1||A_2|)^2 \\ \Rightarrow ||A_1| - |A_2|| &\leq C(t) \leq |A_1| + |A_2|. \end{aligned} \tag{11.16}$$

The amplitude varies periodically in between two values. This phenomenon is known as *beats*.

11.3 Oscillations of systems with many degrees of freedom

We will now study oscillatory behaviour in systems with many degrees of freedom. A Lagrangian describing such a system is of the form

$$L = \frac{1}{2} \sum_{i,j} (m_{ij} \dot{x}_i \dot{x}_j - k_{ij} x_i x_j) \tag{11.17}$$

where

$$m_{ij} = m_{ji} \tag{11.18}$$
$$k_{ij} = k_{ji}. \tag{11.19}$$

To compute the equations of motion, we need first

$$\frac{\partial L}{\partial \dot{x}_i} = \sum_j m_{ij}\dot{x}_j \tag{11.20}$$
$$\frac{\partial L}{\partial x_i} = \sum_j k_{ij}x_j. \tag{11.21}$$

The Euler-Lagrange equations give then:

$$\sum_j (m_{ij}\ddot{x}_j + k_{ij}x_j) = 0, \quad i = 1, \ldots, N. \tag{11.22}$$

This is a system of N differential equations with N unknowns. We can set

$$x_j = A_j \mathrm{e}^{\mathrm{i}\omega t} \tag{11.23}$$

(and only need to remember to keep the real part of this expression at the end of the calculation).
Substituting into (11.22), we obtain:

$$\sum_j (k_{ij} - \omega^2 m_{ij})A_j = 0 \tag{11.24}$$

or, in a matrix form

$$\begin{pmatrix} k_{11} - \omega^2 m_{11} & k_{12} - \omega^2 m_{12} & \ldots \\ k_{21} - \omega^2 m_{21} & k_{22} - \omega^2 m_{22} & \ldots \\ \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{11.25}$$

The condition for the system to have oscillatory solutions of the form of (11.23) is then

$$\det(k_{ij} - \omega^2 m_{ij}) = 0 \tag{11.26}$$

The solutions of (11.26), if they exist, are real and positive. Indeed:

$$\sum_j (k_{ij} - m_{ij}\omega^2) A_j = 0$$

$$\rightsquigarrow \sum_{i,j} (k_{ij} - m_{ij}\omega^2) A_i^\star A_j = 0$$

$$\Rightarrow \omega^2 = \frac{\sum_{i,j} k_{ij} A_i^\star A_j}{\sum_{i,j} m_{ij} A_i^\star A_j} = \frac{\sum_{i,j} k_{ji} A_i^\star A_j}{\sum_{i,j} m_{ji} A_i^\star A_j} = \frac{\sum_{i,j} k_{ij} A_i A_j^\star}{\sum_{i,j} m_{ij} A_i A_j^\star} = \left(\omega^2\right)^\star \qquad \rightsquigarrow \omega^2 \text{ is real.}$$

Notice also that

$$\sum_{i,j} A_i A_j^\star = \sum_{i,j} k_{ij}(a_i + \mathrm{i}b_i)(a_j - \mathrm{i}b_j)$$

$$= \sum_{i,j} k_{ij}(a_i a_j + b_i b_j) + \mathrm{i} \sum_{i,j} \underbrace{k_{ij}}_{\text{symmetric}} \cancelto{0}{(b_i a_j - a_i b_j)}$$

$$= \sum_{i,j} k_{ij}(a_i a_j + b_i b_j) \geq 0.$$

The roots of ω_a^2 of (11.26) give the characteristic frequencies (eigenfrequencies) of the system. To construct solutions $x_j(t)$ after this step becomes a simple exercise in Linear Algebra to find the eigenvectors corresponding to each ω_a. Let us assume, for simplicity, that there is no degeneracy. Then the solutions are

$$x_j(t) = \sum_a \Delta_{ja} \theta_a(t) \tag{11.27}$$

with Δ_{ja} the minors of the determinant and

$$\theta_a(t) = \mathrm{Re}\left[C_a \exp(\mathrm{i}\omega_a t)\right]. \tag{11.28}$$

We can regard the equation

$$x_j = \sum_a \Delta_{ja} \theta_a$$

as a change from one set of generalized coordinates to another. The θ_a coordinates satisfy the differential equations

$$\ddot{\theta}_a + \omega_a^2 \theta_a = 0 \tag{11.29}$$

and are called normal coordinates. They describe independent oscillators. In normal coordinates, the Lagrangian takes the form

$$L = \sum_a \frac{1}{2} m_a (\dot{\theta}_a^2 - \omega_a^2 \omega_a^2).$$

Exercise 11.1. Consider the Lagrangian

$$L = \frac{1}{2}\left[\dot{x}_1^2 + 2\dot{x}_1\dot{x}_2 + \dot{x}_2^2 - 3x_1^2 - 4x_2^2 + 2x_1x_2\right].$$

Find the eigenfrequencies and normal coordinates.

11.4 Damped Oscillations

We now consider oscillatory systems with some friction losses. The equation fo motion for the simplest one-dimensional oscillator is

$$\boxed{m\ddot{x} + kx + \alpha\dot{x} = 0}$$

or, equivalently,

$$\ddot{x} + \omega_0^2 x + 2\lambda x = 0$$

with

$$\omega_0^2 = \frac{k}{m} \quad \text{and} \quad 2\lambda = \frac{a}{m}.$$

The solution is of the form

$$x(t) = \mathrm{e}^{kt},$$

then

$$\begin{aligned} k^2 + \omega_0^2 + 2\lambda k &= 0 \\ \rightsquigarrow (k+\lambda)^2 &= \lambda^2 - \omega_0^2 \\ \Rightarrow k &= -\lambda \pm \sqrt{\lambda^2 - \omega_0^2} \end{aligned}$$

Case 1 $\lambda < \omega_0$

Then, we have two complex solutions

$$k_\pm = -\lambda \pm \mathrm{i}\sqrt{\omega_0^2 - \lambda^2}$$

and the general solution is

$$\begin{aligned} x &= \mathrm{Re}\left[A\mathrm{e}^{-\lambda t + \mathrm{i}\sqrt{\omega_0^2-\lambda^2}t}\right] \\ &= \underbrace{|A|\mathrm{e}^{-\lambda t}}_{C(t)} \cos(\omega t + \theta_0) \end{aligned}$$

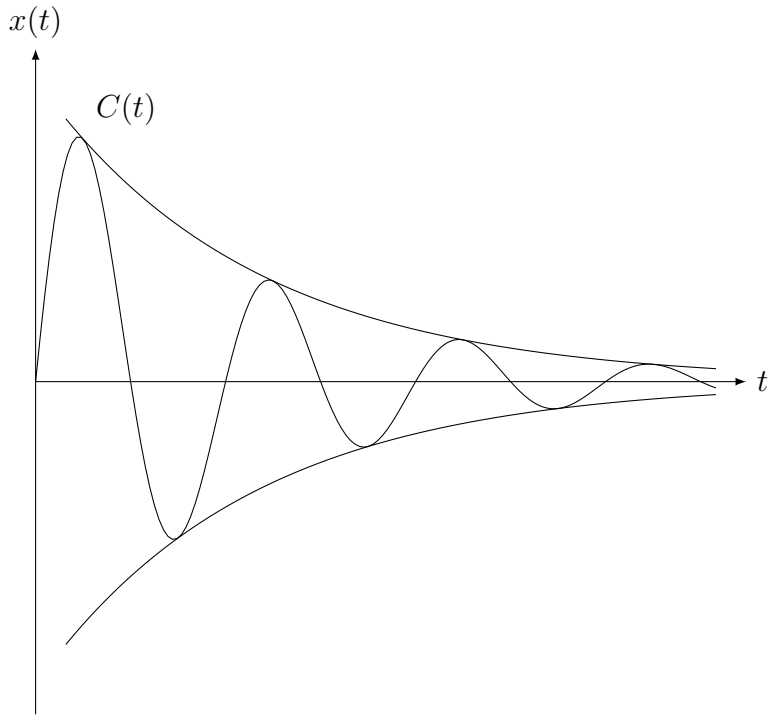


Figure 27: Damped oscillation with time-varying amplitude $C(t)$.

Case 2 $\lambda > \omega_0$

We have two real solutions and no oscillation, which is called *aperiodic*

damping:

$$x = \mathrm{e}^{-\lambda t}\left[c_1 \mathrm{e}^{-\sqrt{\lambda^2-\omega_0^2}t} + c_2 \mathrm{e}^{+\sqrt{\lambda^2-\omega_0^2}t}\right].$$

For many degrees of freedom, the frictional forces take the form:

$$f_i = -\sum_j \alpha_{ij}\dot{x}_j.$$

Then, we can write

$$f_i = -\frac{\partial F}{\partial x_i} \quad \text{with} \quad F = \frac{1}{2}\sum_{i,j}\alpha_{ij}\dot{x}_i\dot{x}_j.$$

Then, the equations of motion can originate from

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{x}_i} - \frac{\partial L}{\partial x_i} = -\frac{\partial F}{\partial \dot{x}_i}}.$$

The function F is known as the dissipative function.
Let us now compute the energy change vote:

$$\begin{aligned}
\frac{\mathrm{d}E}{\mathrm{d}t} &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_i \dot{x}_i\frac{\partial L}{\partial \dot{x}_i} - L\right) \\
&= \ddot{x}\frac{\partial L}{\partial \dot{x}_i} - \frac{\mathrm{d}L}{\mathrm{d}t} + \dot{x}_i\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{x}_i} \\
&= \ddot{x}\frac{\partial L}{\partial \dot{x}_i} - \frac{\partial L}{\partial x_i}\dot{x}_i - \frac{\partial L}{\partial \dot{x}_i}\ddot{x}_i + \dot{x}_i\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial x_i} \\
&= \dot{x}_i\left[\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{x}_i} - \frac{\partial L}{\partial x_i}\right] \\
&= -\dot{x}_i\frac{\partial F}{\partial \dot{x}_i} \\
&= -2F.
\end{aligned}$$

For the system to lose energy, we require that

$$F > 0.$$

12 Rigid Body

So far, we have studied the motion of point-like particles. We will now move to study the motion of macroscopic objects which do not change shape during their motion. We call these object *rigid bodies*. A rigid body is a special case of a system of many point-like particles with the constraint that the distances among the constituent particles is fixed.

Consider a reference point O on the rigid body. We can determine the position of every other point A of the rigid body with respect to the reference point O with a vector

$$\overrightarrow{OA} = \vec{l}_a.$$

The magnitude of $\vec{l}_a$ is fixed, since it connects two points of the rigid body.

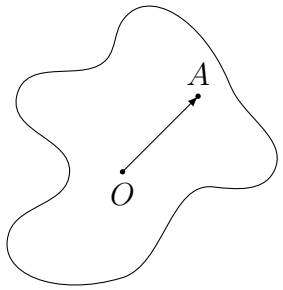


Figure 28: Rigid body with origin O and point A.

As the object moves only the orientation of $\overrightarrow{OA}$ can change, while the length of the vector stays fixed. The motion of $\vec{l}_a$ is therefore a rotation. The reference point O is of course also moving. The position of the point A in the body with respect to a fixed point F outside the body is

$$\overrightarrow{FA} = \overrightarrow{FO} + \overrightarrow{OA}$$

or, equivalently

$$\vec{r}_a(t) = \vec{R}(t) + \vec{l}_a(t). \tag{12.1}$$

For small times $\mathrm{d}t$, the displacement of the point A is

$$\mathrm{d}\vec{r}_a = \mathrm{d}\vec{R} + \mathrm{d}\vec{l}_a. \tag{12.2}$$

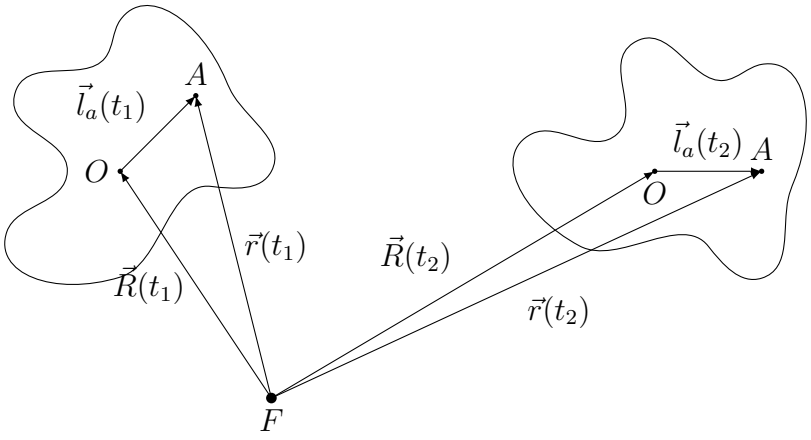


Figure 29: Same body with different orientation.

We can write the change $\mathrm{d}\vec{l}_a$ as an infinitesimal rotation

$$\mathrm{d}\vec{l}_a = \mathrm{d}\vec{\phi} \times \vec{l}_a. \tag{12.3}$$

Thus, we have that

$$\boxed{\frac{\mathrm{d}\vec{r}_a}{\mathrm{d}t} = \frac{\mathrm{d}\vec{R}}{\mathrm{d}t} + \frac{\mathrm{d}\vec{\phi}}{\mathrm{d}t} \times \vec{l}_a}. \tag{12.4}$$

This is an equation which relates the velocity of a particle A in the rigid body to the velocity of a reference point O and the angular velocity

$$\vec{\omega} = \frac{\mathrm{d}\vec{\phi}}{\mathrm{d}t} \tag{12.5}$$

of a point A with respect to the fixed point O. All points in a rigid body rotate simultaneously, as they are required to have fixed distances. Therefore, the angular velocity ω is common to all particles.

Let us call

$$\vec{V} \equiv \frac{\mathrm{d}\vec{R}}{\mathrm{d}t} \tag{12.6}$$

the velocity of the reference point O.
We will assume that a body is composed of $a = 1, \ldots, N$ particles with

masses m_a. The kinetic energy of the body is then

$$\begin{aligned} T &= \sum_a \frac{1}{2} m_a \dot{\vec{r}}_a^2 \\ &= \sum_a \frac{1}{2} m_a \left(\vec{V} + \vec{\omega} \times \vec{l}_a \right)^2 \\ &= \frac{1}{2} \left(\sum_a m_a \right) V^2 + \frac{1}{2} \sum_a m_a \left(\vec{\omega} \times \vec{l}_a \right)^2 + \vec{V} \cdot \left(\vec{\omega} \times \sum_a m_a \vec{l}_a \right). \end{aligned} \tag{12.7}$$

We will now choose our reference point O to be the center of mass

$$O : \sum_a m_a \vec{l}_a = 0. \tag{12.8}$$

In this frame, the last term of (12.7) vanishes. Therefore, in the center of mass frame the kinetic energy is

$$T = \frac{1}{2} m V^2 + \frac{1}{2} \sum_a m_a \left(\vec{l}_a \times \omega \right)^2 \tag{12.9}$$

with

$$m \equiv \sum_a m_a. \tag{12.10}$$

We not that the center of mass frame is not necessarily an inertial system of reference.
We will now prove the identity

$$\left(\vec{a} \times \vec{b} \right)^2 = \vec{a}^2 \vec{b}^2 - \left(\vec{a} \cdot \vec{b} \right)^2. \tag{12.11}$$

Proof.

$$\begin{aligned} \left(\vec{a} \times \vec{b} \right)^2 &= \left(\epsilon_{ijk} a_j b_k \right)^2 \\ &= \left(\epsilon_{ijk} a_j b_k \right) \left(\epsilon_{ilm} a_l b_m \right) \\ &= \epsilon_{ijk} \epsilon_{ilm} a_j b_k a_l b_m \\ &= \begin{vmatrix} \delta_{jl} & \delta_{jm} \\ \delta_{kl} & \delta_{km} \end{vmatrix} a_j b_k a_l b_m \\ &= \left(\delta_{jl} \delta_{km} - \delta_{kl} \delta_{jm} \right) a_j b_k a_l b_m \\ &= a_j^2 b_k^2 - \left(a_j b_j \right) \left(a_k b_k \right) \\ &= \vec{a}^2 \vec{b}^2 - \left(\vec{a} \cdot \vec{b} \right)^2 \end{aligned}$$

□

Therefore, the "rotational" term of the kinetic energy in (12.9) becomes

$$\frac{1}{2}\sum_a m_a \left(\vec{\omega}\times\vec{l}_a\right)^2 = \frac{1}{2}\sum_a m_a \left(\omega^2 l_a^2 - \left(\vec{\omega}\cdot\vec{l}_a\right)^2\right). \tag{12.12}$$

Let us now write the above in terms of the Cartesian components of the position vectors

$$\vec{l}_a \equiv (l_{a1}, l_{a2}, l_{a3}) \equiv (l_{ai})\,.$$

We have

$$\begin{aligned}\omega^2 l_a^2 - \left(\vec{\omega}\cdot\vec{l}_a\right)^2 &= (\omega_i\omega_i)\, l_a^2 - (\omega_i l_{ai})^2 \\ &= \omega_i\omega_j\delta_{ij} l_a^2 - \omega_i\omega_j l_{ai} l_{aj} \\ &= \left(\delta_{ij} l_a^2 - l_{ai} l_{aj}\right)\omega_i\omega_j \end{aligned} \tag{12.13}$$

Therefore, the kinetic energy takes the form

$$T = \frac{1}{2}mV^2 + \frac{1}{2}I_{ij}\omega_i\omega_j \tag{12.14}$$

where $m \equiv \sum_a m_a$ is the total mass of the rigid body and I_{ij} is the so called *tensor of inertia*.

12.1 Tensor of inertia

The tensor of inertia is characteristic of the geometry and mass distribution of the rigid body. From its definition,

$$I_{ij} \equiv \sum_a m_a \left(l_i^2\delta_{ij} - l_{ai} l_{aj}\right) \tag{12.15}$$

we see that it is a symmetric tensor:

$$I_{jk} = I_{kj}.$$

It therefore consists of six components. In a matrix notation, the tensor of inertia is

$$I_{jk} = \sum_a \begin{pmatrix} l_{a2}^2 + l_{a3}^2 & -l_{a1}l_{a2} & -l_{a1}l_{a3} \\ -l_{a1}l_{a2} & l_{a1}^2 + l_{a3}^2 & -l_{a2}l_{a3} \\ -l_{a1}l_{a3} & -l_{a2}l_{a3} & l_{a1}^2 + l_{a2}^2 \end{pmatrix} m_a. \tag{12.16}$$

Using standard linear algebra methods, we can diagonalize this matrix. The diagonalization corresponds to a change of axes for the vectors $\vec{l}_a$. The axes for which the tensor of inertia is diagonal are called *principal axes*.

$$\begin{pmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{pmatrix} \overset{\text{Rotation}}{\underset{\text{of axes}}{\longrightarrow}} \begin{pmatrix} I_1 & 0 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & I_3 \end{pmatrix}. \tag{12.17}$$

The eigenvalues I_1, I_2, I_3 are called the *principal moments of inertia*.

Exercise 12.1. Demonstrate that principal axes always exist.

We can prove the following inequalities:

$$\begin{aligned} I_1 + I_2 &\geq I_3 \\ I_2 + I_3 &\geq I_1 \\ I_3 + I_1 &\geq I_2. \end{aligned} \tag{12.18}$$

Indeed,

$$\begin{aligned} I_1 + I_2 &= \sum_a \left(l_{a2}^2 + l_{a3}^2 + l_{a1}^2 + l_{a3}^2 \right) \\ &\geq \sum_a m_a \left(l_{a1}^2 + l_{a2}^2 \right) \\ &= I_3. \end{aligned}$$

In the definition of the tensor of inertia of (12.16), the position vectors of the rigid body points $\vec{l}_a$ are measured with respect to the center of mass. Let us now consider an analogous expression

$$I'_{ij} \equiv \sum_a m_a \left(r_a^2 \delta_{ij} - r_{ai} r_{aj} \right) \tag{12.19}$$

where the vectors $\vec{r}_a \equiv \vec{l}_a + \vec{c}$ are measured with respect to a different point. Then,

$$\begin{aligned} I'_{ij} =& \sum_a m_a \left(\left(\vec{l}_a + \vec{c} \right)^2 \delta_{ij} - \left(l_{ai} + c_i \right) \left(l_{aj} + c_j \right) \right) \\ =& \sum_a m_a \left(\vec{l}_a^{\,2} \delta_{ij} - l_{ai} l_{aj} \right) + \left(\sum_a m_a \right) \left(\vec{c}^{\,2} \delta_{ij} - c_i c_j \right) \\ &+ 2 \left(\sum_a m_a \vec{l}_a \right) \cdot \vec{c} \delta_{ij} \left(\sum_a m_a l_{ai} \right) c_j - \left(\sum_a m_a l_{aj} \right) c_i. \end{aligned}$$

However, $\sum_a m_a \vec{l}_a = 0$. Therefore, we obtain

$$I'_{ij} = I_{ij} + m\left(\vec{c}^2\delta_{ij} - c_i c_j\right). \tag{12.20}$$

(12.20) is a convenient formula to use if the summation over all particles is performed easier in I'_{ij} for position vectors with a different origin than the center of mass.

Finally, we remark that in the case of a continuous solid body, the sum over all particles must be replaces with an integral over a continuous density:

$$I_{ij} = \int \mathrm{d}^3\vec{l}\rho(\vec{l})\left[\vec{l}^2\delta_{ij} - l_i l_j\right]. \tag{12.21}$$

We can classify rigid objects according to their moments of inertia.

$I_1 \neq I_2 \neq I_3$	asymmetric top
$I_1 = I_2 \neq I_3$	symmetric top
$I_1 = I_2 = I_3$	spherical top

Figure 30: Classification of rigid objects.

12.1.1 Examples of moments of inertia

Example 12.1. Consider a system of particles in the same line.
The tensor of inertia is

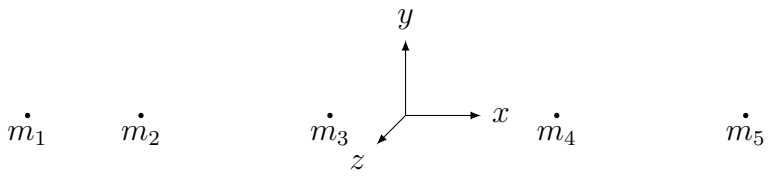


Figure 31: System of particles in the same line.

$$I_{jk} = \sum_a m_a\left(r_a^2\delta_{jk} - r_{ak}r_{aj}\right),$$

but $r_{ai} = r_a \delta_{ix}$, thus

$$\begin{aligned} I_{jk} &= \sum_a m_a \left(r_a^2 \delta_{jk} - r_a^2 \delta_{kx} \delta_{jx} \right) \\ &= \sum_a m_a r_a^2 \left(\delta_{jk} - \delta_{jx} \delta_{kx} \right). \end{aligned}$$

This gives us

$$\begin{aligned} I_{xx} &= 0 \\ I_{yy} &= \sum_a m_a r_a^2 \\ I_{zz} &= \sum_a m_a r_a^2 \\ I_{xy} &= \sum_a m_a r_a^2 \left(\cancelto{0}{\delta_{xy}} - r_a^2 \delta_{xx} \cancelto{0}{\delta_{yx}} \right) = 0 \\ I_{xz} &= \sum_a m_a r_a^2 \left(\cancelto{0}{\delta_{xz}} - r_a^2 \delta_{xx} \cancelto{0}{\delta_{zx}} \right) = 0 \\ I_{yz} &= \sum_a m_a r_a^2 \left(\cancelto{0}{\delta_{yz}} - r_a^2 \cancelto{0}{\delta_{xy}} \cancelto{0}{\delta_{zx}} \right) = 0. \end{aligned}$$

For two atoms,

$$I_2 = I_3 = m_1 r_1^2 + m_2 r_2^2.$$

We want to express this in terms of the distance d of the atoms. We start

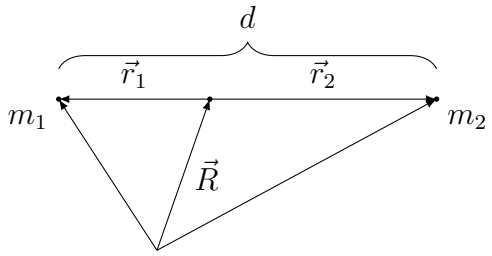


Figure 32: Two atoms with mass m_1 and m_2 and distance d.

with the center of mass equation:

$$\begin{aligned} 0 &= m_1\vec{r}_1 + m_2\vec{r}_2 \\ \Rightarrow \vec{r}_2 &= -\frac{m_1}{m_2}\vec{r}_1. \end{aligned}$$

We obtain for the distance:

$$\begin{aligned} \vec{d} &= \vec{r}_1 - \vec{r}_2 \\ &= \vec{r}_1\left(1 + \frac{m_1}{m_2}\right) \\ &= \vec{r}_1\frac{m_1 + m_2}{m_2} \end{aligned}$$

and thus

$$\boxed{\vec{r}_1 = \frac{m_2}{m_1 + m_2}\vec{d}}$$

$$\boxed{\vec{r}_2 = -\frac{m_1}{m_1 + m_2}\vec{d}}.$$

Then, the moments of inertia are

$$\begin{aligned} I_{2,3} &= m_1 r_1^2 + m_2 r_2^2 \\ &= \left[m_1\left(\frac{m_2}{m_1 + m_2}\right)^2 + m_2\left(\frac{m_1}{m_1 + m_2}\right)^2\right] d^2 \\ &= \frac{m_1 + m_2}{(m_1 + m_2)^2} \cdot m_1 m_2 \\ \Rightarrow &\boxed{I_{yy} = I_{zz} = \frac{m_1 m_2}{m_1 + m_2} d^2} \end{aligned}$$

Now, recall

$$I_{jk} = \sum_a m_a \left(r_a^2 \delta_{jk} - r_{aj} r_{ak}\right)$$

and let us shift

$$\vec{r}_a \to \vec{r}_a + \vec{c},$$

then

$$I'_{jk} = \sum_a \left[(\vec{r}_a + \vec{c}) - (r_{aj} + c_j)(r_{ak} + c_k)\right].$$

The linear term is

$$\sum_a m_a \left[(2\vec{r}_a \cdot \vec{c})\, \delta_{jk} - (r_{aj} c_k - r_{ak} c_j) \right],$$

but it disappears since

$$\sum_a m_a \vec{r}_a = 0$$

and therefore

$$\begin{aligned} I'_{jk} &= \sum_a m_a \left[(\vec{r}_a + \vec{c})^2 \delta_{jk} - (\vec{r}_a + \vec{c})_j (\vec{r}_a + \vec{c})_k \right] \\ &= I_{jk} + \left(\sum_a m_a \right) \left[\vec{c}^2 \delta_{jk} - c_j c_k \right]. \end{aligned}$$

Example 12.2. Consider two points with same mass m_1 and distance a from each other shifted by λ from the x-axis and another point with mass m_2 according to figure 33. Then, the center of mass equation gives us a

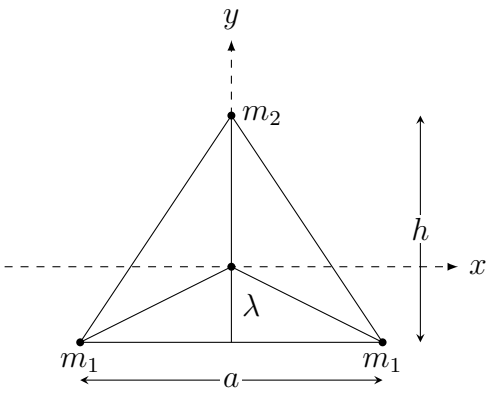


Figure 33: Two masses m_1 with the same distance from the center of mass and one mass m_2 with y-distance h.

relation for λ and h:

$$\begin{aligned} (h - \lambda) m_2 &= 2 m_1 \lambda \\ \Rightarrow \lambda &= \frac{m_2}{2m_1 + m_2} h. \end{aligned}$$

The moments of inertia are

$$\begin{aligned} I_{xx} &= \sum_a m_a \left(r_a^2 \delta_{xx} - r_{ax}^2\right) \\ &= \sum_a m_a r_{ax}^2 \\ &= 2m_1 \left(\frac{a}{2}\right)^2 \\ &= \frac{1}{2} m_1 a^2 \end{aligned}$$

$$\begin{aligned} I_{yy} &= \sum_a m_a r_{ay}^2 \\ &= m_2 \left(h - \lambda\right)^2 + 2m_1 \lambda^2 \\ &= \left[2m_1 \left(\frac{m_2}{m}\right)^2 + m_2 \left(\frac{2m_1}{m}\right)^2\right] h^2 \\ &= \frac{2m_1 m_2}{m} h^2 (m_2 + 2m_1) \\ &= \frac{2m_1 m_2}{m} h^2 \end{aligned}$$

$$\begin{aligned} I_{zz} &= \sum_a m_a \left(r_a^2 \delta_{zz} - r_{az}^2\right) \\ &= \sum_a m_a \left(r_{ax}^2 + r_{ay}^2\right) \\ &= I_{yy} + I_{xx}. \end{aligned}$$

Exercise 12.2. What are the moments of inertia for a cone?

Example 12.3. What is a frequency of a compound pendulum?
Let α, β, γ be the angles between the principal axes and the axis rotation. The velocity of the center of mass is $v = l\dot{\phi}$. Then, the Lagrangian is

$$\begin{aligned} L &= \frac{1}{2} m v^2 + \frac{1}{2} \left(I_1 \omega_1^2 + I_2 \omega_2^2 + I_3 \omega_3^2\right) - V \\ &= \frac{1}{2} m l^2 \dot{\phi}^2 + \frac{1}{2} \left(I_1 \cos^2 \alpha + I_2 \cos^2 \beta + I_3 \cos^2 \gamma\right) \dot{\phi}^2 - V(\phi) \end{aligned}$$

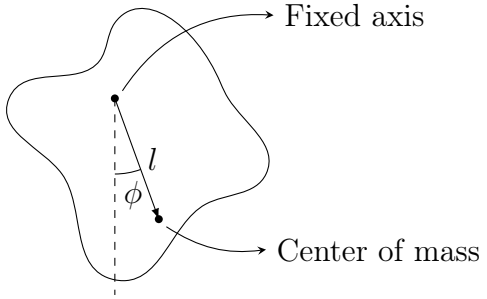


Figure 34: Compound pendulum.

where

$$\begin{aligned} V(\phi) &= \int \mathrm{d}^3\vec{x}\rho(\vec{x})gz \\ &= mgz \\ &= -mgl\cos\phi \\ &\approx -mgl + \frac{1}{2}mgl\phi^2 + \dots \end{aligned}$$

and thus

$$L = \frac{1}{2}\left[ml^2 + I_1\cos^2\alpha + I_2\cos^2\beta + I_3\cos^2\gamma\right]\dot{\phi}^2 - \frac{1}{2}mgl\phi^2 + \dots .$$

The frequency of the pendulum is

$$\begin{aligned} \omega &= \sqrt{\frac{k}{m}} \\ &= \sqrt{\frac{mgl}{ml^2 + I_1\cos^2\alpha + I_2\cos^2\beta + I_3\cos^2\gamma}}. \end{aligned}$$

12.2 Angular momentum

The angular momentum of a rigid body is

$$\begin{aligned} \overrightarrow{M} &= \sum_a m_a\left(\vec{r}_a \times \dot{\vec{r}}_a\right) \\ &= \sum_a m_a\left(\left(\overrightarrow{R} + \vec{l}_a\right) \times \left(\dot{\overrightarrow{R}} + \vec{\omega} \times \vec{l}_a\right)\right). \end{aligned}$$

In the center of mass frame,

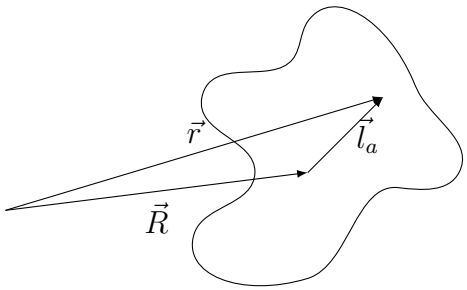


Figure 35: Rigid body.

$$\sum_a m_a \vec{l}_a = 0,$$

the angular momentum is

$$\overrightarrow{M} = \sum_a m_a \vec{l}_a \times \left(\vec{\omega} \times \vec{l}_a \right). \tag{12.22}$$

We have

$$\begin{aligned} \left[\vec{a} \times \left(\vec{b} \times \vec{a} \right) \right]_i &= \epsilon_{ijk} a_j \epsilon_{klm} b_l a_m \\ &= \epsilon_{kij} \epsilon_{klm} a_j a_m b_l \\ &= \begin{vmatrix} \delta_{il} & \delta_{in} \\ \delta_{jl} & \delta_{jn} \end{vmatrix} a_j b_l a_m \\ &= b_i a^2 - a_i \left(\vec{a} \cdot \vec{b} \right) \\ \Rightarrow \vec{a} \times \left(\vec{b} \times \vec{a} \right) &= \vec{b} a^2 - \vec{a} \left(\vec{a} \cdot \vec{b} \right). \end{aligned} \tag{12.23}$$

Thus,

$$\begin{aligned}
\overrightarrow{M} &= \sum_a m_a \left[l_a^2 \overrightarrow{\omega} - \vec{l}_a \left(\overrightarrow{\omega} \cdot \vec{l}_a \right) \right] \\
M_i &= \sum_a m_a \left[l_a^2 \omega_i - l_{ai} \left(l_{aj} \omega_j \right) \right] \\
&= \left(\sum_a m_a \left[l_a^2 \delta_{ij} - l_{ai} l_{aj} \right] \right) \omega_j \\
\Rightarrow & \boxed{M_i = I_{ij} \omega_j}. \qquad (12.24)
\end{aligned}$$

Changing our axes to the principal axes of inertia, we have

$$M_i = I_i \omega_i. \qquad (12.25)$$

We observe that the angular momentum is not necessarily aligned to the direction of the angular velocity. This happens for a spherical top,

$$I_1 = I_2 = I_3 = I,$$

for which

$$\overrightarrow{M} = I \overrightarrow{\omega}. \qquad (12.26)$$

12.2.1 Precession

We now examine the motion of a symmetrical top,

$$I_1 = I_2 \neq I_3,$$

which moves freely, without the influence of external forces. In the center of mass frame, the motion to the symmetrical top is purely rotational. We can determine it completely by using angular momentum conservation.

One principal axis is the axis of symmetry $\hat{x}_3$, as shown in Figure 37. The other two principal axes are perpendicular to it. Due to the symmetry, we have the freedom to choose their orientation. We choose $\hat{x}_2$-axis to be perpendicular to the plane defined by $\hat{x}_3$ and the angular momentum $\overrightarrow{M}$. Then,

$$\overrightarrow{M} \perp \hat{x}_2 \Rightarrow M_2 \equiv \overrightarrow{M} \cdot \hat{x}_2 = 0,$$

but

$$M_2 = I_2 \omega_2 \Rightarrow \boxed{\omega_2 = 0},$$

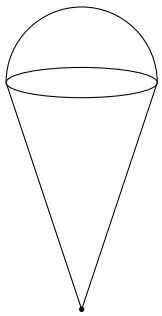


Figure 36: A symmetrical top.

thus $\overrightarrow{M}$, $\vec{\omega}$, $\hat{x}_3$ are all on the same plane.
Consider now the points of the top along the symmetry axis $\hat{x}_3$. Their velocity is

$$\begin{aligned} \overrightarrow{V} &= \vec{\omega} \times \vec{r} \qquad \vec{r} \parallel \hat{x}_3 \\ \Rightarrow \overrightarrow{V} &\perp \left(\vec{\omega}, \overrightarrow{M}, \hat{x}_3\right). \end{aligned}$$

The velocity of the points of the $\hat{x}_3$-axis is perpendicular to the plan of $(\vec{\omega}, \overrightarrow{M}, \hat{x}_3)$. We have

$$\begin{aligned} \overrightarrow{V} &= (\vec{\omega} \times \hat{x}_3)\, x_3 \\ &= \left[\left(\omega_1 \hat{x}_1 + \cancelto{0}{\omega_2} \hat{x}_2 + \omega_3 \hat{x}_3\right) \times \hat{x}_3\right] x_3 \\ &= \left[\omega_1 \left(\hat{x}_1 \times \hat{x}_3\right)\right] x_3 \\ &= \frac{M_1}{I_1} x_3 \hat{x}_2. \end{aligned}$$

Since $\overrightarrow{M}$ is conserved, then

$$\frac{M_1}{I_1} \equiv \omega_1$$

is constant. Therefore, the axis of the top *rotates uniformly around the direction of* $\overrightarrow{M}$, describing a circular cone. We have therefore identified two rotations:

1. Rotation around $\overrightarrow{M}$. This is called *regular precession*. The frequency

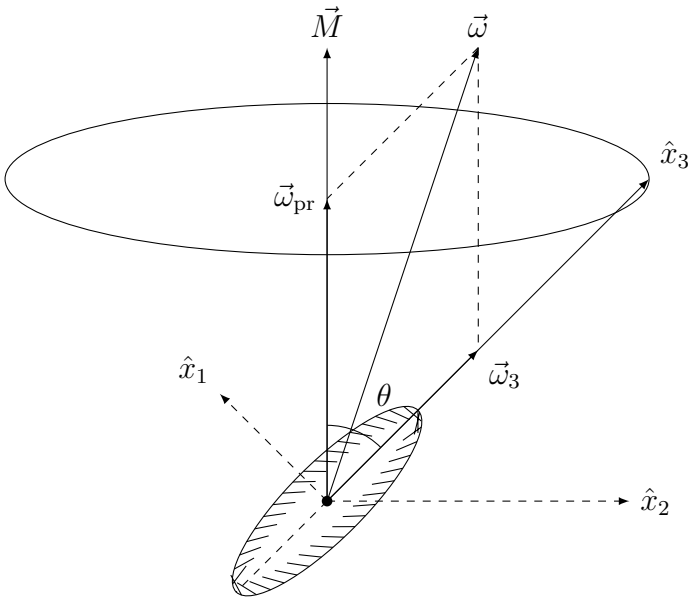


Figure 37: Rigid body with angular momentum $\overrightarrow{M}$ perpendicular to the principle axis $\hat{x}_2$ and angle θ between the angular momentum and the symmetry axis $\hat{x}_3$ of the body.

of this rotation is

$$\begin{aligned}\omega_{\rm pr} &= \omega_1 \sin\theta \\ &= \frac{M_1}{I_1}\sin\theta \\ &= \frac{\frac{M}{\sin\theta}\sin\theta}{I_1} \\ \Rightarrow &\boxed{\omega_{\rm pr} = \frac{M}{I_1}}.\end{aligned}$$

2. Rotation around symmetry axis. The frequency of this rotation is

$$\boxed{\omega_3 = \frac{M_3}{I_3} = \frac{M\cos\theta}{I_3}}.$$

12.3 Equations of motion of rigid body

We need to determine $\vec{R}(t)$ and $\vec{\omega}(t)$. Then geometry of the rigid body will give us the positions $\vec{l}_a(t)$ relative to a reference point at the rigid body. We have

$$\begin{aligned}
\vec{f}_a &= \frac{\mathrm{d}\vec{P}_a}{\mathrm{d}t} \quad \forall \text{particle } a \\
\Rightarrow \sum_a \vec{f}_a &= \frac{\mathrm{d}}{\mathrm{d}t} \sum_a \vec{P}_a \\
&= \frac{\mathrm{d}}{\mathrm{d}t} \sum_a m_a \left(\dot{\vec{R}} + \vec{\omega} \times \vec{l}_a \right) \\
&= \frac{\mathrm{d}}{\mathrm{d}t} \left[\left(\sum_a m_a \right) \vec{V} \right] + \frac{\mathrm{d}}{\mathrm{d}t} \left[\cancelto{0}{\left(\sum_a m_a \vec{l}_a \right)} \times \vec{\omega} \right] \\
&= m \frac{\mathrm{d}}{\mathrm{d}t} \vec{V}.
\end{aligned}$$

The sum of the forces

$$\vec{F} = \sum_a \vec{f}_a$$

includes only external forces, because internal forces cancel out. Indeed, in the absence of external forces, internal forces cannot give a momentum to the body, thus

$$\sum_a \vec{f}_a^{(\mathrm{int})} = 0.$$

Notice that the Lagrangian describing a rigid body is

$$L = \frac{1}{2} m \vec{V}^2 + \frac{1}{2} I_{ij} \omega_i \omega_j - V(\vec{R}, \phi).$$

Then, the Euler-Lagrange equations are

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L}{\partial \vec{V}} - \frac{\partial L}{\partial \vec{R}} &= 0 \\
\Rightarrow m \dot{\vec{V}} &= -\frac{\partial V}{\partial \vec{R}} \equiv \vec{F}
\end{aligned}$$

where we have used the shorthand notation

$$\frac{\partial}{\partial \vec{x}} = \vec{\nabla}_{\vec{x}} \equiv \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3} \right).$$

Now let us look at the rate of change of the total angular momentum. Let us specify the reference frame to be so that the center of mass does not move:

$$\vec{V} = 0.$$

This is not necessarily an inertial frame.
Then,

$$\begin{aligned}
\frac{\mathrm{d}\vec{M}}{\mathrm{d}t} &= \frac{\mathrm{d}}{\mathrm{d}t} \left(\sum_a \vec{l}_a \times \vec{P}_a \right) \\
&= \sum_a \frac{\mathrm{d}}{\mathrm{d}t} \vec{l}_a \times \vec{P}_a + \sum_a \vec{l}_a \times \frac{\mathrm{d}\vec{P}_a}{\mathrm{d}t} \\
&= \sum_a \cancel{\frac{\mathrm{d}}{\mathrm{d}t} \vec{l}_a \times \left(m_a \frac{\mathrm{d}\vec{l}_a}{\mathrm{d}t} \right)}^{0} + \sum_a \vec{l}_a \times \frac{\mathrm{d}\vec{P}_a}{\mathrm{d}t} \\
&= \sum_a \vec{l}_a \times \vec{f}_a.
\end{aligned}$$

The sum

$$\vec{K} \equiv \sum_a \vec{l}_a \times \vec{f}_a$$

is the total torque. The above sum includes only external forces, since the internal forces alone cannot produce angular momentum.

$$\frac{\mathrm{d}\vec{M}}{\mathrm{d}t} = \vec{K}$$

can be derived from the Lagrangian:

$$\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L}{\partial \vec{\omega}} - \frac{\partial L}{\partial \vec{\phi}} = 0,$$

where

$$\begin{aligned}
\frac{\partial L}{\partial \omega_i} &= \frac{\partial}{\partial \omega_i} \left(\frac{1}{2} \omega_i \sum_j I_{ij} \omega_j \right) = \sum_j I_{ij} \omega_j = M_i \\
\frac{\partial L}{\partial \phi_i} &= -\frac{\partial V}{\partial \phi_i}.
\end{aligned}$$

For

$$\begin{aligned}\partial V &= -\sum_a \underbrace{\vec{f}_a \cdot \partial\vec{l}_a}_{\text{work}} \\ &= -\sum_a \vec{f}_a \cdot \left(\partial\vec{\phi} \times \vec{l}_a\right) \\ &= -\partial\vec{\phi} \cdot \underbrace{\sum_a \left(\vec{l}_a \times \vec{f}_a\right)}_{=\overrightarrow{K}} \\ &= -\partial\vec{\phi} \cdot \overrightarrow{K},\end{aligned}$$

we obtain

$$-\frac{\partial V}{\partial\vec{\phi}} = \overrightarrow{K}.$$

The equations of motion

$$\begin{aligned}\frac{\mathrm{d}\overrightarrow{P}}{\mathrm{d}t} &= \overrightarrow{F} \qquad \text{(total force)} \\ \frac{\mathrm{d}\overrightarrow{M}}{\mathrm{d}t} &= \overrightarrow{K} \qquad \text{(total torque)}\end{aligned}$$

have been derived for a reference frame where the axes have a fixed orientation.
However, it is often convenient to use a reference frame with rotating axes. Typically, these are the principal axes of inertia. Consider two coordinate systems where $(\hat{x}_1, \hat{x}_2, \hat{x}_3)$ are fixed and $(\hat{y}_1(t), \hat{y}_2(t), \hat{y}_3(t))$ are rotating.
If we express a vector $\vec{A}$ in terms of the rotating system, we have

$$\vec{A} = \sum_i A_i \hat{y}_i(t).$$

The total derivative is

$$\begin{aligned}\frac{\mathrm{d}\vec{A}}{\mathrm{d}t} &= \sum_i \frac{\mathrm{d}A_i}{\mathrm{d}t}\hat{y}_i(t) + \sum_i A_i(t)\frac{\mathrm{d}\hat{y}_i(t)}{\mathrm{d}t} \\ &= \sum_i \frac{\mathrm{d}A_i}{\mathrm{d}t}\hat{y}_i(t) + \sum_i A_i(t)\omega \times \hat{y}_i(t) \\ \Rightarrow \boxed{\frac{\mathrm{d}\vec{A}}{\mathrm{d}t} &= \left(\frac{\mathrm{d}\vec{A}}{\mathrm{d}t}\right)_{\text{rot}} + \vec{\omega} \times \vec{A}}. \qquad (12.27)\end{aligned}$$

In (12.27), $\left(\dfrac{\mathrm{d}\vec{A}}{\mathrm{d}t}\right)_{\mathrm{rot}}$ is the rate of change of the vector $\vec{A}$ as it is perceived in the rotating frame.

Therefore, in a rotating reference frame with a fixed (or freely moving) center the equations of motion take the form

$$\left(\frac{\mathrm{d}\vec{P}}{\mathrm{d}t}\right)_{\mathrm{rot}} + \vec{\omega} \times \vec{P} = \vec{F}$$
$$\left(\frac{\mathrm{d}\vec{M}}{\mathrm{d}t}\right)_{\mathrm{rot}} + \vec{\omega} \times \vec{M} = \vec{K}$$

where $\vec{\omega}$ is the angular velocity vector for the rotation of the rotating frame. In terms of components, we have

$$m\left(\frac{\mathrm{d}V_1}{\mathrm{d}t} + \omega_2 v_3 - \omega_3 v_2\right) = F_1$$
$$m\left(\frac{\mathrm{d}V_2}{\mathrm{d}t} + \omega_3 v_1 - \omega_1 v_3\right) = F_2$$
$$m\left(\frac{\mathrm{d}V_3}{\mathrm{d}t} + \omega_1 v_2 - \omega_2 v_1\right) = F_3.$$

We can choose the axes of the rotating system to be the principal axes of inertia. Then, using

$$M_1 = I_1\omega_1 \qquad M_2 = I_2\omega_2 \qquad M_3 = I_3\omega_3$$

and substituting into

$$\left(\frac{\mathrm{d}\vec{M}}{\mathrm{d}t}\right)_{\mathrm{rot}} + \vec{\omega} \times \vec{M} = \vec{K},$$

we have

$$\left.\begin{aligned} I_1\frac{\mathrm{d}\omega_1}{\mathrm{d}t} + (I_3 - I_1)\omega_2\omega_3 &= K_1 \\ I_1\frac{\mathrm{d}\omega_2}{\mathrm{d}t} + (I_1 - I_3)\omega_3\omega_1 &= K_2 \\ I_1\frac{\mathrm{d}\omega_3}{\mathrm{d}t} + (I_2 - I_1)\omega_1\omega_2 &= K_3. \end{aligned}\right\} \text{ Euler's equations}$$

12.3.1 Applications: Free rotational motion

Example 12.4 (Symmetrical top)**.** In this case, we have

$$I_1 = I_2 \neq I_3 \qquad \overrightarrow{K} = 0$$

and for

$$\frac{\mathrm{d}\omega_3}{\mathrm{d}t} = 0 \quad \Rightarrow \quad \omega_3 = const,$$

we obtain

$$\frac{\mathrm{d}\omega_1}{\mathrm{d}t} + \Omega\omega_2 = 0 \tag{12.28}$$

$$\frac{\mathrm{d}\omega_2}{\mathrm{d}t} - \Omega\omega_1 = 0 \tag{12.29}$$

with

$$\Omega \equiv \omega_3 \frac{I_3 - I_1}{I_1}.$$

This leads to

$$\dot{\omega}_1 = -\Omega\omega_2$$
$$\dot{\omega}_2 = \Omega\omega_1$$

and we can decouple this system of differential equations:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}(\omega_1 + \mathrm{i}\omega_2) &= \Omega(-\omega_2 + \mathrm{i}\omega_1) \\
&= \mathrm{i}\Omega(\omega_1 + \mathrm{i}\omega_2) \\
\Rightarrow \omega_1 + \mathrm{i}\omega_2 &= A\mathrm{e}^{\mathrm{i}\Omega t} \\
&= A(\cos\omega t + \mathrm{i}\sin\omega t) \\
\Rightarrow \boxed{\omega_1 = A\cos\omega t \qquad \omega_2 = A\sin\omega t}&.
\end{aligned}$$

Notice that

$$\omega_1^2 + \omega_2^2 + \omega_3^2 = A^2 + \omega_3^2 = \text{const.} \quad \Rightarrow \quad |\vec{\omega}| = \text{const.},$$

i.e. the magnitude of the angular velocity is constant.
Since ω_3 is constant too, then $\vec{\omega}$ rotates around the symmetry axis of the top with angular velocity Ω.

Example 12.5 (Asymmetrical top)**.** Let us now consider the general case of an asymmetric top:

$$I_3 > I_2 > I_1.$$

Energy and angular momentum conservation give

$$\begin{aligned} \frac{1}{2}I_1\omega_1^2 + \frac{1}{2}I_2\omega_2^2 + \frac{1}{2}I_3\omega_3^2 &= E \\ \Rightarrow \frac{M_1^2}{I_1} + \frac{M_2^2}{I_2} + \frac{M_3^2}{I_3} &= 2E \end{aligned}$$

and

$$M_1^2 + M_2^2 + M_3^2 = M^2.$$

This is the intersection of *a sphere and an ellipsoid*.
We can prove easily the inequalities

$$\boxed{2EI_1 \leq M^2 \leq 2EI_3}.$$

If $M^2 = 2EI_1$ (smallest value),

$$M_2 = M_3 = 0.$$

For $M^2 = 2EI_3$,

$$\begin{aligned} \frac{M_1^2}{I_1} + \frac{M_2^2}{I_2} + \frac{M_3^2}{I_3} &= 2E \\ M_1^2 + M_2^2 + M_3^2 &= M^2. \end{aligned}$$

For $M^2 = 2EI_2$,

$$\begin{aligned} \frac{M_1^2}{I_1} + \frac{M_2^2}{I_2} + \frac{M_3^2}{I_3} &= 2E \\ \frac{M_1^2}{I_2} + \frac{M_2^2}{I_2} + \frac{M_3^2}{I_2} &= 2E \\ \Rightarrow M_1^2\left(\frac{1}{I_1} - \frac{1}{I_2}\right) + M_3^2\left(\frac{1}{I_3} - \frac{1}{I_2}\right) &= 0 \\ \Rightarrow M_1^2 &= M_3^2. \end{aligned}$$

12.3.2 Euler angles

Euler equations give the angular velocity $\omega(t)$ in a rotating system with axes the principal axes of inertia. What is then the motion of this rotating system as we observe it from a fixed-axes system?
To relate a fixed and a rotating system with a common center, we can use what is called *Euler angles*.

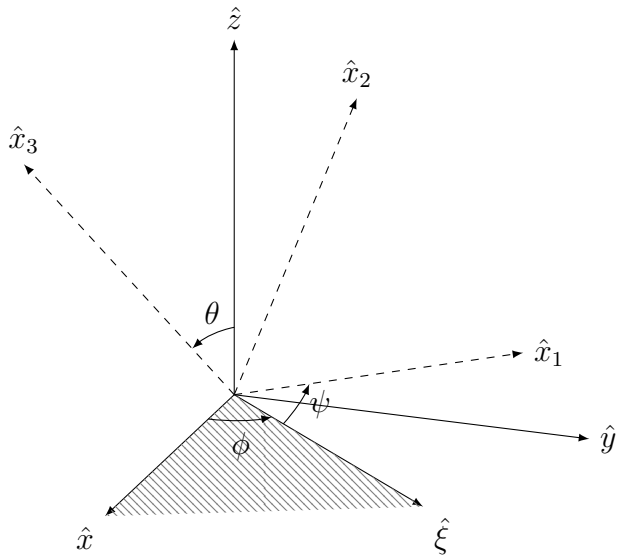


Figure 38: Rotation of the coordinate system $(\hat{x}, \hat{y}, \hat{z})$ by the Euler angles ϕ, θ, ψ to the system $(\hat{x}_1, \hat{x}_2, \hat{x}_3)$.

(x, y, z) is a fixed axes system and (x_1, x_2, x_3) is a rotating one. To go from (x, y, z) to (x_1, x_2, x_3) we perform a series of rotations.

1. Consider the intersection $\overrightarrow{ON}$ of the (x_1, x_2)-plane and the (x, y)-plane. Align the x-axis to $\overrightarrow{ON}$ by rotating counter-clockwise by an angle ϕ. We then go to coordinates

$$(x, y, z) \to (\xi, \eta, z)$$
$$\begin{pmatrix} \xi \\ \eta \\ z \end{pmatrix} = \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

2. Then we rotate around the $\xi(\equiv \overrightarrow{ON})$-axis counter-clockwise:

$$\begin{pmatrix} \xi \\ \eta' \\ z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \xi \\ \eta \\ z \end{pmatrix}.$$

3. Finally, we rotate around the z'-axis by ψ:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \xi \\ \eta' \\ z' \end{pmatrix}.$$

Collectively, we can write:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

We can now relate the angular velocity

$$\vec{\omega} = (\dot{\theta}_1, \dot{\theta}_2, \dot{\theta}_3)$$

as seen in the (x_1, x_2, x_3)-frame in terms of the Euler angles and angular velocities.
For the Euler angular velocities we have:

$$\begin{aligned} \dot{\vec{\phi}} &= \dot{\phi}\hat{z} \\ \dot{\vec{\theta}} &= \dot{\theta}\hat{\xi} \\ \dot{\vec{\psi}} &= \dot{\psi}\hat{z}'. \end{aligned}$$

We want to express these rotation axes in terms of $\hat{x}_1$, $\hat{x}_2$ and $\hat{x}_3$:

$$\begin{aligned} \hat{\xi} &= (\hat{\xi}\cdot\hat{x}_1)\cdot\hat{x}_1 + (\hat{\xi}\cdot\hat{x}_2)\cdot\hat{x}_2 + \cancelto{0}{(\hat{\xi}\cdot\hat{x}_3)}\cdot\hat{x}_3 \\ &= \cos\psi\hat{x}_1 + \cos\left(\psi + \frac{\pi}{2}\right)\hat{x}_2 \\ &= \cos\psi\hat{x}_1 - \sin\psi\hat{x}_2 \end{aligned}$$

$$\begin{aligned}\hat{z} &= \cos\theta\hat{x}_3 + \sin\theta\cos\left(\frac{\pi}{2} - \psi\right)\hat{x}_1 + \sin\theta\cos\psi\hat{x}_2 \\ &= \cos\theta\hat{x}_3 + \sin\theta\sin\psi\hat{x}_1 + \sin\theta\cos\psi\hat{x}_2\end{aligned}$$

$$\hat{z}' = \hat{x}_3,$$

then

$$\begin{aligned}\vec{\omega} &= \dot{\vec{\theta}} + \dot{\vec{\phi}} + \dot{\vec{\psi}} \\ &= (\dot{\phi}\sin\theta\sin\psi + \dot{\theta}\cos\psi)\hat{x}_1 + (\dot{\psi}\sin\theta\cos\psi - \dot{\theta}\sin\psi)\hat{x}_2 + (\dot{\phi}\cos\theta + \dot{\psi})\hat{x}_3.\end{aligned}$$

13 Hamiltonian mechanics

As we established in Sect. 4.6, the motion of a holonomic system with n degrees of freedom is determined by the Lagrange equations, that is a set of n second order differential equations in n unknown variables q_i, where we specify the initial conditions given by q_i and $\dot{q}_i$ at a given time t.

It has been known since Lagrange's times that higher order differential equations can be reduced to systems of first order differential equations. Thus, we expect that a system of n second order differential equations can be reduced to a system of $2n$ first order differential equations. In general, it is more convenient to deal with a system of first order differential equations, because an approximate solution may be found even when an exact solution is not known. To illustrate this point, let us consider a point whose motion is described by the first order differential equation,

$$\dot{\vec{r}}(t) = \vec{f}\left(\vec{r}(t); t\right) . \tag{13.1}$$

Given an initial condition at a time t_0, eq. (13.1) can be integrated to,

$$\vec{r}(t) = \vec{r}(t_0) + \int_{t_0}^{t} \vec{f}\left(\vec{r}(\tau); \tau\right) \mathrm{d}\tau \, . \tag{13.2}$$

The integral equation (13.2) is more convenient than the differential one (13.1) because it can be approximated. In fact, given the initial condition $\vec{r}(t_0)$, we can build a series of successive approximations through the iterative formula,

$$\vec{r}_{k+1}(t) = \vec{r}(t_0) + \int_{t_0}^{t} \vec{f}\left(\vec{r}_k(\tau); \tau\right) \mathrm{d}\tau \, , \qquad k = 0, 1, 2, \ldots \, . \tag{13.3}$$

For example, let us take the non-linear differential equation in one dimension,

$$\dot{x}(t) = x^2 + t^2 \, , \tag{13.4}$$

with $t_0 = 0$ and initial condition $x(t_0) = x_0(t) = 0$. Then,

$$\begin{aligned} x_1(t) &= \int_0^t \left[x_0^2(\tau) + \tau^2\right] \mathrm{d}\tau = \int_0^t \tau^2 \mathrm{d}\tau = \frac{t^3}{3} \, , \\ x_2(t) &= \int_0^t \left[x_1^2(\tau) + \tau^2\right] \mathrm{d}\tau = \int_0^t \left(\frac{\tau^6}{9} + \tau^2\right) \mathrm{d}\tau = \frac{t^3}{3} + \frac{t^7}{63} \, , \end{aligned} \tag{13.5}$$

and so on. One can continue like this until one gets eq. (13.2) to the desired accuracy. Thus, a first order differential equation can be transformed into an integral equation which can be solved to the desired accuracy.

Another reason to go to first order differential equations is that they have a straightforward geometric interpretation. To see this, let us consider the motion of a particle on the $x-y$ plane, described by a set of first order differential equations,

$$\dot{\vec{r}}(t) = \vec{v}(x, y; t)\,, \tag{13.6}$$

which can be written in Cartesian coordinates as,

$$\begin{aligned} \dot{x} &= f(x, y; t)\,, \\ \dot{y} &= g(x, y; t)\,. \end{aligned} \tag{13.7}$$

Then to find $\vec{r}$ means to look for the curve which at every point has $\dot{\vec{r}}$ as its tangent vector, as in Fig. 39.

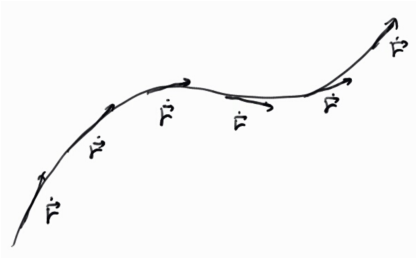


Figure 39: A curve in the $x-y$ plane and the vectors tangent to it.

Conversely, if we have a system of second order differential equations,

$$\begin{aligned} \ddot{x} &= f(x, y, \dot{x}, \dot{y}; t)\,, \\ \ddot{y} &= g(x, y, \dot{x}, \dot{y}; t)\,, \end{aligned} \tag{13.8}$$

in order to associate a point $\ddot{\vec{r}} = \vec{a}(\vec{r}, \dot{\vec{r}}; t)$ to any point $\vec{r}$, we must introduce a four-dimensional space with coordinates $(x, y, \dot{x}, \dot{y})$ whose tangent vector is $(\dot{x}, \dot{y}, \ddot{x}, \ddot{y})$. Thus a geometrical interpretation is much less obvious.

13.1 Hamilton equations

In 1834, William Rowan Hamilton transformed the Lagrange equations,

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0\,, \qquad i = 1, \ldots, n\,, \tag{13.9}$$

into a system of $2n$ first order differential equations. To do so, he introduced the *conjugate momenta* p_i to the generalized coordinates q_i,

$$p_i = \frac{\partial L}{\partial \dot{q}_i}\,, \qquad i = 1, \ldots, n\,, \tag{13.10}$$

and proposed to use q_i, p_i and t as the canonical variables, rather than the variables q_i, $\dot{q}_i$ and t used by Lagrange. If the Jacobian of p_i in eq. (13.10) is non-vanishing,

$$\left|\frac{\partial p_i}{\partial \dot{q}_j}\right| = \left|\frac{\partial^2 L}{\partial \dot{q}_i \partial \dot{q}_j}\right| \neq 0\,, \tag{13.11}$$

then we can invert eq. (13.10), and write that

$$\dot{q}_i = \dot{q}_i(q_k, p_k, t)\,, \tag{13.12}$$

that is, we can express the Lagrange variables in terms of Hamilton variables and viceversa. In particular, for natural systems, for which the Lagrangian is a quadratic function of the generalized velocities, eq. (13.10) expresses a linear relation between conjugate momenta and generalized velocities.

Hamilton introduced a function $H(q_k, p_k, t)$, called the *Hamiltonian*, which is defined through the *Legendre transformation*,

$$H = \sum_{i=1}^{n} p_i \dot{q}_i - L\,, \tag{13.13}$$

and showed that one can transform the Lagrange equations into a system of first order differential equations,

$$\dot{q}_i = \frac{\partial H}{\partial p_i}\,, \qquad \dot{p}_i = -\frac{\partial H}{\partial q_i}\,, \qquad i = 1, \ldots, n\,, \tag{13.14}$$

which are called the *Hamilton equations*.

In order to prove Hamilton equations, we shall use *Donkin's theorem*, which states that given a function of n variables, $X(x_1, \ldots, x_n)$, with a non-vanishing Hessian,

$$\left|\frac{\partial^2 X}{\partial x_i \partial x_j}\right| \neq 0\,, \tag{13.15}$$

and given a set of auxiliary variables,

$$y_i = \frac{\partial X}{\partial x_i}\,, \qquad i = 1, \ldots, n\,, \tag{13.16}$$

there is a transformation which is the inverse of eq. (13.16) and is generated by a function $Y(y_1, \ldots, y_n)$, such that,

$$x_i = \frac{\partial Y}{\partial y_i}\,, \qquad i = 1, \ldots, n\,. \tag{13.17}$$

The function Y generating the inverse transformation is related to the function X by the Legendre transformation,

$$Y = \sum_{i=1}^{n} x_i y_i - X\,. \tag{13.18}$$

Furthermore, if X depends on m more variables, $X = X(x_1, \ldots, x_n; \alpha_1, \ldots, \alpha_m)$, then also Y depends on m more variables, $Y = Y(y_1, \ldots, y_n; \alpha_1, \ldots, \alpha_m)$, with the relation,

$$\frac{\partial X}{\partial \alpha_k} = -\frac{\partial Y}{\partial \alpha_k}\,, \qquad k = 1, \ldots, m\,. \tag{13.19}$$

Proof of Donkin's theorem. Because the Hessian of X, which is the Jacobian of y_i is non-vanishing,

$$\left|\frac{\partial^2 X}{\partial x_i \partial x_j}\right| = \left|\frac{\partial y_i}{\partial x_j}\right| \neq 0\,, \tag{13.20}$$

eq. (13.16) can be inverted, and we can write $x_i = x_i(y_1, \ldots, y_n)$. Let us take a function $Y(y_1, \ldots, y_n)$ to be given by a Legendre transformation (13.18). Then,

$$\begin{aligned}
\frac{\partial Y}{\partial y_i} &= \frac{\partial}{\partial y_i}\left(\sum_{k=1}^{n} x_k y_k - X\right) \\
&= \frac{\partial x_k}{\partial y_i} y_k + x_i - \frac{\partial X}{\partial y_i} \\
&= \frac{\partial x_k}{\partial y_i} y_k + x_i - \sum_k \frac{\partial X}{\partial x_k}\frac{\partial x_k}{\partial y_i} \\
&= x_i\,,
\end{aligned} \tag{13.21}$$

where in the last step above we used the chain rule and eq. (13.16).

If X depends on m more variables, $X = X(x_1, \ldots, x_n; \alpha_1, \ldots, \alpha_m)$, those variables appear in eq. (13.16) and in the inverse transformation as well,

$$x_i = x_i(y_1, \ldots, y_n; \alpha_1, \ldots, \alpha_m)\,. \tag{13.22}$$

The function Y is given by the Legendre transformation (13.18) with the variables x_i given by the inverse transformation (13.22),

$$\begin{aligned} \left.\frac{\partial Y}{\partial \alpha_i}\right|_{y_j} &= \frac{\partial}{\partial \alpha_i}\left(\sum_{k=1}^{n} x_k y_k - X\right) \\ &= \sum_k \frac{\partial x_k}{\partial \alpha_i} y_k - \sum_k \frac{\partial X}{\partial x_k}\frac{\partial x_k}{\partial \alpha_i} - \frac{\partial X}{\partial \alpha_i} \\ &= -\frac{\partial X}{\partial \alpha_i}\,, \end{aligned} \tag{13.23}$$

which completes the proof of Donkin's theorem.

In order to obtain Hamilton equations, we identify,

$$\begin{aligned} X &\equiv L\,, \\ x_i &\equiv \dot{q}_i\,, \\ y_i &\equiv p_i\,, \\ \alpha_i &\equiv q_i, t\,, \end{aligned} \tag{13.24}$$

with $i = 1, \ldots, n$. Then Donkin's theorem implies that given a function $L(q_1, \ldots, q_n; \dot{q}_1, \ldots, \dot{q}_n; t)$, and a set of auxiliary variables (13.10), there is a function $H(q_1, \ldots, q_n; p_1, \ldots, p_n; t)$ which generates the inverse transformation of eq. (13.10),

$$\dot{q}_i = \frac{\partial H}{\partial p_i}\,, \qquad i = 1, \ldots, n\,. \tag{13.25}$$

The Hamiltonian H is related to the Lagrangian L by the Legendre transformation (13.18),

$$H = \sum_{i=1}^{n} p_i \dot{q}_i - L\,, \tag{13.26}$$

with

$$\begin{aligned} \frac{\partial L}{\partial q_i} &= -\frac{\partial H}{\partial q_i}\,, \\ \frac{\partial L}{\partial t} &= -\frac{\partial H}{\partial t}\,. \end{aligned} \tag{13.27}$$

Lagrange equations (13.9) and eq. (13.10) imply that,

$$\dot{p}_i = \frac{\partial L}{\partial q_i}\,, \tag{13.28}$$

and using eq. (13.27), we get that,

$$\dot{p}_i = -\frac{\partial H}{\partial q_i}\,. \tag{13.29}$$

Eqs. (13.25) and (13.29) form a system of $2n$ first order differential equations, the Hamilton equations (13.14).

Hamilton equations imply that the total time derivative of the Hamiltonian H equals the partial time derivative,

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \sum_i \left(\frac{\partial H}{\partial p_i}\dot{p}_i + \frac{\partial H}{\partial q_i}\dot{q}_i \right) + \frac{\partial H}{\partial t} = \frac{\partial H}{\partial t}\,. \tag{13.30}$$

From eq. (7.1) and from the Legendre transformation (13.13), we realize that for a conservative system the Hamiltonian is the total energy of the system. But for a conservative system the total energy is conserved, so

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \frac{\partial H}{\partial t} = 0\,, \tag{13.31}$$

that is, the Hamiltonian $H = H(q_1, \ldots, q_n; p_1, \ldots, p_n)$ is a constant of motion.

For elementary systems, solving a system of Hamilton equations is not usually simpler than solving the analogous system of Lagrange equations. However, the strength of the Hamilton equations may be appreciated when dealing with complex dynamic systems, like the planetary orbits in celestial mechanics, which can be treated through approximate methods of perturbation theory using Hamiltonian mechanics. Furthermore, Hamilton equations naturally incorporate the concept of symmetry: if a coordinate q is associated to a symmetry, its conjugate momentum p is conserved and the Hamiltonian does not depend on q, effectively reducing the system to $n-1$ variables. That is not apparent with Lagrange equations, which in that case will still depend on n velocities $\dot{q}_i$. Also, although Hamilton equations cannot be used as such in quantum mechanics, their structure inspires the equations of quantum mechanics. Finally, Hamiltonian mechanics has a natural geometric structure in terms of simplectic manifolds (to be defined later). In the following lectures, we will briefly explore some of the benefits of Hamiltonian mechanics listed above.

13.2 The Hamiltonian as a conserved quantity and/or the total energy

The simplest Hamiltonian we can consider is for a particle moving on a plane xy under the action of a central potential. The Lagrangian is

$$L = \frac{m}{2}\left(\dot{x}^2 + \dot{y}^2\right) - V(x, y).$$

The conjugate momenta are

$$p_x = \frac{\partial L}{\partial \dot{x}} = m\dot{x}$$
$$p_y = \frac{\partial L}{\partial \dot{y}} = m\dot{y}$$

which we invert to

$$\dot{x} = \frac{p_x}{m} \qquad \dot{y} = \frac{p_y}{m}$$

such that the Lagrangian becomes

$$L = \frac{1}{2m}\left(p_x^2 + p_y^2\right) - V(x, y).$$

The Hamiltonian is

$$\begin{aligned} H &= p_x\dot{x} + p_y\dot{y} - L \\ &= \frac{1}{m}\left(p_x^2 + p_y^2\right) - \frac{1}{2m}\left(p_x^2 + p_y^2\right) + V(x, y) \\ &= \frac{1}{2m}\left(p_x^2 + p_y^2\right) + V(x, y). \end{aligned}$$

Note that the Hamiltonian is the total energy of the particle, and it is also a conserved quantity, since

$$\frac{\partial L}{\partial t} = -\frac{\partial H}{\partial t} = 0.$$

Let us examine these concept more in detail. We said that for a conservative system, the kinetic and potential energies do not depend on time. Then the Lagrangian does not either,

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

and thus

$$\sum_i \frac{\partial L}{\partial \dot{q}_i}\dot{q}_i - L$$

is a conserved quantity which we identify with the Hamiltonian

$$H = \sum_i p_i \dot{q}_i - L.$$

We just saw that for a particle in central potential, H is also the total energy E, and that is true in fact for any conservative system.

In the tutorials, we showed that for a conservative system, defined as a holonomic system with:

- only stationary contraints,
- the potential energy which does not depend explicitly on time,
- no non-potential forces

the total energy is conserved:

$$\frac{\mathrm{d}E}{\mathrm{d}t} = 0.$$

Let us examine the concepts of *total energy* and *conserved quantity* in the Hamiltonian language. We know that if

$$\frac{\partial L}{\partial t} = -\frac{\partial H}{\partial t} = 0,$$

H is a conserved quantity. Under what conditions is H the total energy of the system?

In the Lagrangian formalism, we established that for a holonomic system, the Euler-Lagrange equations are

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = \tilde{Q}_i$$

with $L = T - V$ and the non-potential force $\tilde{Q}_i(q_j, \dot{q}_j, t)$.
Let us suppose that the non-potential force $\tilde{Q}_i$ (called so because it does not

come from an ordinary potential $V(q_i, t)$) comes actually from a generalized potential $V(q_i, \dot{q}_i, t)$, through the equation

$$\tilde{Q}_i = \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial V}{\partial \dot{q}_i} - \frac{\partial V}{\partial q_i} \tag{13.32}$$

then we can still write the Euler-Lagrange equations as

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0$$

if we require that $L = T - V$. Then (13.32) implies that

$$\tilde{Q}_i = \sum_n \frac{\partial^2 V}{\partial \dot{q}_i \partial \dot{q}_n}\ddot{q}_n + \dots .$$

However, in the generalized force there is no explicit dependence on accelerations $\ddot{q}_n$, since

$$\tilde{Q}_i = \tilde{Q}_i(q_n, \dot{q}_n, t).$$

Thus

$$\frac{\partial^2 V}{\partial \dot{q}_i \partial \dot{q}_n} = 0,$$

so the potential V can be at most linear in the velocities:

$$\begin{aligned} V &= \sum_i a_i \dot{q}_i + U \\ &= V_1 + U \end{aligned}$$

where U is the ordinary potential.
We have already seen an example of such a potential: a particle in an electromagnetic field has potential

$$V = e\phi - \frac{e}{c}\dot{\vec{r}} \cdot \vec{A}.$$

Next, we restrict ourselves to "natural" systems, i.e. systems for which the kinetic energy, and thus the Lagrangian, are at most quadratic in the velocities (so far we have only met "natural" systems, but there is an important "non-natural" system: a relativistic particle). However, apart from

this restriction, the following discussion is general. The kinetic energy of a natural system is

$$T = T_2 + T_1 + T_0$$

with

$$T_2 = \sum_{i,n} a_{in}\dot{q}_i\dot{q}_n \qquad T_1 = \sum_i b_i\dot{q}_i$$

and the Lagrangian is

$$L = T - V = L_2 + L_1 + L_0$$

with

$$L_2 = T_2 \qquad L_1 = T_1 - V_1 \qquad L_0 = T_0 - U.$$

The Hamiltonian is

$$H = \sum_i p_i\dot{q}_i - L = \sum_i \frac{\partial L}{\partial \dot{q}_i}\dot{q}_i - L.$$

Then we use Euler's theorem on homogeneous functions,

$$\sum_i \frac{\partial L_2}{\partial \dot{q}_i}\dot{q}_i = 2L_2 \qquad \sum_i \frac{\partial L_1}{\partial \dot{q}_i}\dot{q}_i = L_1,$$

thus

$$\begin{aligned} H &= 2L_2 + L_1 - (L_2 + L_1 + L_0) \\ &= L_2 - L_0 \end{aligned}$$

which is independent of whether a potential dependent on velocities is present or not. In particular,

$$\boxed{H = T_2 - T_0 + U}.$$

As we saw on the tutorial on the energy of a holonomic system, for a system with constraints which do not depend explicitly on time, $T_0 = 0$, $T = T_2$, then

$$\boxed{H = T + U}.$$

Thus we have established that the most general condition for a natural system to have the Hamiltonian as the total energy is that the constraints do not depend explicitly on time.

Example 13.1. For a particle in an electromagnetic field, the Lagrangian is

$$L = \frac{m\dot{r}^2}{2} - e\phi + \frac{e}{c}\dot{\vec{r}} \cdot \vec{A}.$$

The conjugate momenta are

$$p_x = \frac{\partial L}{\partial \dot{x}} = m\dot{x} + \frac{e}{c}A_x \quad \text{and likewise for } y,z.$$

The Hamiltonian is

$$\begin{aligned} H &= p_x\dot{x} + p_y\dot{y} + p_z\dot{z} - L \\ &= \frac{m\dot{r}^2}{2} + e\phi \\ &= \frac{1}{2m}\left(\vec{p} - \frac{e}{c}\vec{A}\right)^2 + e\phi \end{aligned}$$

i.e. the total energy of the particle.

Now, the difference between the Hamiltonian being a *conserved quantity* or the *total energy* is clear:
A system might have the Hamiltonian

$$H = H(q_i, p_i)$$

which does not depend explicitly on time, then

$$\frac{\partial H}{\partial t} = 0$$

and H is a conserved quantity. However, the constraints might be non-stationary,

$$\vec{r}_m = \vec{r}_m(q_1, \ldots, q_m, t),$$

then

$$H = T_2 - T_0 + U$$

is not the total energy.

13.3 Cyclic coordinates

When studying conservation laws, we saw that if a coordinate q_n is *cyclic*, i.e. such that

$$\frac{\partial L}{\partial q_n} = 0$$

the Euler-Lagrange equations imply that

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_n} = 0.$$

Becuase the conjugate momentum is

$$p_n = \frac{\partial L}{\partial \dot{q}_n}$$

that means that p_n is conserved $p_n = c_n$.

In the Hamiltonian formalism, this means that H depends on $2(n-1)$ variables and a constant c_n, to be fixed by the initial conditions

$$H = H(q_1, \ldots, q_{n-1}; p_1, \ldots, p_{n-1}; c_n; t)$$

since q_n does not occur in the Legendre transform

$$H = \sum_i p_i \dot{q}_i - L(q_1, \ldots, q_{n-1}; \dot{q}_1, \ldots, \dot{q}_n; t)$$

and

$$\dot{p}_n = -\frac{\partial H}{\partial q_n} = 0 \quad \Rightarrow \quad p_n = c_n.$$

Effectively, the system has gone from n to $n-1$ degrees of freedom (DOF). In contrast, in the Lagrangian formalism the system still depends on n DOF:

$$L = L(q_1, \ldots, q_{n-1}; \dot{q}_1, \ldots, \dot{q}_n; t).$$

So a Hamiltonian formulation is particularly convenient when cyclic coordinates are present.

Example 13.2. The motion of a particle in a central potential $V(r)$ in polar coordinates r, θ.

The Lagrangian is

$$L = \frac{m}{2}\left(\dot{r}^2 + r^2\dot{\theta}^2\right) - V(r).$$

Because the system is conservative, the Hamiltonian is

$$H = T + V(r).$$

The conjugate momenta are

$$p_r = \frac{\partial L}{\partial \dot{r}} = m\dot{r}$$
$$p_\theta = \frac{\partial L}{\partial \dot{\theta}} = mr^2\dot{\theta}$$

which can be inverted,

$$\dot{r} = \frac{p_r}{m} \qquad \dot{\theta} = \frac{p_\theta}{mr^2}.$$

Thus, the Hamiltonian is

$$H = \frac{1}{2m}\left(p_r^2 + \frac{p_\theta^2}{r^2}\right) + V(r).$$

The Hamiltonian equations are

$$\dot{r} = \frac{\partial H}{\partial p_r} = \frac{p_r}{m}$$
$$\dot{\theta} = \frac{\partial H}{\partial p_\theta} = \frac{p_\theta}{mr^2}$$
$$\dot{p}_r = -\frac{\partial H}{\partial r} = \frac{p_\theta^2}{mr^3} - \frac{\partial V}{\partial r} \qquad \text{i.e. the equation of the force in a central potential}$$
$$\dot{p}_\theta = -\frac{\partial H}{\partial \theta} = 0 \qquad \text{angular momentum conservation.}$$

Note that θ is cyclic and $p_\theta = M$ is is conserved. The Hamiltonian depends on one DOF and one constant, the angular momentum M:

$$H = H(r; p_r; M).$$

13.4 Hamilton's equations and Hamilton's principle

Hamilton's principle was to require the action

$$S = \int_{t_i}^{t_f} \mathrm{d}t\, L\left[\{q_j(t)\}, \{\dot{q}_j(t)\}; t\right]$$

to be minimal:

$$\delta S = 0$$

where $\{q_i\}$ is a short-hand for $q_1, \ldots, q_n$.
In the Hamiltonian formulation, we require that

$$\delta S = \delta \int_{t_i}^{t_f} \mathrm{d}t \left(\sum_i p_i \dot{q}_i - H\left[\{p_j\}, \{q_j\}, t\right] \right) = 0$$

is minimal on the $4n$-dimensional phase space which depends on the coordinates q and p, and their time derivatives $\dot{q}$ and $\dot{p}$.
It is like in variational calculus: we must minimize the integral

$$I\left[\{q\}, \{p\}\right] = \int_{t_i}^{t_f} \mathrm{d}t\, f\left(\{q\}, \{p\}, \{\dot{q}\}, \{\dot{p}\}, t\right)$$

where $f = \sum_i p_i \dot{q}_i - H$.
Then $\delta I = 0$ leads to the Euler-Lagrange equations

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial f}{\partial \dot{q}_i} - \frac{\partial f}{\partial q_i} = 0 \tag{13.33}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial f}{\partial \dot{p}_i} - \frac{\partial f}{\partial p_i} = 0 \tag{13.34}$$

which, for f as given above, yield:

$$(13.33) \qquad \Rightarrow \qquad \dot{p}_i + \frac{\partial H}{\partial q_i} = 0$$

$$(13.34) \qquad \Rightarrow \qquad -\dot{q}_i + \frac{\partial H}{\partial p_i} = 0.$$

So, in a very concise way, we obtained Hamilton's equations!

13.5 Canonical transformations

The motion of a particle in a plane under the action of a central potential can be described using Cartesian coordinates x, y or polar coordinates r, θ as generalized coordinates. The advantage with polar coordinates is that θ is cyclic.

In general, given a Hamiltonian

$$H(q_1, \ldots, q_n; p_1, \ldots, p_n; t)$$

the coordinates $q_1, \ldots, q_n$ might not be cyclic, but there might be another set of coordinates $Q_1, \ldots, Q_n$ for which some or all of them are cyclic.
It is important then to have a procedure to go from one set $(\{q\}, \{p\})$in the $2n$-dimensional space to another:

$$\begin{aligned} Q_i &= Q_i\left(\{q\}; \{p\}; t\right) \\ P_i &= P_i\left(\{q\}; \{p\}; t\right). \end{aligned}$$

The coordinates Q_i, P_i must also be *canonical*, i.e. there must be a function

$$K(Q_1, \ldots, Q_n; P_1, \ldots, P_n; t)$$

such that

$$\dot{Q}_i = \frac{\partial K}{\partial P_i} \qquad \dot{P}_i = -\frac{\partial K}{\partial Q_i}.$$

So they must satisfy Hamilton's principle

$$\delta \int_{t_i}^{t_f} \mathrm{d}t \left(\sum_i P_i \dot{Q}_i - K\left(\{Q\}; \{P\}; t\right) \right) = 0$$

just like for the old coordinates

$$\delta \int_{t_i}^{t_f} \mathrm{d}t \left(\sum_i p_i \dot{q}_i - H\left(\{q\}; \{p\}; t\right) \right) = 0.$$

Then the two integrands must be the same, up to a variation on the end points, whose integral vanishes:

$$\sum_i p_i \dot{q}_i - H = \sum_i P_i \dot{Q}_i - K + \frac{\mathrm{d}F}{\mathrm{d}t} \tag{13.35}$$

where F cannot be a function of the phase space coordinates $(\{q\},\{p\})$ or $(\{Q\},\{P\})$ since they have null variation on the end points, but it can be a function of a mixture of old and new variables, for example $F = F(\{q\},\{Q\};t)$. Then (13.35) implies that

$$\sum_i p_i\dot{q}_i - H = \sum_i P_i\dot{Q}_i - K + \sum_i \left(\frac{\partial F}{\partial q_i}\dot{q}_i + \frac{\partial F}{\partial Q_i}\dot{Q}_i\right) + \frac{\partial F}{\partial t}.$$

Since q_i and Q_i are independent, so are their time derivatives $\dot{q}_i$ and $\dot{Q}_i$, whose coefficients must vanish. Thus, we obtain the transformation equations:

$$p_i = \frac{\partial F}{\partial q_i} \qquad P_i = -\frac{\partial F}{\partial Q_i}$$

and

$$K = H + \frac{\partial F}{\partial t}.$$

The function F is called the *generating function* of the canonical transformation above. If we set in particular,

$$F(\{q\},\{Q\}) = \sum_{k=1}^{n} q_k Q_k,$$

then the transformation equations are

$$p_i = \frac{\partial F}{\partial q_i} = Q_i$$
$$P_i = -\frac{\partial F}{\partial Q_i} = -q_i$$

i.e. this canonical transformation interchanges momenta and coordinates. In the Hamiltonian formalism momenta and space coordinates are coordinates on an equal footing of the $2n$-dimensional phase space. There is no memory of q being a space coordinate and p a mass times a velocity.

If p_i cannot be taken as a function of q_j and Q_j, it is suitable to take a function of the type $F = F(\{q\},\{P\};t)$ as a generating function. But since in eq. (13.35), $\dot{P}_i$ does not occur, it is more convenient to set

$$F = G(\{q\},\{P\};t) - \sum_k Q_k P_k$$

such that eq. (13.35) becomes

$$\sum_i p_i \dot{q}_i - H = -\sum_i Q_i \dot{P}_i - K + \frac{\mathrm{d}G}{\mathrm{d}t} \tag{13.36}$$

Making the total derivative explicit,

$$\sum_i p_i \dot{q}_i - H = -\sum_i Q_i \dot{P}_i - K + \sum_i \left(\frac{\partial G}{\partial q_i} \dot{q}_i + \frac{\partial G}{\partial P_i} \dot{P}_i \right) + \frac{\partial G}{\partial t},$$

and considering that q_i and P_i are independent, we obtain the transformation equations:

$$p_i = \frac{\partial G}{\partial q_i} \qquad Q_i = \frac{\partial G}{\partial P_i}$$

and

$$K = H + \frac{\partial G}{\partial t}.$$

Note that if we take in particular,

$$G\left(\{q\}, \{P\}\right) = \sum_{k=1}^{n} q_k P_k, \tag{13.37}$$

then the transformation equations are

$$p_i = \frac{\partial G}{\partial q_i} = P_i$$
$$Q_i = \frac{\partial G}{\partial P_i} = q_i$$

and

$$K = H,$$

i.e. the new coordinated equal the old ones: the function (13.37) generates the *identity transformation.*

One can also consider

$$F = F\left(\{p\}, \{Q\}, t\right)$$
$$F = F\left(\{p\}, \{P\}, t\right)$$

as generating functions (a classification of, and relations among, the F's is given in Goldstein's book), and any combinations of them. The classification above is not exhaustive: there might be generating functions which do not conform to any of the types given above.

13.6 A canonical transformation for the harmonic oscillator

Consider a harmonic oscillator in one dimension, with force constant k. The Hamiltonian is

$$H = \frac{p^2}{2m} + \frac{kq^2}{2} = \frac{1}{2m}\left(p^2 + m^2\omega^2 q^2\right)$$

where one can set $k = m\omega^2$ with ω the angular frequency of the oscillation. Because H is a sum of two squares, if we find a canonical transformation

$$\begin{aligned} p &= f(P)\cos Q \\ q &= \frac{f(P)}{m\omega}\sin Q \end{aligned} \tag{13.38}$$

then

$$\begin{aligned} K = H &= \frac{f^2(P)}{2m}\left(\cos^2 Q + \sin^2 Q\right) \\ &= \frac{f^2(P)}{2m} \end{aligned}$$

so that Q is cyclic. We must find a function $f(P)$ which makes the transformation canonical.
As a generating function, we try

$$F = \frac{m\omega q^2}{2}\cot Q,$$

then the transformation equations are

$$\begin{aligned} p &= \frac{\partial F}{\partial q} = m\omega q\cot Q \\ P &= -\frac{\partial F}{\partial Q} = -\frac{m\omega q^2}{2}\left(-1 - \frac{\cos^2 Q}{\sin^2 Q}\right) = \frac{m\omega q^2}{2\sin^2 Q} \end{aligned}$$

which we can solve for q and p:

$$\begin{aligned} p &= \sqrt{2m\omega P}\cos Q \\ q &= \sqrt{\frac{2P}{m\omega}}\sin Q. \end{aligned}$$

Comparing to (13.38), we see that the transformation is canonical if we choose

$$f(P) = \sqrt{2m\omega P}.$$

Then the Hamiltonian is

$$H = \frac{2m\omega P}{2m} = \omega P.$$

The harmonic oscillator is a conservative system, thus the Hamiltonian is the total energy. Since the Hamiltonian is cyclic in Q, the momentum P is constant:

$$P = \frac{E}{\omega}.$$

Hamilton's equation for $\dot{Q}$ is

$$\dot{Q} = \frac{\partial H}{\partial P} = \omega$$

which is trivially integrated:

$$Q(t) = \omega t + \alpha.$$

Returning then to the original variables gives us

$$q(t) = \sqrt{\frac{2E}{m\omega^2}} \sin(\omega t + \alpha)$$
$$p(t) = \sqrt{2mE} \cos(\omega t + \alpha).$$

Accordingly, also the phase space plots look different for the old q, p and the new Q, P variables. Since

$$\frac{q^2}{\frac{2E}{m\omega^2}} + \frac{p^2}{2mE} = 1$$

the (q, p) plot is an ellipse with

major semi-axis	$A = \sqrt{\frac{2E}{m\omega^2}}$
minor semi-axis	$B = \sqrt{2mE}$

and the (Q, P) plot is a straight line.

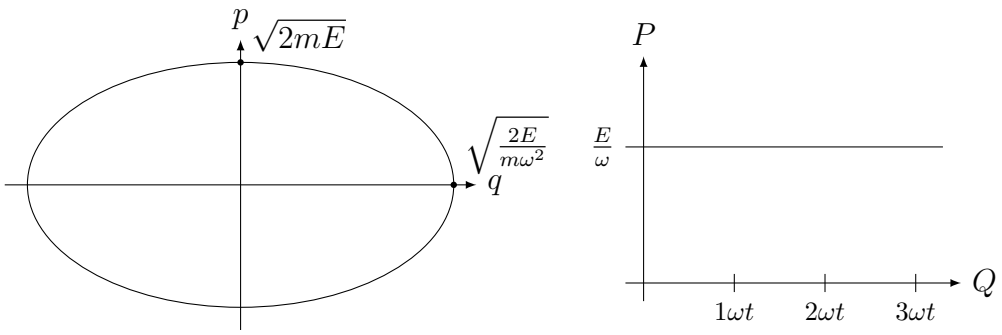


Figure 40: Plots in the phase space of (q, p) (ellipse) and (Q, P) (straight line).

13.7 Canonical transformations in symplectic notation

For a system of n degrees of freedom, we can cast Hamilton's equations in a compact and useful notation with a deep mathematical meaning.

Let us write the coordinates q_i and the momenta p_i as a vector with $2n$ elements:

$$\omega_i = q_i \qquad \omega_{i+n} = p_i \qquad i = 1, \ldots, n.$$

Likewise, we define

$$\left(\frac{\partial H}{\partial \vec{\omega}}\right)_i = \frac{\partial H}{\partial q_i} \qquad \left(\frac{\partial H}{\partial \vec{\omega}}\right)_{i+n} = \frac{\partial H}{\partial p_i}.$$

Then we introduce the $2n \times 2n$ *symplectic* matrix

$$\mathbf{J} = \left(\begin{array}{c|c} 0 & \mathbf{1}_n \\ \hline -\mathbf{1}_n & 0 \end{array}\right) \qquad \text{with } \mathbf{1}_n \text{ the } n \times n \text{ identity matrix.}$$

Note that the transpose matrix equals the inverse

$$\mathbf{J}^T = \mathbf{J}^{-1} = \left(\begin{array}{c|c} 0 & -\mathbf{1}_n \\ \hline \mathbf{1}_n & 0 \end{array}\right) = -\mathbf{J}$$

thus

$$\mathbf{J}\mathbf{J}^T = \mathbf{J}^T\mathbf{J} = \left(\begin{array}{c|c} \mathbf{1}_n & 0 \\ \hline 0 & \mathbf{1}_n \end{array}\right) = \mathbf{1}_{2n}$$

and

$$\mathbf{J}^2 = -\mathbf{1}_{2n}.$$

Then Hamilton's equations can be written as

$$\dot{\vec{\omega}} = \mathbf{J}\frac{\partial H}{\partial \vec{\omega}}$$

or in components:

$$\begin{pmatrix} \dot{q}_i \\ \hline \dot{p}_i \end{pmatrix} = \left(\begin{array}{c|c} 0 & \mathbf{1}_n \\ \hline -\mathbf{1}_n & 0 \end{array}\right) \begin{pmatrix} \dfrac{\partial H}{\partial q_i} \\ \hline \dfrac{\partial H}{\partial p_i} \end{pmatrix} \quad \Rightarrow \quad \begin{cases} \dot{q}_i = \dfrac{\partial H}{\partial p_i} \\ \dot{p}_i = -\dfrac{\partial H}{\partial q_i} \end{cases} .$$

Now let us consider a *restricted* canonical transformation (i.e. one with no explicit dependence on time):

$$\begin{aligned} Q_i &= Q_i\left(\{q\},\{p\}\right) \\ P_i &= P_i\left(\{q\},\{p\}\right). \end{aligned} \tag{13.39}$$

Because there is no dependence on time, the new Hamiltonian will equal the old one,

$$K = H.$$

The time derivative of Q_i is

$$\begin{aligned} \dot{Q}_i &= \sum_j \left(\frac{\partial Q_i}{\partial q_j}\dot{q}_j + \frac{\partial Q_i}{\partial p_j}\dot{p}_j\right) \\ &= \sum_j \left(\frac{\partial Q_i}{\partial q_j}\frac{\partial H}{\partial p_j} - \frac{\partial Q_i}{\partial p_j}\frac{\partial H}{\partial q_j}\right). \end{aligned} \tag{13.40}$$

We can also invert (13.39),

$$\begin{aligned} q_i &= q_i\left(\{Q\},\{P\}\right) \\ p_i &= p_i\left(\{Q\},\{P\}\right) \end{aligned}$$

and write the Hamiltonian as a function of Q and P. Thus, we can write

$$\frac{\partial H}{\partial P_i} = \sum_j \left[\frac{\partial H}{\partial q_j}\frac{\partial q_j}{\partial P_i} + \frac{\partial H}{\partial p_j}\frac{\partial p_j}{\partial P_i}\right]$$

and equating it to (13.40), we obtain

$$\left(\frac{\partial Q_i}{\partial q_j}\right)_{q,p} = \left(\frac{\partial p_j}{\partial P_i}\right)_{Q,P} \qquad \left(\frac{\partial Q_i}{\partial p_j}\right)_{q,p} = -\left(\frac{\partial q_j}{\partial P_i}\right)_{Q,P} \tag{13.41}$$

where the indices remind us that on the l.h.s. of the equations Q_i is a function of q, p and on the r.h.s. p_j is a function of Q, P.
A comparison of $\dot{P}_i$ with $\dfrac{\partial H}{\partial Q_i}$ yields

$$\left(\frac{\partial P_i}{\partial q_j}\right)_{q,p} = -\left(\frac{\partial p_j}{\partial Q_i}\right)_{Q,P} \qquad \left(\frac{\partial P_i}{\partial p_j}\right)_{q,p} = \left(\frac{\partial q_j}{\partial Q_i}\right)_{Q,P}. \tag{13.42}$$

Eqs. (13.41) and (13.42) characterize a (restricted) canonical transformation, and are equivalent to the method of the generating function introduced before. Thus, a canonical transformation can be obtained finding a suitable generating function or through eqs. (13.41) and (13.42). Although the procedure above has been displayed for restricted canonical transformations it is not less general, because it can be extended to canonical transformations in general.

Remembering that in symplectic notation, Hamilton's equations are

$$\dot{\vec{\omega}} = \mathbf{J}\frac{\partial H}{\partial \vec{\omega}} \quad \text{with } \mathbf{J} = \left(\begin{array}{c|c} 0 & \mathbf{1} \\ \hline -1 & 0 \end{array}\right).$$

The restricted canonical transformation can be written as

$$\vec{\Omega} = \vec{\Omega}(\vec{\omega}).$$

Then, taking the time derivative

$$\dot{\vec{\Omega}} = \mathbf{M}\dot{\vec{\omega}}$$

or, in components,

$$\dot{\Omega}_i = \frac{\partial \Omega_i}{\partial \omega_j}\dot{\omega}_j \quad \text{with } M_{ij} = \frac{\partial \Omega_i}{\partial \omega_j}.$$

Using Hamilton's equations, we write

$$\dot{\vec{\Omega}} = \mathbf{MJ}\frac{\partial H}{\partial \vec{\omega}}.$$

By the inverse transformation, we can see H as a function of Ω_i and write

$$\frac{\partial H}{\partial \omega_i} = \frac{\partial H}{\partial \Omega_j}\frac{\partial \Omega_j}{\partial \omega_i} \quad \text{or} \quad \frac{\partial H}{\partial \vec{\omega}} = \mathbf{M}^T\frac{\partial H}{\partial \vec{\Omega}},$$

then

$$\dot{\vec{\Omega}} = \mathbf{MJM}^T\frac{\partial H}{\partial \vec{\Omega}}. \tag{13.43}$$

But the Hamiltonian is the same in the old and the new variables, and Hamilton's equations do not change form, so

$$\dot{\vec{\Omega}} = \mathbf{J}\frac{\partial H}{\partial \vec{\Omega}}$$

and equating with (13.43), we obtain the condition

$$\boxed{\mathbf{MJM}^T = \mathbf{J}}$$

which defines the symplectic matrices, and is called the *symplectic* condition for a canonical transformation.

13.8 Infinitesimal canonical transformations

The canonical transformation

$$\vec{\Omega} = \vec{\Omega}\left(\vec{\omega}, t\right)$$

with

$$\omega_i = q_i \qquad \omega_{i+n} = p_i \qquad \Omega_i = Q_i \qquad \Omega_{i+n} = P_i$$

changes continuously with time.
If the transformation $\vec{\omega} \to \vec{\Omega}\left(t\right)$ is canonical, so is the transformation $\vec{\omega} \to \vec{\Omega}\left(t_0\right)$, thus $\vec{\Omega}\left(t_0\right) \to \vec{\Omega}\left(t\right)$ should be canonical as well.
To show that, let us consider an infinitesimal canonical transformation (ICT)

$$\vec{\Omega} = \vec{\omega} + \delta\vec{\omega},$$

or in components,

$$Q_i = q_i + \delta q_i \qquad P_i = p_i + \delta p_i.$$

Let us take a generating function of type

$$G\left(\{q\},\{P\},t\right)$$

as a continuous deformation of the identity transformation:

$$G = \sum_k q_k P_k + \epsilon \tilde{G}\left(\{q\},\{P\},t\right) \quad \text{with } \epsilon \ll 1.$$

Then, the transformation equations are

$$p_i = \frac{\partial G}{\partial q_i} = P_i + \epsilon \frac{\partial \tilde{G}}{\partial q_i}$$
$$Q_i = \frac{\partial G}{\partial P_i} = q_i + \epsilon \frac{\partial \tilde{G}}{\partial P_i}$$

so the deformations of the canonical variables are

$$\delta p_i = P_i - p_i = -\epsilon \frac{\partial \tilde{G}}{\partial q_i}$$
$$\delta q_i = Q_i - q_i = \epsilon \frac{\partial \tilde{G}}{\partial P_i} = \epsilon \frac{\partial \tilde{G}}{\partial p_i}$$

where in the equation for δq_i we replaced P_i with p_i because the difference is $O(\epsilon)$. Likewise, the generating function of the ICT, $\tilde{G}\left(\{q\},\{P\},t\right)$, can be replaced by $\tilde{G}\left(\{q\},\{p\},t\right)$.
In symplectic notation,

$$\delta\vec{\omega} = \epsilon \mathbf{J} \frac{\partial \tilde{G}}{\partial \vec{\omega}}. \tag{13.44}$$

Now, as an ICT consider

$$\vec{\Omega}(t) = \vec{\Omega}\left(t_0 + \delta t\right).$$

The symplectic matrix is

$$M_{ij} = \frac{\partial \Omega_i}{\partial \omega_j} = \delta_{ij} + \frac{\partial \delta\omega_i}{\partial \omega_j}$$

or using (13.44)

$$\begin{aligned}\mathbf{M} &= \mathbf{1} + \frac{\partial \delta\vec{\omega}}{\partial\vec{\omega}} \\ &= \mathbf{1} + \epsilon\mathbf{J}\frac{\partial^2\tilde{G}}{\partial\vec{\omega}\partial\vec{\omega}}\end{aligned}$$

or in components

$$M_{ij} = \delta_{ij} + \epsilon J_{ik}\frac{\partial^2\tilde{G}}{\partial\omega_k\partial\omega_j}$$

with

$$\frac{\partial^2\tilde{G}}{\partial\omega_k\partial\omega_j} = \frac{\partial^2\tilde{G}}{\partial\omega_j\omega_k}.$$

But

$$\mathbf{J}^T = -\mathbf{J} \quad \Rightarrow \quad J_{ik} = -J_{ki}$$

then

$$\left(M_{ji}\right)^T = M_{ij} = \delta_{ij} - \epsilon\frac{\partial^2\tilde{G}}{\partial\omega_j\partial\omega_k}J_{ki}$$

i.e.

$$\mathbf{M}^T = \mathbf{1} - \epsilon\frac{\partial^2\tilde{G}}{\partial\vec{\omega}\partial\vec{\omega}}\mathbf{J}$$

so

$$\begin{aligned}\mathbf{M}\mathbf{J}\mathbf{M}^T &= \left(\mathbf{1} + \epsilon\mathbf{J}\frac{\partial^2\tilde{G}}{\partial\vec{\omega}\partial\vec{\omega}}\right)\mathbf{J}\left(\mathbf{1} - \epsilon\frac{\partial^2\tilde{G}}{\partial\vec{\omega}\partial\vec{\omega}}\mathbf{J}\right) \\ &= \mathbf{J} + \mathcal{O}\left(\epsilon^2\right)\end{aligned}$$

which is the symplectic condition, so the ICT $\vec{\Omega}(t) = \vec{\Omega}\left(t_0 + \delta t\right)$ is indeed canonical, thus we conclude that any canonical transformation, with or without time dependence, fulfills the symplectic condition.

We have seen that:

- we can continuously transform the canonical variables through a canonical transformation,
- the identity transformation is canonical.

Now, the inverse of a canonical transformation is also a canonical transformation; the composition of two canonical transformations is also a canonical transformation, and that composition is associative:

$$a(bc) = (ab)c.$$

Therefore, the canonical transformations form a (continuous) group.

13.9 Integrals of the equations of motion, Poisson brackets and canonical invariants

The primary goal of a canonical transformation is to find a set of canonical variables with as many cyclic coordinates as possible.

More in general, suppose that $f(\{q\},\{p\},t)$ is an integral of the equations of motion (IEM),

$$\frac{\mathrm{d}f}{\mathrm{d}t} = 0,$$

i.e. f is a conserved quantity:

$$f(\{q\},\{p\},t) = c.$$

For example, we know that for a conservative system the Hamiltonian is a conserved quantity, or IEM, and that for a system with cyclic coordinates q_i, the conjugate momenta p_i are conserved quantities, or IEM.

Suppose that we have k IEM $f_1,\ldots,f_k$ with $k < 2n$, then any arbitrary combination of those k IEM is also an IEM. Thus, it is clear that it suffices to consider only the independent IEM.

If we know $2n$ independent IEM,

$$f_k(\{q\},\{p\},t) = c_k \quad k = 1,\ldots,2n,$$

then the equations of motion can be written as

$$\begin{aligned} q_i &= q_i(c_1,\ldots,c_{2n},t) \\ p_i &= p_i(c_1,\ldots,c_{2n},t) \end{aligned}$$

i.e. the system is completely integrable. Then the broader goal is to find as many independent IEM as possible.

If $f(\{q\},\{p\},t)$ is an IEM, then

$$\frac{\mathrm{d}f}{\mathrm{d}t}=0 \quad \Rightarrow \quad \sum_i\left(\frac{\partial f}{\partial q_i}\dot{q}_i+\frac{\partial f}{\partial p_i}\dot{p}_i\right)+\frac{\partial f}{\partial t}=0,$$

using Hamilton's equations

$$\frac{\partial f}{\partial t}+\sum_i\left(\frac{\partial f}{\partial q_i}\frac{\partial H}{\partial p_i}-\frac{\partial f}{\partial p_i}\frac{\partial H}{\partial q_i}\right)=0$$

$$\text{or} \quad \frac{\partial f}{\partial t}+[f,H]_{q,p}=0,$$

where we have introduced the *Poisson bracket*

$$[\phi,\psi]_{q,p}=\sum_i\left(\frac{\partial\phi}{\partial q_i}\frac{\partial\psi}{\partial p_i}-\frac{\partial\phi}{\partial p_i}\frac{\partial\psi}{\partial q_i}\right).$$

Thus, to say that f is an IEM is equivalent to require that

$$\frac{\partial f}{\partial t}+[f,H]_{q,p}=0.$$

This offers a criterion to test candidate IEM: In order for f to be an IEM, it must fulfill

$$\frac{\partial f}{\partial t}=[H,f]_{q,p}$$

and if a particular f does not depend explicitly on time, it must fulfill

$$[H,f]_{q,p}=0.$$

The Poisson bracket $[\phi,\psi]_{q,p}$ has a symplectic structure. In fact,

$$[\phi,\psi]_{\vec{\omega}}=\left(\frac{\partial\phi}{\partial\vec{\omega}}\right)^T\mathbf{J}\frac{\partial\psi}{\partial\vec{\omega}} \quad \text{for} \quad \omega_i=q_i \qquad \omega_{i+n}=p_i$$

as can be seen by writing $[\phi,\psi]_{\vec{\omega}}$ in components (please check!).
If ϕ, ψ are taken among the variables q_i, p_i themselves, then from the definition of the Poisson bracket

$$[q_j,q_k]_{q,p}=[p_j,p_k]_{q,p}=0$$
$$[q_j,p_k]_{q,p}=\delta_{jk},$$

thus we can write the antisymmetric matrix $\mathbf{J}$ as the Poisson bracket matrix

$$[\vec{\omega},\vec{\omega}]_{\vec{\omega}} = \mathbf{J} \quad \text{with} \quad \left([\vec{\omega},\vec{\omega}]_{\vec{\omega}}\right)_{jk} = [\omega_j,\omega_k]$$

(again, write it explicitly in components).
Now, take a set of new canonical variables,

$$\Omega_i = Q_i \qquad \Omega_{i+n} = P_i,$$

then by definition

$$\left[\vec{\Omega},\vec{\Omega}\right]_{\vec{\omega}} = \left(\frac{\partial\vec{\Omega}}{\partial\vec{\omega}}\right)^T \mathbf{J}\frac{\partial\vec{\Omega}}{\partial\vec{\omega}},$$

but $\dfrac{\partial\vec{\Omega}}{\partial\vec{\omega}}$ is the transformation matrix $\mathbf{M}$, in components

$$\frac{\partial\Omega_i}{\partial\omega_j} = M_{ij},$$

then

$$\left[\vec{\Omega},\vec{\Omega}\right]_{\vec{\omega}} = \mathbf{M}^T\mathbf{J}\mathbf{M}.$$

If the transformation is canonical,

$$\mathbf{M}^T\mathbf{J}\mathbf{M} = \mathbf{J} \quad \Rightarrow \quad \left[\vec{\Omega},\vec{\Omega}\right]_{\vec{\omega}} = \mathbf{J},$$

i.e. the *fundamental* Poisson brackets (i.e. those made out of canonical variables) are invariant under canonical transformations, i.e. are *canonical invariants*.

Now, we show that any Poisson bracket is a canonical invariant.
Consider the Poisson bracket

$$[\phi,\psi]_{\vec{\omega}}.$$

We know that

$$\frac{\partial H}{\partial\vec{\omega}} = \mathbf{M}^T\frac{\partial H}{\partial\vec{\Omega}}.$$

Likewise, we can write

$$\frac{\partial\psi}{\partial\vec{\omega}} = \mathbf{M}^T\frac{\partial\psi}{\partial\vec{\Omega}},$$

analogously

$$\left(\frac{\partial\phi}{\partial\vec{\omega}}\right)^T = \left(\mathbf{M}^T\frac{\partial\phi}{\partial\vec{\Omega}}\right)^T = \left(\frac{\partial\phi}{\partial\vec{\Omega}}\right)^T\mathbf{M},$$

so

$$\begin{aligned}[\phi,\psi]_{\vec{\omega}} &= \left(\frac{\partial\phi}{\partial\vec{\omega}}\right)^T\mathbf{J}\frac{\partial\psi}{\partial\vec{\omega}}\\ &= \left(\frac{\partial\psi}{\partial\vec{\Omega}}\right)^T\mathbf{M}\mathbf{J}\mathbf{M}^T\frac{\partial\psi}{\partial\vec{\Omega}}\\ &= \left(\frac{\partial\psi}{\partial\vec{\Omega}}\right)^T\mathbf{J}\frac{\partial\psi}{\partial\vec{\Omega}}\\ &= [\phi,\psi]_{\vec{\Omega}},\end{aligned}$$

thus any Poisson bracket is a canonical invariant. Then it does not matter which set of canonical variables we use to compute it, and we can drop the index to the right of the $[\cdot,\cdot]$ symbol.
The poisson brackets sport the properties:

(i) $[\phi,\psi] = -[\psi,\phi]$ antisymmetry

(ii) $[a\phi + b\psi,\chi] = a[\phi,\chi] + b[\psi,\chi]$ with a, b constants linearity

(iii) $[\phi\psi,\chi] = [\phi,\chi]\psi + \phi[\psi,\chi]$

(iv) $\dfrac{\partial}{\partial t}[\phi,\psi] = \left[\dfrac{\partial\phi}{\partial t},\psi\right] + \left[\phi,\dfrac{\partial\psi}{\partial t}\right]$

(v) $[[\phi,\psi],\chi] + [[\psi,\chi],\phi] + [[\chi,\phi],\psi] = 0.$

All, but (v), stem directly from the definition of Poisson bracket. (v) is called *Jacobi identity*. To prove it, let us write

$$[\phi,\psi] = \frac{\partial\phi}{\partial\omega_i}J_{ij}\frac{\partial\psi}{\partial\omega_i} \qquad \text{with } J_{ij} = -J_{ji},$$

where sums are understood over repeated indices, usually called dummy in-

dices. Then

$$\begin{aligned}
[[\phi,\psi],\chi] &= \frac{\partial\,[\phi,\psi]}{\partial\omega_i} J_{ij} \frac{\partial\chi}{\partial\omega_j} \\
&= \frac{\partial}{\partial\omega_i}\left[\frac{\partial\phi}{\partial\omega_k} J_{kl} \frac{\partial\psi}{\partial\omega_l}\right] J_{ij} \frac{\partial\chi}{\partial\omega_j} \\
&= \left(\frac{\partial^2\phi}{\partial\omega_i\partial\omega_k} J_{kl} \frac{\partial\psi}{\partial\omega_l} + \frac{\partial\phi}{\partial\omega_k} J_{kl} \frac{\partial^2\psi}{\partial\omega_i\partial\omega_l}\right) J_{ij} \frac{\partial\chi}{\partial\omega_j}.
\end{aligned}$$

Likewise, we can write $[[\psi,\chi],\phi]$ and $[[\chi,\phi],\psi]$. In total, we will have six terms, each with a fourfold sum over i, j, k, l.
In particular,

$$[[\chi,\phi],\psi] = \left(\frac{\partial^2\chi}{\partial\omega_k\partial\omega_j} J_{ji} \frac{\partial\phi}{\partial\omega_i} + \frac{\partial\chi}{\partial\omega_j} J_{ji} \frac{\partial^2\phi}{\partial\omega_k\partial\omega_i}\right) J_{kl} \frac{\partial\psi}{\partial\omega_l}.$$

Let us collect the second derivatives of ϕ:

$$J_{kl}\left(J_{ij} + J_{ji}\right) \frac{\partial^2\phi}{\partial\omega_i\partial\omega_k} \frac{\partial\psi}{\partial\omega_l} \frac{\partial\chi}{\partial\omega_j} = 0$$

where we used that $\partial^2\phi$ is symmetric in its indices and J_{ij} is antisymmetric. Likewise, one shows that the coefficients of $\partial^2\psi$ and $\partial^2\chi$ vanish, which proves Jacobi identity.

Note that the Jacobi identity is not an associative composition rule. In fact

$$\begin{aligned}
[\phi,[\psi,\chi]] &= -\,[[\psi,\chi],\phi] \\
&= [[\phi,\psi],\chi] + [[\chi,\phi],\psi] \\
&\neq [[\phi,\psi],\chi].
\end{aligned}$$

In fact, the Poisson brackets form a non-associative algebra. In particular, the properties of the Poisson brackets (antisymmetry, linearity, Jacobi identity) define a *Lie algebra*.

We have studied in some depth the properties of the Poisson brackets. We can now prove *Jacobi-Poisson Theorem*, which states that if f and g are IEM then also $[f,g]$ is an IEM.

Proof. Since f and g are IEM, we can write

$$\frac{\partial f}{\partial t} + [f, H] = 0 \qquad \frac{\partial g}{\partial t} + [g, H] = 0,$$

then

$$\begin{aligned}\frac{\partial [f, g]}{\partial t} &= \left[\frac{\partial f}{\partial t}, g\right] + \left[f, \frac{\partial g}{\partial t}\right] \\ &= -[[f, H], g] - [f, [g, H]] \\ &= [[H, f], g] + [[g, H], f],\end{aligned}$$

so thanks to Jacobi identity,

$$\frac{\partial [f, g]}{\partial t} + [[f, g], H] = 0,$$

so also $[f, g]$ is an IEM. □

If we have k independent IEM, with $k < 2n$, the repeated application of Poisson brackets between independent IEM may in principle generate the missing IEM.

13.10 Conservation theorems & canonical transformations through the Poisson bracket

In the Lagrangian formalism, we established through Noether's theorem that if the equations of motion are invariant under symmetry transformations there are corresponding conserved quantities. We saw that if the symmetry transformation is a time shift the conserved quantity is the energy; if it is a space shift the conserved quantity is the momentum; if it is a rotation, angular momentum is conserved.

We also saw that continuous symmetry transformations form a group, and that the generators of the transformation form a Lie algebra.

Now we look at the conservation laws through the Hamiltonian formalism and the Poisson brackets.

Recall that a continuous deformation of the identity transformation is generated by the function

$$G = \sum_k q_k P_k + \epsilon \tilde{G}(\{q\}, \{p\}, t).$$

The ICT is

$$\overrightarrow{\Omega} = \vec{\omega} + \delta\vec{\omega} \quad \text{with } \delta\vec{\omega} = \epsilon \mathbf{J} \frac{\partial \tilde{G}}{\partial \vec{\omega}}.$$

Recall that the Poisson bracket is defined as

$$[\phi, \psi] = \left(\frac{\partial \phi}{\partial \vec{\omega}} \right)^T \mathbf{J} \frac{\partial \psi}{\partial \vec{\omega}}.$$

Then in particular,

$$[\vec{\omega}, \psi] = \mathbf{J} \frac{\partial \psi}{\partial \vec{\omega}}$$

so we can write the ICT as

$$\delta\vec{\omega} = \epsilon \left[\vec{\omega}, \tilde{G} \right].$$

Now, suppose that the continuous parameter of the ICT is the time t, such that $\epsilon = \mathrm{d}t$, then the change in the Hamiltonian under the ICT is

$$K = H + \frac{\partial G}{\partial t} = H + \epsilon \frac{\partial \tilde{G}}{\partial t} \quad \Rightarrow \quad \delta H = \mathrm{d}t \frac{\partial \tilde{G}}{\partial t}$$

and we can take the Hamiltonian as the generator $\tilde{G}$. then

$$\delta\vec{\omega} = \mathrm{d}t \left[\vec{\omega}, H \right] = \mathrm{d}t \dot{\vec{\omega}} = \mathrm{d}\vec{\omega}$$

i.e. the motion from a time t to $t + \mathrm{d}t$ is described by an ICT generated by the Hamiltonian. So the Hamiltonian is the generator of the time shifts.

If $\tilde{G}$ is a constant of motion, it generates an ICT which does not change the Hamiltonian. Thus, the constants of motion are the generators of the canonical transformations which leave the Hamiltonian invariant.

We know that if a coordinate q_i is cyclic, the Hamiltonian is independent of q_i, so H will be invariant under an ICT that changes q_i only.
Now, take

$$\tilde{G}\left(\{q\}, \{p\}, t \right) = p_i,$$

then the transformation equations are

$$\begin{aligned} \delta q_j &= \epsilon \delta_{ij} \\ \delta p_j &= 0, \end{aligned}$$

so only q_i changes as required, and the conjugate momentum p_i is a constant of motion.
In symplectic notation, we can write the generator $\tilde{G}$ as

$$\tilde{G}_i = (\mathbf{J}\vec{\omega})_i = J_{ik}\omega_k,$$

then $\delta\vec{\omega} = \epsilon\mathbf{J}\dfrac{\partial\tilde{G}}{\partial\vec{\omega}}$ implies

$$\begin{aligned}\delta\omega_j &= \epsilon J_{jl}\frac{\partial\tilde{G}_i}{\partial\omega_l}\\ &= \epsilon J_{jl}J_{ik}J_{kl}\\ &= \epsilon J_{jk}J_{ik} = \epsilon J_{jk}J^T_{ki}\\ &= \epsilon\delta_{ij}.\end{aligned}$$

So a shift of any canonical variable ω_i is generated by the conjugate variable through $\tilde{G}$ as above. In particular, if $\omega_i = q_i$, $\tilde{G}_i = p_i$; if $\omega_i = p_i$, $\tilde{G}_i = -q_i$. Thus we can say that the momentum is the generator of the space shifts, or in full generality the conjugate variable is the generator of a phase space variable shifts.

Now, consider the ICT which yields a rotation of θ about the z-axis. Through the rotation, the changes in coordinates are

$$\delta x = -y\delta\theta \qquad \delta y = x\delta\theta \qquad \delta z = 0.$$

Also the conjugate momentum rotates in the same way, so

$$\delta p_x = -p_y\delta\theta \qquad \delta p_y = p_x\delta\theta \qquad \delta p_z = 0.$$

The transformation equations are

$$\begin{aligned}\delta p &= -\epsilon\frac{\partial\tilde{G}}{\partial q}\\ \delta q &= \epsilon\frac{\partial\tilde{G}}{\partial p}\end{aligned}$$

and it's easy to check that the rotations above are generated by

$$\tilde{G} = xp_y - yp_x,$$

which is the z-component of angular momentum:

$$\tilde{G} = M_z = (\vec{r} \times \vec{p})_z .$$

So the generating function of an infinitesimal rotation is in general

$$\tilde{G} = \overrightarrow{M} \cdot \vec{n}$$

about some axis $\vec{n}$.

Let us consider now finite canonical transformations. Let us take the ICT of a function $\phi(\{q\}, \{p\}, t)$ where the trajectory in phase space is parametrized by α:

$$\delta\phi = \phi(\vec{\omega} + \delta\vec{\omega}) - \phi(\vec{\omega}) = \left(\frac{\partial\phi}{\partial\vec{\omega}}\right)^T \delta\omega$$

but $\delta\vec{\omega} = \mathrm{d}\alpha \mathbf{J} \dfrac{\partial\tilde{G}}{\partial\vec{\omega}}$ so

$$\delta\phi = \mathrm{d}\alpha \left(\frac{\partial\phi}{\partial\vec{\omega}}\right)^T \mathbf{J} \frac{\partial\tilde{G}}{\partial\vec{\omega}}.$$

Recalling the definition of Poisson bracket, we can write

$$\delta\phi = \mathrm{d}\alpha \left[\psi, \tilde{G}\right] \quad \text{or} \quad \frac{\partial\phi}{\partial\alpha} = \left[\phi, \tilde{G}\right].$$

We can take the variation of $\left[\phi, \tilde{G}\right]$ and likewise we find that

$$\frac{\partial^2\phi}{\partial\alpha^2} = \frac{\partial\left[\phi, \tilde{G}\right]}{\partial\alpha} = \left[\left[\phi, \tilde{G}\right], \tilde{G}\right].$$

Then, if we Taylor expand the function ϕ about some initial value for which $\alpha = 0$, we get

$$\begin{aligned}\phi(\alpha) &= \phi(0) + \left.\frac{\partial\phi}{\partial\alpha}\right|_{\alpha=0} + \left.\frac{\partial^2\phi}{\partial\alpha^2}\right|_{\alpha=0} \frac{\alpha^2}{2} + \dots \\ &= \phi(0) + \left[\phi, \tilde{G}\right]\alpha + \left[\left[\phi, \tilde{G}\right], \tilde{G}\right]\frac{\alpha^2}{2} + \dots\end{aligned}$$

which can be formally exponentiated

$$\phi(\alpha) = \phi(0)\, \mathrm{e}^{\hat{G}\alpha}\Big|_{\alpha=0}$$

where the exponential is just a short-hand for the power series of nested Poisson brackets above. The power series can be operatively used, though.

Example 13.3. Take the ICT of a rotation about the z-axis. Then

$$\delta\vec{\omega} = \mathrm{d}\theta\,[\vec{\omega}, M_z]$$

and we set

$$\delta x = -y\,\mathrm{d}\theta \qquad \delta y = x\,\mathrm{d}\theta \qquad M_z = xP_y - yP_x$$

and it's immediate to check that

$$[x, M_z] = -y \qquad [y, M_z] = x.$$

Then we can build the nested Poisson brackets

$$\begin{aligned}
[[x, M_z], M_z] &= -[y, M_z] = -x \\
[[[x, M_z], M_z], M_z] &= -[x, M_z] = y \\
[[[[x, M_z], M_z], M_z], M_z] &= [y, M_z] = x \quad \text{and so on}
\end{aligned}$$

and the power series for a rotated variable is

$$\begin{aligned}
X &= x - y\theta - x\frac{\theta^2}{2} + y\frac{\theta^3}{3!} + x\frac{\theta^4}{4!} + \dots \\
&= x\left(1 - \frac{\theta^2}{2} + \frac{\theta^4}{4!} + \dots\right) - y\left(\theta - \frac{\theta^3}{3!} + \dots\right) \\
&= x\cos\theta - y\sin\theta
\end{aligned}$$

which corresponds to a finite rotation about the z-axis.

Finally, if the generator of the ICT is the Hamiltonian and the parameter is the time t, the formal power series becomes

$$\phi(t) = \phi(0) + [\phi, H]\,t + [[\phi, H], H]\,\frac{t^2}{2} + [[[\phi, H], H], H]\,\frac{t^3}{3!} + \dots$$

which is formally exponentiated as

$$\phi(t) = \phi(0)\,\mathrm{e}^{\hat{H}t}\Big|_{t=0}$$

which has formal analogies with the Heisenberg picture of quantum mechanics.

13.11 More on angular momentum and Poisson brackets

Let us take the ICT of a rotation of an angle θ about an axis $\vec{n}$. We can write

$$\delta\vec{\omega} = \mathrm{d}\theta \left[\vec{\omega}, \overrightarrow{M} \cdot \vec{n}\right]$$

where $\overrightarrow{M}$ is the angular momentum.
However, we know that under a rotation θ about $\vec{n}$, a generic vector $\vec{v}$ changes as

$$\delta\vec{v} = \mathrm{d}\theta\, \vec{n} \times \vec{v},$$

then taking the change in $\vec{v}$ as the ICT, we obtain

$$\left[\vec{v}, \overrightarrow{M} \cdot \vec{n}\right] = \vec{n} \times \vec{v},$$

writing it in components,

$$\begin{aligned} [v_i, M_j n_j] &= \epsilon_{ijk} n_j v_k \\ \text{i.e. } [v_i, M_j] &= \epsilon_{ijk} v_k. \end{aligned}$$

In particular, if we take $\vec{v} = \overrightarrow{M}$ we have the relations

$$\left[\overrightarrow{M}, \overrightarrow{M} \cdot \vec{n}\right] = \vec{n} \times \overrightarrow{M} \tag{13.45}$$

or in components

$$[M_i, M_j] = \epsilon_{ijk} M_k \tag{13.46}$$

Eq. (13.45) implies that

$$\left[M^2, \overrightarrow{M} \cdot \vec{n}\right] = 0$$

and eq. (13.46) is formally equivalent to the relation for the generators of the Lie algebra of SO(3), the group of rotations in 3 dimensions (recall that for the Lie algebra generators, $[\cdot, \cdot]$ is a commutator of matrices which represent the generators, while here $[\cdot, \cdot]$ is a Poisson bracket, but they share the same formal properties). More on this later.

Note that if M_x and M_y are constants of motion, Jacobi-Poisson theorem says that also M_z is a constant of motion, since

$$[M_x, M_y] = M_z.$$

In addition, if we take the generic vector $\vec{v}$ to be the momentum $\vec{p}$, then

$$\left[\vec{p}, \overrightarrow{M} \cdot \vec{n}\right] = \vec{n} \times \vec{p},$$

or in components

$$[p_i, M_j] = \epsilon_{ijk} p_k,$$

so if p_z is conserved, then

$$[p_z, M_x] = p_y$$
$$[p_z, M_y] = -p_x$$

and we see that also p_x and p_y are conserved.
So if we started from M_x, M_y, p_z as constants of motion, or IEM, then Jacobi-Poisson theorem says that also M_z, p_x, p_y are IEM, thus $\overrightarrow{M}$ and $\vec{p}$ are IEM.

However, note that had we started from p_x, p_y, M_z as IEM, we would have found no new IEM through Jacobi-Poisson theorem, so one must rely on the kindness of Nature, in what IEM are given to begin with!

Exercise 13.1. Check the statement above.

Also note that for two canonical momenta

$$[p_i, p_j] = 0.$$

But

$$[M_i, M_j] = \epsilon_{ijk} M_k,$$

so two components of angular momentum cannot be canonical variables. However, the total angular momentum

$$\overrightarrow{M} = \sum_n \vec{r}_n \times \vec{p}_n$$

and one of the components can be, since

$$\left[M^2, \overrightarrow{M} \cdot \vec{n}\right] = 0.$$

This also has formal analogies in quantum mechanics, where M_i and M_j cannot be quantized together (we say that they cannot be simultaneously eigenvalues), but M^2 and M_i can.

13.12 Kepler problem in the Hamiltonian formalism

We saw that in an ICT as a rotation of angle θ about an axis $\vec{n}$, angular momentum $\overrightarrow{M}$ is conserved, and we have the relation

$$[M_i, M_j] = \epsilon_{ijk} M_k.$$

We also saw previously that the generators of the Lie algebra of SO(3), the group of rotations in 3 dimensions, are obtained from the rotation of a vector

$$\delta\vec{v} = \delta\theta\vec{n} \times \vec{v}$$

and fulfill the relation

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

The generators have the matrix representation

$$(J_i)_{jk} = -\mathrm{i}\epsilon_{ijk}$$

$$J_1 = \mathrm{i}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \qquad J_2 = \mathrm{i}\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} \qquad J_3 = \mathrm{i}\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

We noted the formal analogy between the Poisson bracket

$$[M_i, M_j] = \epsilon_{ijk} M_k$$

on the angular momentum and the commutation of matrices in

$$[J_i, J_j] = \mathrm{i}\epsilon_{ijk} J_k.$$

Now, let us consider again Kepler problem in a central potential

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

The angular momentum $\overrightarrow{M}$ is conserved, but we know that also another quantity, the Laplace-Runge-Lenz (LRL) vector,

$$\overrightarrow{A} = \vec{p} \times \overrightarrow{M} - \frac{ma\vec{r}}{r}$$

is conserved.
Because $\overrightarrow{A}$ is a vector, we can take the relation

$$[v_i, M_j] = \epsilon_{ijk} v_k$$

and immediately establish that

$$[A_i, M_j] = \epsilon_{ijk} A_k.$$

With some more work, it can be shown that

$$[A_1, A_2] = -\left(p^2 - \frac{2ma}{r}\right) M_3.$$

Exercise 13.2. Check this relation.

Note that the energy E is

$$E = T + V = \frac{p^2}{2m} - \frac{a}{r}$$

so

$$[A_1, A_2] = -2mEM_3 \quad \text{where } E < 0$$

for a bound orbit.
We can rescale the LRL vector,

$$\overrightarrow{D} \equiv \frac{\vec{A}}{\sqrt{-2mE}},$$

thus we have

$$[D_1, D_2] = M_3.$$

Cyclically permuting the indices, we obtain

$$[D_i, D_j] = \epsilon_{ijk} M_k,$$

so the complete set of Poisson brackets for Kepler bound orbits is

$$[M_i, M_j] = \epsilon_{ijk} M_k \qquad [D_i, D_j] = \epsilon_{ijk} M_k \qquad [D_i, M_j] = \epsilon_{ijk} D_k.$$

The same relations are fulfilled by the generators of the Lie algebra of SO(4), the group of rotations in 4 dimensions:

$$[J_i, J_j] = \mathrm{i}\epsilon_{ijk} J_k \qquad [N_i, N_j] = \mathrm{i}\epsilon_{ijk} J_k \qquad [N_i, J_j] = \mathrm{i}\epsilon_{ijk} N_k.$$

To see it, let us add a zeroth row and column to J_i:

$$J_1 = \mathrm{i}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \qquad J_2 = \mathrm{i}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \qquad J_3 = \mathrm{i}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The generators N_i of rotations in the $x_0 - x_1$-plane are

$$N_1 = \mathrm{i}\begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad N_2 = \mathrm{i}\begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad N_3 = \mathrm{i}\begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

It is straightforward to check that they indeed satisfy the commutation relations above.
For unbound orbits, the energy $E > 0$ is positive and we rescale the LRL vector as

$$\overrightarrow{C} \equiv \frac{\overrightarrow{A}}{\sqrt{2mE}}$$

and the complete set of Poisson brackets is

$$[M_i, M_j] = \epsilon_{ijk} M_k \qquad [C_i, M_j] = \epsilon_{ijk} C_k \qquad [C_i, C_j] = -\epsilon_{ijk} M_k.$$

The same relations are fulfilled by the generators of the Lie algebra of the restricted Lorentz group. However, the analogy is purely formal: Kepler problem is non relativistic Newtonian mechanics, nothing to do with special relativity.

13.13 Poincare (Cartan) integral invariants

Euler-Lagrange equations are necessary and sufficient condition for the variation of the action

$$\delta S = \int_{t_i}^{t_f} \mathrm{d}t \, L\left(\{q_i\}, \{\dot{q}_i\}, t\right) = 0$$

of a holonomic system to vanish. We also now that in the $4n$-dimensional space spanned by q_i, p_i and their derivatives $\dot{q}_i, \dot{p}_i$, Hamilton equations stem from the variation of the action

$$\delta S = \int_{t_i}^{t_f} \mathrm{d}t \left(\sum_i p_i \dot{q}_i - H\left(\{p_i\}, \{q_i\}, t\right) \right) = 0.$$

In all those variational-calculus manipulations, we always require that the initial and final times are fixed. Let us relax that conditions, and consider the $(2n+1)$-dimensional space also called *extended phase space* with coordinates

$(\{q\},\{p\},t)$ and suppose that we can parametrize them through a parameter ζ:

$$q_k = q_k(\zeta) \qquad p_k = p_k(\zeta) \qquad t = t(\zeta).$$

Then also the initial and final times depend on ζ:

$$t_i = t_i(\zeta) \qquad t_f = t_f(\zeta).$$

It can be shown that the variation of the action

$$S = \int_{t_i(\zeta)}^{t_f(\zeta)} \mathrm{d}t\, L\left(\{q\},\{\dot{q}\},t\right)$$

equals

$$\delta S = \left| \sum_k p_k \delta q_k - H\delta t \right|_{t_i(\zeta)}^{t_f(\zeta)} + \int_{t_i(\zeta)}^{t_f(\zeta)} \mathrm{d}t \left(\sum_k p_k \delta q_k - H\delta t \right).$$

On the trajectories of the extended phase space where Hamilton equations are fulfilled, the integral term vanishes, and we have

$$\delta S = \left| \sum_k p_k \delta q_k - H\delta t \right|_{t_i(\zeta)}^{t_f(\zeta)}$$

(if, in particular t_i and t_f are fixed, then $\delta S = 0$).
Now, let us take an arbitrary closed curve C_i given by the equations

$$q_k = q_k^i(\zeta) \qquad p_k = p_k^i(\zeta) \qquad t^i = t^i(\zeta) \quad \text{with } 0 \le \zeta \le T$$

and

$$q_k^i(0) = q_k^i(T) \qquad p_k^i(0) = p_k^i(T) \qquad t^i(0) = t^i(T).$$

For every point on the contour C_i, we can draw a line using Hamilton's equations. Thus we obtain a cylinder of lines. Let us take another arbitrary closed curve C_f with

$$q_k^f(0) = q_k^f(T) \qquad p_k^f(0) = p_k^f(T) \qquad t^f(0) = t^f(T).$$

Next, we integrate the variation of the action

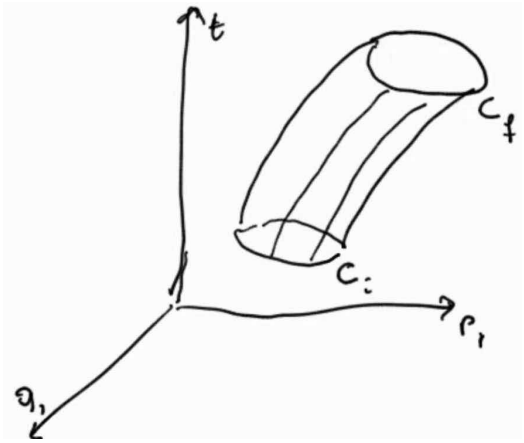


Figure 41: Cylinder of lines.

$$\delta S = \frac{\partial S}{\partial \zeta}\delta\zeta = \left|\sum_k p_k \delta q_k - H\delta t\right|_{t_i(\zeta)}^{t_f(\zeta)}$$

between 0 and T:

$$\begin{aligned} S(T) - S(0) &= \int_0^T \left|\sum_k p_k \delta q_k - H\delta t\right|_{t_i(\zeta)}^{t_f(\zeta)} \\ &= \int_0^T \left(\sum_k p_k^f \delta q_k^f - H_f \delta t_f\right) - \int_0^T \left(\sum_k p_k^i \delta q_k^i - H_i \delta t_i\right) \\ &= \oint_{C_f} \left(\sum_k p_k \delta q_k - H\delta t\right) - \oint_{C_i} \left(\sum_k p_k \delta q_k - H\delta t\right), \end{aligned}$$

but $S(T) = S(0)$

$$\Rightarrow \oint_{C_f} \left(\sum_k p_k \delta q_k - H\delta t\right) = \oint_{C_i} \left(\sum_k p_k \delta q_k - H\delta t\right).$$

Thus we have established that

$$I = \oint \left(\sum_k p_k \delta q_k - H\delta t\right)$$

is an integral invariant. It is *Poincare-Cartan integral invariant* (which in some, more geometric, formulations, is considered as the fundamental object of mechanics). It is possible to show that the canonical transformations can be obtained from the invariance of the Poincare-Cartan integral. Several theorems of flows in hydrodynamics also follow from Poincare-Cartan integral.

If, in particular, we consider contours C of simultaneous states, i.e. we take the hyperplane of constant time in the cylinder above, then $\delta t = 0$ and we obtain

$$I_1 = \oint \sum_k p_k \delta q_k$$

which is called *Poincare integral invariant.*

I_1 is invariant with respect to any Hamiltonian system. In fact, if for any system given by the equations

$$\dot{q}_k = Q_k\left(\{q\},\{p\},t\right) \qquad \dot{p}_k = P_k\left(\{q\},\{p\},t\right)$$

I_1 is invariant, it follows that the system is Hamiltonian. Indeed, using $\dfrac{\mathrm{d}}{\mathrm{d}t}\delta q = \delta\dot{q}$:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}I_1 = 0 &= \oint \sum_k \left(\dot{p}_k \delta q_k + p_k \frac{\mathrm{d}}{\mathrm{d}t}\delta q_k\right) \\ &= \oint \sum_k \left(\dot{p}_k \delta q_k + p_k \delta \dot{q}_k\right) \\ &= \oint \sum_k \left(\dot{p}_k \delta q_k - \dot{q}_n \delta p_k\right),\end{aligned}$$

thus

$$\oint \sum_k \left(P_k \delta q_k - Q_k \delta p_k\right) = 0.$$

The integrand may be the total differential of some function. If we take that function to be

$$-H\left(\{q\},\{p\},t\right),$$

then

$$\sum_k \left(p_k \delta q_k - Q_k \delta p_k\right) = -\delta H = -\sum_k \left(\frac{\partial H}{\partial q_k}\delta q_k + \frac{\partial H}{\partial p_k}\delta p_k\right).$$

It follows that

$$Q_k = \frac{\partial H}{\partial p_k} \qquad P_k = -\frac{\partial H}{\partial q_k}$$

which implies that the system is Hamiltonian.

Now, think of Stokes' theorem in 3 dimensions:

$$\oint_{\partial S} \vec{A} \cdot \mathrm{d}\vec{l} = \iint_S \left(\vec{\nabla} \times \vec{A}\right) \cdot \mathrm{d}\vec{S}$$

where ∂S is a contour C which is the boundary of a surface S. In particular, in 2 dimensions,

$$\oint_{\partial S} A_x \,\mathrm{d}x + A_y \,\mathrm{d}y = \iint_S \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right) \mathrm{d}x \,\mathrm{d}y$$

which is also called *Green's theorem.*

Stokes' and Green's theorems (as well as Gauss' theorem) are all particular cases of the generalized Stokes' theorem for *differential forms*, which states that

$$\int_{C=\partial S} \omega = \iint_S \mathrm{d}\omega$$

i.e. the integral of a *form* ω on a closed contour C equals the integral of the form $\mathrm{d}\omega$ over the surface S bounded by the contour C.
In a nutshell, consider the differential form in 2 dimensions:

$$\omega = \vec{A} \cdot \mathrm{d}\vec{l} = A_x \,\mathrm{d}x + A_y \,\mathrm{d}y,$$

then the derivative

$$\mathrm{d}\omega = \mathrm{d}\vec{A} \wedge \mathrm{d}\vec{l}$$

is defined through the wedge product $\wedge$, which is linear and antisymmetric, just like the cross product:

$$a \wedge b = -b \wedge a.$$

Then

$$\begin{aligned}
\mathrm{d}\vec{A} \wedge \mathrm{d}\vec{l} &= \mathrm{d}A_x \wedge \mathrm{d}x + \mathrm{d}A_y \wedge \mathrm{d}y \\
&= \frac{\partial A_x}{\partial x} \cancelto{0}{\mathrm{d}x \wedge \mathrm{d}x} + \frac{\partial A_x}{\partial y} \mathrm{d}y \wedge \mathrm{d}x + \frac{\partial A_y}{\partial x} \mathrm{d}x \wedge \mathrm{d}y + \cancelto{0}{\frac{\partial A_y}{\partial y} \mathrm{d}y \wedge \mathrm{d}y} \\
&= \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right) \mathrm{d}x \wedge \mathrm{d}y
\end{aligned}$$

so

$$\int \vec{A} \cdot \mathrm{d}\vec{l} = \iint \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \right) \mathrm{d}x \wedge \mathrm{d}y$$

where $\mathrm{d}x \wedge \mathrm{d}y$ is the (oriented) area element. For our purposes, we can just take it to be $\mathrm{d}x\,\mathrm{d}y$. Thus, we got Green's theorem.

Now, take the form

$$\omega = \vec{p} \cdot \mathrm{d}\vec{q} = p_1\,\mathrm{d}q_1 + \cdots + p_n\,\mathrm{d}q_n$$

in the $2n$-dimensional phase space, then

$$\mathrm{d}\omega = \mathrm{d}\vec{p} \wedge \mathrm{d}\vec{q} = \mathrm{d}p_1 \wedge \mathrm{d}q_1 + \cdots + \mathrm{d}p_n \wedge \mathrm{d}q_n.$$

Then Stokes' theorem implies

$$\int_{\partial S = C} \omega = \iint_S \mathrm{d}\omega$$
$$\Rightarrow I_1 = \oint_{\partial S = C} \sum_i p_i\,\mathrm{d}q_i = \iint_S \sum_i \mathrm{d}p_i \wedge \mathrm{d}q_i \equiv J_2$$

where $\mathrm{d}p_i \wedge \mathrm{d}q_i$ are area elements.

We have related Poincare integral invariant I_1 to a 2-dimensional integral invariant J_2. In fact, this structure can be iterated, i.e. there is an invariant integral

$$I_3 = \iiint \sum_{i,k} p_i \delta q_i \delta p_k \delta q_k$$

which can be related to a

$$J_4 = \iiiint \sum_{i,k} \delta p_i \delta q_i \delta p_k \delta q_k$$

and so on, until one gets to I_{2n-1} and

$$J_{2n} = \int \ldots \int \mathrm{d}q_1 \ldots \mathrm{d}q_n\ \mathrm{d}p_1 \ldots \mathrm{d}p_n$$

which is of particular interest, because it is the volume element of the phase space.
It is particularly simple to see that J_{2n} is an invariant. In fact, if

$$\mathrm{d}\vec{\omega} = \mathrm{d}q_1 \ldots \mathrm{d}q_n\ \mathrm{d}p_1 \ldots \mathrm{d}p_n$$

through a canonical transformation the volume becomes

$$\mathrm{d}\vec{\Omega} = \mathrm{d}Q_1 \dots \mathrm{d}Q_n \; \mathrm{d}P_1 \dots \mathrm{d}P_n$$

but $\mathrm{d}\vec{\omega}$ and $\mathrm{d}\vec{\Omega}$ are related by the Jacobian of $\mathbf{M}$:

$$\delta\vec{\Omega} = |\det \mathbf{M}| \, \mathrm{d}\vec{\omega} \quad \text{with } M_{ij} = \frac{\mathrm{d}\Omega_i}{\mathrm{d}\omega_j}$$

and we know that

$$\mathbf{M}^T \mathbf{J} \mathbf{M} = \mathbf{J} \Rightarrow (\det \mathbf{M})^2 = 1$$

so

$$|\det \mathbf{M}| = 1 \Rightarrow \mathrm{d}\vec{\Omega} = \mathrm{d}\vec{\omega}.$$

The fact that a volume element is invariant under a canonical transformation immediately implies *Liouville's theorem* in statistical mechanics, i.e. that the density of systems around some given system in phase space remains constant in time.
In fact, the number of systems in a closed neighbourhood of a given system does not change with time; the closed shape just moves about in phase space. Canonical variables at different times are related by a canonical transformation, but we know that the volume element J_{2n} is a canonical invariant, so the volume size cannot change with time. Therefore the density of systems is a constant of motion.

Example 13.4. See fig. (42).

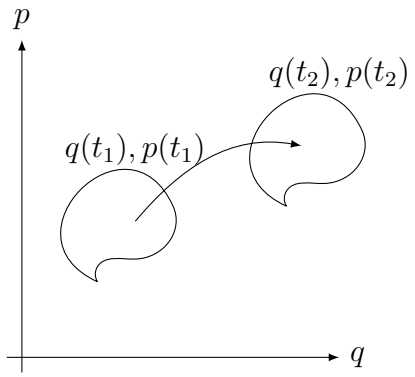


Figure 42: A closed shape moves with time in a 2-dimensional phase space.

14 Hamilton-Jacobi formalism

14.1 Hamilton-Jacobi equation

Canonical transformations help us to understand the dynamics of a mechanical system, i.e. to solve its equations of motion.
They can do that in 2 ways:

(a) By transforming the Hamiltonian $H(q_1, \ldots, q_n; p_1, \ldots, p_n; t)$ to a Hamiltonian $K(Q_1, \ldots, Q_n; P_1, \ldots, P_n; t)$ where all the coordinates Q_i are cyclic.
In the new variables, Hamilton's equations are

$$\dot{Q}_i = \frac{\partial K}{\partial P_i} \qquad \dot{P}_i = -\frac{\partial K}{\partial Q_i}.$$

Then, because $\dfrac{\partial K}{\partial Q_i} = 0$, $P_i = \text{const.} = c_i$, so the Hamiltonian becomes

$$K = K(c_1, \ldots, c_n; t).$$

Suppose that the Hamiltonian is conserved,

$$\frac{\mathrm{d}K}{\mathrm{d}t} = \frac{\partial K}{\partial t} = 0,$$

then

$$\dot{Q}_i = \frac{\partial K}{\partial c_i} = f_i(c_1, \ldots, c_n)$$

which can be trivially integrated to

$$Q_i = f_i(\{c\})\, t + \gamma_i$$

so the system is fully specified in terms of $2n$ constants of motion, c_i and γ_i.

(b) The second way is to transform to a set of $2n$ independent constants of motion (IEM)

$$f_k(\{q\}, \{p\}, t) = c_k \quad k = 1, \ldots, 2n$$

such that

$$q_i = q_i\left(c_1, \ldots, c_{2n}, t\right)$$
$$p_i = p_i\left(c_1, \ldots, c_{2n}, t\right)$$

so the system is completely integrable.
The $2n$ independent IEM might also be the initial conditions such that

$$q_i = q_i\left(\{q_0\}, \{p_0\}, t\right)$$
$$p_i = p_i\left(\{q_0\}, \{p_0\}, t\right).$$

Then one looks for a canonical transformation that maps the set of variables $(\{q\}, \{p\})$ to the set $(\{q_0\}, \{p_0\})$.

We shall follow the second avenue: Let us consider a canonical transformation generated by the function $F\left(\{q\}, \{P\}, t\right)$ (although the discussion is fully general, and can be done with other generating functions).
The transformation equations are:

$$p_i = \frac{\partial F}{\partial q_i} \qquad Q_i = \frac{\partial F}{\partial P_i} \qquad K = H + \frac{\partial F}{\partial t}.$$

Suppose that we have found a function F, such that the new Hamiltonian vanishes:

$$K = 0 = H + \frac{\partial F\left(\{q\}, \{P\}, t\right)}{\partial t}.$$

Then the new equations of motions can be trivially solved:

$$\dot{Q}_i = \frac{\partial K}{\partial P_i} = 0 \quad \Rightarrow \quad Q_i = \text{const.} = \beta_i$$
$$\dot{P}_i = -\frac{\partial K}{\partial Q_i} = 0 \quad \Rightarrow \quad P_i = \text{const.} = \alpha_i.$$

The equation for $K = 0$ is called *Hamilton-Jacobi-equation*:

$$H\left(\{q\}\left\{\frac{\partial F}{\partial q}\right\}, t\right) + \frac{\partial F}{\partial t} = 0.$$

It is a 1st-order partial differential equation in $(n+1)$ variables: $q_1, \ldots, q_n, t$. Its solution,

$$F \equiv S\left(q_1, \ldots, q_n; t; \alpha_1, \ldots, \alpha_{n+1}\right)$$

must be given in terms of $n+1$ independent IEM. S is called *Hamilton's principal function.*
Because S appears in the HJ-equation only through the derivatives of Q, if S is a solution also $S+\alpha$ is a solution. So S is determined up to an additive constant, which is irrelevant since the transformation equations are given through derivatives of S,then

$$S = S\left(q_1, \ldots, q_n; \alpha_1, \ldots, \alpha_n; t\right).$$

Because the generating function is $F\left(\{q\}, \{P\}; t\right)$ we may take the new momenta P_i to be the constants α_i

$$P_i = \alpha_i \quad \text{or any combination of them} \quad P_i = f_i\left(\alpha_1, \ldots, \alpha_n\right).$$

Then the transformation equations are

$$p_i = \frac{\partial S\left(\{q\}, \{\alpha\}, t\right)}{\partial q_i} \tag{14.1}$$

$$Q_i = \beta_i = \frac{\partial S\left(\{q\}, \{\alpha\}, t\right)}{\partial \alpha_i} \tag{14.2}$$

At $t = t_0$, eqs. (14.1) are n equations connecting α_i to $\{q_0\}$, $\{p_0\}$. The β_i from eqs. (14.2) can also be obtained from the initial conditions, by computing the r.h.s. at $t = t_0$, so one gets

$$q_i = q_i\left(\{\alpha\}, \{\beta\}, t\right)$$

and through eq. (14.1) also

$$p_i = p_i\left(\{\alpha\}, \{\beta\}, t\right),$$

which solves the system.
To understand the physical meaning of S, let us take the time derivative:

$$\begin{aligned} \frac{\mathrm{d}S}{\mathrm{d}t} &= \sum_i \frac{\partial S}{\partial q_i}\dot{q}_i + \frac{\partial S}{\partial t} \\ &= \sum_i p_i \dot{q}_i - H = L. \end{aligned}$$

So, up to a constant, S is the action:

$$S = \int L \, \mathrm{d}t + \text{const.}.$$

Finally, if the Hamiltonian is conserved,

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \frac{\partial H}{\partial t} = 0,$$

then the HJ-equation is

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

and one can separate the variable t from the others,

$$S\left(\{q\}, \{\alpha\}, t\right) = W\left(\{q\}, \{\alpha\}\right) - \alpha_n t$$

where W is called *Hamilton's characteristic function.*
Substituting it into HJ-equation, we get the reduced HJ-equation

$$H\left(\{q\}, \left\{\frac{\partial W}{\partial q}\right\}\right) = \alpha_n.$$

For conservative systems, $\alpha = E$ the energy.
The physical meaning of W is understood again by taking the time derivative:

$$\frac{\mathrm{d}W}{\mathrm{d}t} = \sum_i \frac{\partial W}{\partial q_i}\dot{q}_i = \sum_i p_i \dot{q}_i.$$

Integrating gives us

$$W = \sum_i \int \mathrm{d}t\, p_i \dot{q}_i = \sum_i \int \mathrm{d}q_i p_i.$$

Note that also W generates a canonical transformation, different from the one generated by S. The transformation is

$$p_i = \frac{\partial W}{\partial q_i} \qquad Q_i = \frac{\partial W}{\partial P_i} = \frac{\partial W}{\partial \alpha_i} \qquad K = H = \alpha_n.$$

Thus W generates a canonical transformation where all coordinates are cyclic. Hamilton's equations are

$$\begin{aligned}
\dot{P}_i = -\frac{\partial K}{\partial Q_i} = 0 \quad &\Rightarrow \quad P_i = \text{const.} = \alpha_i \\
\dot{Q}_i = \frac{\partial K}{\partial \alpha_i} = 0 \quad i = 1, \ldots, n-1 \quad &\Rightarrow \quad Q_i = \text{const.} = \beta_i \\
\dot{Q}_n = \frac{\partial K}{\partial \alpha_n} = 1 \quad &\Rightarrow \quad Q_n = t + \beta_n.
\end{aligned}$$

Note that $W(\{q\},\{\alpha\})$ depends on n constants, but one of them is additive since in the reduced HJ-equation

$$H = H\left(\{q\},\left\{\frac{\partial W}{\partial q}\right\}\right) = \alpha_n.$$

W appears only through derivatives.
So W depends on $n-1$ independent constants, say $\alpha_1,\ldots,\alpha_{n-1}$ which together with α, form a set of n constants, the new momenta, P_i.
To see the difference between the canonical transformations generated by S and W, draw the prospect:

The Hamiltonian is

$H = (\{q\},\{p\},t)$	an IEM: $H = H(\{q\},\{p\})$

Look for canonical transformations to $(\{Q\},\{P\})$ such that

$\{Q\},\{P\}$ are constants	$\{P\}$ are constants

Require that the new Hamiltonian

vanishes: $K = 0$	has $K = H = \alpha_n$, so Q_i are cyclic

Hamilton's equations in the new variables are

$\dot{Q}_i = \dfrac{\partial K}{\partial P_i} = 0$	$\dot{Q}_i = \dfrac{\partial K}{\partial P_i} = v_i(\alpha_i)$
$\dot{P}_i = -\dfrac{\partial K}{\partial Q_i} = 0$	$\dot{P}_i = -\dfrac{\partial K}{\partial Q_i} = 0$

with solutions

$Q_i = \beta_i$	$Q_i = \beta_i + v_i(\alpha_i)\,t$
$P_i = \alpha_i$	$P_i = \alpha_i$

The generating function is Hamilton's

principal function $S(\{q\},\{P\},t)$	characteristic func. $W(\{q\},\{P\})$

HJ-equation is

$H\left(\{q\},\left\{\dfrac{\partial S}{\partial q}\right\},t\right)$	$H\left(\{q\},\left\{\dfrac{\partial W}{\partial q}\right\}\right) = \alpha_n$

A complete solution has

n non additive constants $\alpha_1, \ldots, \alpha_n$	$n - 1$ non-additive constants $\alpha_1, \ldots, \alpha_{n-1}$ which together with α_n make n independent constants

14.2 Simple harmonic oscillator

The Hamiltonian for the simple harmonic oscillator is

$$H = \frac{1}{2m}\left(p^2 + m^2\omega^2 q^2\right) = E \qquad \text{with } \omega = \sqrt{\frac{k}{m}}.$$

The Hamilton-Jacobi equation is

$$\frac{1}{2m}\left[\left(\frac{\partial S}{\partial q}\right)^2 + m^2\omega^2 q^2\right] + \frac{\partial S}{\partial t} = 0.$$

Since $\dfrac{\mathrm{d}H}{\mathrm{d}t} = 0$, we have the r.h.s. case of the table above, so

$$S = W - \alpha t \qquad \text{with } \alpha = E.$$

The reduced HJ-equation is

$$\frac{1}{2m}\left[\left(\frac{\partial W}{\partial q}\right)^2 + m^2\omega^2 q^2\right] = \alpha$$

which can be integrated easily to

$$W = \sqrt{2m\alpha}\int \mathrm{d}q\,\sqrt{1 - \frac{m\omega^2 q^2}{2\alpha}} \quad \Rightarrow \quad S = \sqrt{2m\alpha}\int \mathrm{d}q\,\sqrt{1 - \frac{m\omega^2 q^2}{2\alpha}} - \alpha t.$$

Also the transformation equation

$$\beta = \frac{\partial S}{\partial \alpha} = \sqrt{\frac{m}{2\alpha}}\int \frac{\mathrm{d}q}{\sqrt{1 - \frac{m\omega^2 q^2}{2\alpha}}} - t$$

can be integrated easily to

$$\beta + t = \frac{1}{\omega}\arcsin q\sqrt{\frac{m\omega^2}{2\alpha}}$$

that is

$$q = \sqrt{\frac{2\alpha}{m\omega^2}} \sin(\omega t + \gamma) \quad \text{with } \gamma = \beta\omega$$

which is the usual solution for the harmonic oscillator.
The other transformation equation yields

$$p = \frac{\partial S}{\partial q} = \sqrt{2m\alpha} \cos(\omega t + \gamma).$$

α and β are related to the initial conditions at $t = t_0$:

$$\alpha = \frac{1}{2m}\left(p_0^2 + m^2\omega^2 q_0^2\right) \qquad \frac{q_0}{p_0} = \frac{1}{m\omega}\tan(\beta\omega)$$

so Hamilton's principal function generates a canonical transformation to a new momentum, i.e. the energy, and a new coordinate, i.e. the phase constant of the oscillation.

14.3 2-dimensional harmonic oscillator

In two dimensions, the Hamiltonian for the harmonic oscillator is

$$H = \frac{1}{2m}\left(p_x^2 + p_y^2 + m^2\omega_x^2 x^2 + m^2\omega_y^2 y^2\right) \quad \text{with } \omega_x = \sqrt{\frac{k_x}{m}} \quad \omega_y = \sqrt{\frac{k_y}{m}}$$

and we have again

$$\frac{\mathrm{d}H}{\mathrm{d}t} = 0 \quad \Rightarrow \quad S\left(x, y, \alpha_x, \alpha_y; t\right) = W\left(x, y, \alpha_x, \alpha_y\right) - \alpha t \quad \text{with } \alpha = E.$$

The reduced HJ-equation is

$$\frac{1}{2m}\left[\left(\frac{\partial W}{\partial x}\right)^2 + \left(\frac{\partial W}{\partial y}\right)^2 + m^2\omega_x^2 x^2 + m^2\omega_y^2 y^2\right] = \alpha$$

which can be separated into

$$\frac{1}{2m}\left[\left(\frac{\partial W}{\partial x}\right)^2 + m^2\omega_x^2 x^2\right] = \alpha_x$$

$$\frac{1}{2m}\left[\left(\frac{\partial W}{\partial y}\right)^2 + m^2\omega_y^2 y^2\right] = \alpha_y \quad \text{with } \alpha = \alpha_x + \alpha_y$$

that can be easily solved.

14.4 Separation of variables

In sect. 14.1, we assumed (very wishfully!) that we know Hamilton's principal function S. But in general it may be very difficult to solve HJ-equation: partial differential equations of many variables are hard to deal with.
But with the harmonic oscillator, we saw that a solution can be found if we can separate the variables. In fact, that is the case when HJ-equation becomes a useful tool.
Suppose that in the HJ-equation, the coordinate q_1 enters only through the function $H_1\left(q_1, \dfrac{\partial S}{\partial q_1}\right)$, such that

$$H\left(H_1\left(q_1, \frac{\partial S}{\partial q_1}\right); q_2, \ldots, q_n; \frac{\partial S}{\partial q_2}, \ldots, \frac{\partial S}{\partial q_n}; t\right) + \frac{\partial S}{\partial t} = 0,$$

then q_1 is called *separable*, and we look for a solution of the type

$$S = S_1\left(q_1; \alpha_1, \ldots, \alpha_n\right) + \overline{S}\left(q_2, \ldots, q_n; \alpha_1, \ldots, \alpha_n; t\right).$$

If we replace it into the HJ-equation, we get the identity

$$H\left(H_1\left(q_1, \frac{\partial S_1}{\partial q_1}\right); q_2, \ldots, q_n; \frac{\partial \overline{S}}{\partial q_2}, \ldots, \frac{\partial \overline{S}}{\partial q_n}; t\right) + \frac{\partial \overline{S}}{\partial t} = 0.$$

For this to be true for any q_1, it's necessary and enough to take

$$H_1 = \text{const.} = \alpha_1.$$

So we get 2 equations:

$$H_1\left(q_1, \frac{\partial S_1}{\partial q_1}\right) = \alpha_1$$

$$\overline{H}\left(q_2, \ldots, q_n; \frac{\partial \overline{S}}{\partial q_2}, \ldots, \frac{\partial \overline{S}}{\partial q_n}; \alpha_1; t\right) + \frac{\partial \overline{S}}{\partial t} = 0.$$

The first equation is a first-order ordinary differential equation, which is easy to solve. The second equation is still a partial differential equation, but with one less variable. This procedure can be iterated if we have k separable coordinates $q_1, \ldots, q_n$.
Then the principal function can be written as

$$S\left(q_1, \ldots, q_n; \alpha_1, \ldots, \alpha_n; t\right) = \sum_{i=1}^{k} S_i\left(q_i; \alpha_1, \ldots, \alpha_n\right) + \overline{S}\left(q_{k+1}, \ldots, q_n; \alpha_1, \ldots, \alpha_n; t\right)$$

and HJ-equation splits into k first order differential equations

$$H_i\left(q_i, \frac{\partial S_i}{\partial q_i}\right) = \alpha_i \quad i = 1, \ldots, k$$

and the partial differential equation

$$\overline{H}\left(q_{k+1}, \ldots, q_n; \frac{\partial \overline{S}}{\partial q_{k+1}}, \ldots \frac{\partial \overline{S}}{\partial q_n}; \alpha_1, \ldots, \alpha_k; t\right) + \frac{\partial \overline{S}}{\partial t} = 0.$$

If all the variables q_i and the time t are separable, the principal function can be written as

$$S\left(q_1, \ldots, q_n; \alpha_1, \ldots, \alpha_n; t\right) = \sum_{i=1}^{n} W_i\left(q_i; \alpha_1, \ldots, \alpha_n\right) - \alpha t.$$

The reduced HJ-equation is *completely separable* and can be written as

$$H\left(H_1\left(q_1, \frac{\partial W_1}{\partial q_1}\right); \ldots; H_n\left(q_n, \frac{\partial W_n}{\partial q_n}\right)\right) = \alpha.$$

We get n first-order differential equations

$$H_i\left(q_i, \frac{\partial W_i}{\partial q_i}\right) = \alpha_i \quad i = 1, \ldots, n$$

and the condition

$$H\left(\alpha_1, \ldots, \alpha_n\right) = \alpha$$

where α_i are called *separation constants.*

Note that if q_1 is separable and cyclic: $\dfrac{\partial H}{\partial q_1} = 0$, then

$$H_1\left(q_1, \frac{\partial S_1}{\partial q_1}\right) = \frac{\partial S_1}{\partial q_1} = \alpha_1 \quad \Rightarrow \quad S_1 = \alpha_1 q_1.$$

So if k variables are cyclic, the principal function becomes

$$S = \sum_{i=1}^{k} \alpha_i q_i + \overline{S}\left(q_{k+1}, \ldots, q_n; \alpha_1, \ldots, \alpha_n; t\right).$$

14.5 Action-angle variables

Consider a conservative system with a periodic motion. For simplicity, let us take it as a function of 2 variables (q, p). The Hamiltonian is

$$H(q, p) = \text{const.} = \alpha.$$

As new (constant) momentum, let us take Poincare invariant integral

$$J = \oint p \, \mathrm{d}q.$$

J is called *action variable*.
Hamilton's characteristic function is $W = W(q, J)$. The coordinate conjugate to J is the *angle variable* w.
The transformation equation is

$$w = \frac{\partial W}{\partial J} \qquad \text{and } w \text{ is cyclic.}$$

Hamilton's equation of motion is

$$\dot{w} = \frac{\partial H(J)}{\partial J} = \nu(J)$$

with ν a constant, and a function of J.
As expected, w integrates to

$$w = \nu t + \beta.$$

Consider the change in w through a period

$$\Delta w = \oint \frac{\partial W}{\partial q} \, \mathrm{d}q = \oint \frac{\partial^2 W}{\partial q \partial J} \, \mathrm{d}q.$$

Because J is a constant, we can take it out of the integral:

$$\Delta w = \frac{\mathrm{d}}{\mathrm{d}J} \oint \frac{\partial W}{\partial q} \, \mathrm{d}q = \frac{\mathrm{d}}{\mathrm{d}J} \oint \underbrace{p \, \mathrm{d}q}_{=J} = 1.$$

So, as q goes through a cycle, w changes by 1. Then

$$\Delta w = \nu \tau = 1$$

where τ is the period. So

$$\nu = \frac{1}{\tau}$$

is the frequency of the periodic motion which we have found *without solving the equations of motion of the system.*
The discussion can be directly extended to a periodic motion in a $2n$-dimensional phase space, in terms of n action variables

$$J_i = \oint p_i \, \mathrm{d}q_i.$$

15 Special Relativity

Special relativity is based on the assumption that the laws of nature are the same for inertial observers where their co-ordinates are related via Lorentz transformations:

$$x^\mu \to x^{\mu\prime} = \Lambda^\mu_{\ \nu} x^\nu + \rho^\mu. \tag{15.1}$$

where

$$x^\mu = \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} \tag{15.2}$$

a "four-vector" comprising space-time coordinates with $x^0 = ct$, ρ^μ is a constant four-vector and $\Lambda^\mu_{\ \nu}$ satisfies:

$$\Lambda^\mu_{\ \rho} \Lambda^\nu_{\ \sigma} g_{\mu\nu} = g_{\rho\sigma}. \tag{15.3}$$

The 4×4 matrix $g_{\mu\nu}$ is the so-called *metric*, defined as:

$$g_{\mu\nu} = \begin{cases} +1, & \mu = \nu = 0, \\ -1, & \mu = \nu = 1, 2, 3 \\ 0, & \mu \neq \nu \end{cases} \tag{15.4}$$

In the above we have used Einstein's summation convention. For example, one would write explicitly

$$\Lambda^\mu_{\ \nu} x^\nu = \Lambda^\mu_{\ 0} x^0 + \Lambda^\mu_{\ 1} x^1 + \Lambda^\mu_{\ 2} x^2 + \Lambda^\mu_{\ 3} x^3. \tag{15.5}$$

This is a convention that we will use extensively from now on.

15.1 Proper time

Lorentz transformations leave invariant "proper-time" intervals. These are defined as:

$$\mathrm{d}\tau^2 \equiv c^2\,\mathrm{d}t^2 - \mathrm{d}\vec{x}^2 = g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu. \tag{15.6}$$

Indeed, in a different reference frame we have from Eq. 15.1:

$$\mathrm{d}x^{\mu\prime} = \Lambda^\mu_{\ \nu}\,\mathrm{d}x^\nu. \tag{15.7}$$

A proper-time interval in the new frame is

$$\begin{aligned} \mathrm{d}\tau'^2 &= g_{\mu\nu}\,\mathrm{d}x'^{\mu}\,\mathrm{d}x'^{\nu} \\ &= g_{\mu\nu}\left(\Lambda^{\mu}_{\ \rho}\,\mathrm{d}x^{\rho}\right)\left(\Lambda^{\nu}_{\ \sigma}\,\mathrm{d}x^{\sigma}\right) \\ &= \left(g_{\mu\nu}\Lambda^{\mu}_{\ \rho}\Lambda^{\nu}_{\ \sigma}\right)\,\mathrm{d}x^{\rho}\,\mathrm{d}x^{\sigma} \\ &= g_{\rho\sigma}\,\mathrm{d}x^{\rho}\,\mathrm{d}x^{\sigma} = \mathrm{d}\tau^2. \end{aligned} \tag{15.8}$$

As a consequence of the invariance of proper-time intervals, *the speed of light is the same in all inertial frames.* Indeed, for light we have:

$$\left|\frac{\mathrm{d}\vec{x}}{\mathrm{d}t}\right| = c \rightsquigarrow \mathrm{d}\tau^2 = c^2\,\mathrm{d}t^2 - \mathrm{d}\vec{x}^2 = 0 \tag{15.9}$$

In a new frame,

$$\mathrm{d}\tau'^2 = \mathrm{d}\tau^2 = 0 \rightsquigarrow \left|\frac{\mathrm{d}\vec{x}'}{\mathrm{d}t'}\right| = c \tag{15.10}$$

Lorentz transformations are the only non-singular transformations which preserve proper-time intervals:

$$\begin{aligned} \mathrm{d}\tau^2 &= \mathrm{d}\tau'^2 \\ \rightsquigarrow g_{\rho\sigma}\,\mathrm{d}x^{\rho}\,\mathrm{d}x^{\sigma} &= g_{\mu\nu}\,\mathrm{d}x'^{\mu}\,\mathrm{d}x'^{\nu} \\ \rightsquigarrow g_{\rho\sigma}\,\mathrm{d}x^{\rho}\,\mathrm{d}x^{\sigma} &= g_{\mu\nu}\frac{\partial x^{\mu}}{\partial x_{\rho}}\frac{\partial x^{\nu}}{\partial x_{\sigma}}\,\mathrm{d}x^{\rho}\,\mathrm{d}x^{\sigma}, \end{aligned} \tag{15.11}$$

concluding that:

$$g_{\rho\sigma} = g_{\mu\nu}\frac{\partial x^{\mu}}{\partial x_{\rho}}\frac{\partial x^{\nu}}{\partial x_{\sigma}}. \tag{15.12}$$

Differentiating with $\mathrm{d}x^{\epsilon}$, we obtain:

$$0 = g_{\mu\nu}\left[\frac{\partial^2 x'^{\mu}}{\partial x^{\epsilon}\partial x^{\rho}}\frac{\partial x'^{\nu}}{\partial x^{\sigma}} + \frac{\partial^2 x'^{\mu}}{\partial x^{\epsilon}\partial x^{\sigma}}\frac{\partial x'^{\nu}}{\partial x^{\rho}}\right]. \tag{15.13}$$

To this, we add the same equation with $\epsilon \leftrightarrow \rho$ and subtract the same equation with $\epsilon \leftrightarrow \sigma$. We obtain:

$$0 = 2g_{\mu\nu}\frac{\partial^2 x'^{\mu}}{\partial x^{\epsilon}\partial x^{\rho}}\frac{\partial x'^{\nu}}{\partial x^{\sigma}} \tag{15.14}$$

Assuming that the transformation $x^{\mu} \to x'^{\mu}$ is a well behaved differentiable function and that the inverse of the transformation also exists,

$$\frac{\partial x'^{\mu}}{\partial x^{\sigma}}\frac{\partial x^{\sigma}}{\partial x'^{\nu}} = \delta_{\mu\nu}, \tag{15.15}$$

we obtain that

$$\frac{\partial^2 x'^\mu}{\partial x^\epsilon \partial x^\rho} = 0. \tag{15.16}$$

Therefore, the transformation $x^\mu \to x'^\mu$ ought to be linear:

$$x'^\mu = \Lambda^\mu\ \nu + \rho^\mu. \tag{15.17}$$

15.2 Subgroups of Lorentz transformations

The set of all Lorentz transformations

$$x'^\mu = \Lambda^\mu_{\ \nu} x^\nu + \rho^\mu. \quad g_{\mu\nu}\Lambda^\mu_{\ \rho}\Lambda^\nu_{\ \sigma} = g_{\rho\sigma} \tag{15.18}$$

form a group (**exercise:** prove it), which is known as the group of *inhomogeneous Lorentz group* or the Poincare' group. The subset of transformations with $\rho^\mu = 0$ is known as the *homogeneous Lorentz group.*

From

$$g_{\mu\nu}\Lambda^\mu_{\ \rho}\Lambda^\nu_{\ \sigma} = g_{\rho\sigma}$$

and for $\rho = \sigma = 0$, we have:

$$(\Lambda^0_{\ 0})^2 - \sum_{i=1}^{3}(\Lambda^i_{\ 0})^2 = 1 \rightsquigarrow (\Lambda^0_{\ 0})^2 \geq 1. \tag{15.19}$$

Also, in matrix form the definition of the Lorentz transformation becomes:

$$g = \Lambda^T g \Lambda \rightsquigarrow \det g = \det(\Lambda^T g \Lambda) \rightsquigarrow (\det \Lambda)^2 = 1. \tag{15.20}$$

The subgroup of transformations with

$$\det \Lambda = 1,\ \Lambda^0_{\ 0} \geq 1,$$

which contains the unity $\mathbf{1} = \delta^\mu_\nu$, is known as the proper group of Lorentz transformations. All other transformations are known as *improper Lorentz transformations.* It is impossible with a continuous change of parameters to change

$$\det \Lambda = 1 \to \det \Lambda = -1 \text{ or } \Lambda^0_{\ 0} \geq 1 \to \Lambda^0_{\ 0} \leq -1.$$

Improper Lorentz transformations involve either space-reflection ($\det \Lambda = -1, \Lambda^0_{\ 0} \geq 1$) or time-inversion ($\det \Lambda = 1, \Lambda^0_{\ 0} \leq -1$) or both ($\det \Lambda = -1, \Lambda^0_{\ 0} \leq -1$).

Proper homogeneous or inhomogeneous Lorentz transformations have a further subgroup: the group of rotations,

$$\Lambda^0_{\ 0} = 1,\ \Lambda^i_{\ 0} = \Lambda^0_{\ i} = 0,\ \Lambda^i_{\ j} = R_{ij}, \tag{15.21}$$

with

$$\det R = 1,\ R^T R = 1. \tag{15.22}$$

Thus, for rotations and translations ($x^\mu \to x'^\mu = x^\mu + \rho^\mu$) Lorentz transformations are no different than Galilei transformations.

A difference with Galilei transformations arises in *boosts*. Assume a reference frame O in which a certain particle appears at rest, and O' a reference frame where the particle appears to move with a velocity $\vec{v}$. Space-time intervals in the two frames are related via

$$\mathrm{d}x'^\mu = \Lambda^\mu_{\ \nu}\,\mathrm{d}x^\nu = \Lambda^\mu_{\ 0} c\,\mathrm{d}t, \tag{15.23}$$

given that $\mathrm{d}\vec{x} = 0$ in the frame O. For $\mu = 0$, this equation gives

$$\mathrm{d}t' = \Lambda^0_{\ 0}\,\mathrm{d}t. \tag{15.24}$$

For $\mu = i = 1, 2, 3$ we have:

$$\mathrm{d}x'^i = \Lambda^i_{\ 0} c\,\mathrm{d}t \tag{15.25}$$

Dividing the two, we have

$$v^i \equiv \frac{\mathrm{d}x'^i}{\mathrm{d}t'} = c\frac{\Lambda^i_{\ 0}}{\Lambda^0_{\ 0}} \rightsquigarrow \Lambda^i_{\ 0} = \frac{v^i}{c}\Lambda^0_{\ 0}. \tag{15.26}$$

From

$$g_{\mu\nu}\Lambda^\mu_{\ \rho}\Lambda^\nu_{\ \sigma} = g_{\rho\sigma}$$

and for $\rho = \sigma = 0$, we have:

$$\begin{aligned} & (\Lambda^0_{\ 0})^2 - (\Lambda^i_{\ 0})^2 = 1 \\ \rightsquigarrow \quad & \Lambda^0_{\ 0} = \gamma = \left(1 - \frac{\vec{v}^2}{c^2}\right)^{-\frac{1}{2}}. \end{aligned} \tag{15.27}$$

and thus

$$\Lambda^i_{\ 0} = \gamma\frac{v^i}{c}. \tag{15.28}$$

The remaining components are not determined uniquely by knowing the velocity $\vec{v}$ of the particle. Indeed, two Lorentz transformations

$$\Lambda^{\mu}_{\ \nu} \quad \text{and} \quad \Lambda^{\mu}_{\ \rho} R^{\rho}_{\ \nu}$$

where R is a rotation, boost a particle to the same velocity. For coordinate systems O and O' with parallel axes we find that (**exercise**)

$$\Lambda^{i}_{\ j} = \delta^{i}_{j} + \frac{v^i v^j}{v^2}(\gamma - 1) \tag{15.29}$$

and

$$\Lambda^{0}_{\ j} = \gamma \frac{v^j}{c}. \tag{15.30}$$

15.3 Time dilation

Consider an inertial observer O which looks as a clock at rest. Two ticks of the clock correspond to a space-time interval

$$\mathrm{d}\vec{x} = 0, \quad \mathrm{d}t = \Delta t. \tag{15.31}$$

The proper time interval is

$$\mathrm{d}\tau = (c^2\,\mathrm{d}t^2 - \mathrm{d}\vec{x}^2)^{\frac{1}{2}} = c\Delta t. \tag{15.32}$$

A second observer sees the clock with velocity $\vec{v}$. Two ticks of the clock define a space-time interval

$$\mathrm{d}t' = \Delta t', \quad \mathrm{d}\vec{x}' = \vec{v}\,\mathrm{d}t'. \tag{15.33}$$

The proper-time interval in the new frame is:

$$\mathrm{d}\tau' = (c^2\,\mathrm{d}t'^2 - \mathrm{d}\vec{x}'^2)^{\frac{1}{2}} = c\Delta t' \sqrt{1 - \left|\frac{\mathrm{d}\vec{x}'}{c\mathrm{d}t'}\right|^2} = c\Delta t' \sqrt{1 - \frac{\vec{v}^2}{c^2}}. \tag{15.34}$$

The proper-time is invariant under the change of inertial reference frames. Thus we conclude that

$$\Delta t' = \frac{\Delta t}{\sqrt{1 - \frac{\vec{v}^2}{c^2}}} = \gamma \Delta t \tag{15.35}$$

15.4 Doppler effect

Take our clock to be a source of light with a frequency

$$\omega = \frac{2\pi}{\Delta t}.$$

For an observer where the light-source is moving with velocity $\vec{v}$ this time interval is measured to be

$$\mathrm{d}t' = \gamma \Delta t.$$

In the same period, the distance of the observer from the light source increases by

$$v_r \, \mathrm{d}t'$$

where v_r is the component of the velocity of the light-source along the direction of sight of the observer. The time elapsing between the reception of two successive light wave-fronts from the observer is

$$c \, \mathrm{d}t_0 = c \, \mathrm{d}t' + v_r \, \mathrm{d}t'. \tag{15.36}$$

The frequency measured by the observer is

$$\omega' = \frac{2\pi}{dt_0} = \frac{\sqrt{1 - \frac{v^2}{c^2}}}{1 + \frac{v_r}{c}} \omega. \tag{15.37}$$

If the light-source is moving along the line of sight, $v_r = v$, we have

$$\omega' = \sqrt{\frac{1 - \frac{v}{c}}{1 + \frac{v}{c}}} \omega. \tag{15.38}$$

If the light-source moves away from the observer, $v_r > 0$, the frequency decreases and the light appears to be more red (red shift). If the source moves towards the observer, the frequency increases (violet shift).

Exercise 15.1. Calculate the angle of the direction of motion of the light-source with respect to the line of sight of the observer for which there is no shift in the frequency.

For an application of the Doppler effect in cosmology, read about Hubble's law.

15.5 Particle dynamics

How can we compute the force of a particle which moves with a relativistic velocity $\vec{v}$? We should expect that our classical formulae from Newtonian mechanics need to be modified. On the other hand, Newtonian expressions for the force should be valid if a particle is at rest. We can always change reference frame with Lorentz transformations to bring a particle at rest and calculate the change in its velocity for a small time interval using Newtonian mechanics. However, we will need to perform these changes of reference frame at every small increase of the velocity of the particle during its acceleration due to the force.

In a more elegant solution to the problem, we define a relativistic force acting on a particle as

$$f^\mu = mc^2 \frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tau^2}, \tag{15.39}$$

where m is the mass of a particle [3]. If the particle is at rest, the proper-time interval $d\tau$ coincides with the common time-interval dt

$$\mathrm{d}\tau = c\,\mathrm{d}t.$$

Therefore, in the rest frame of the particle, the"space"-components of the force four-vector become

$$f^i_{\mathrm{rest}} = m\frac{\mathrm{d}^2 x^i}{\mathrm{d}t^2} = F^i_{\mathrm{Newton}}, \quad \text{for } i = 1, 23, \tag{15.40}$$

where $\overrightarrow{F}_{\mathrm{Newton}}$ is the force-vector as we know it from Newtonian mechanics. The "time" component of the force four-vector vanishes:

$$f^0_{rest} = mc\frac{\mathrm{d}^2 t}{\mathrm{d}t^2} = 0. \tag{15.41}$$

Under a Lorentz transformation, f^μ transforms as

$$f^\mu = mc^2\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tau^2} \to mc^2\frac{\mathrm{d}^2 x'^\mu}{\mathrm{d}\tau^2} = mc^2\frac{\mathrm{d}^2\left(\Lambda^\mu_\nu x^\nu + \rho^\mu\right)}{\mathrm{d}\tau^2} = \Lambda^\mu_{\ \nu}\, mc^2\frac{\mathrm{d}^2 x^\nu}{\mathrm{d}\tau^2} \tag{15.42}$$

Therefore,

$$f'^\mu = \Lambda^\mu_\nu f^\nu. \tag{15.43}$$

[3] With mass, we mean the mass of a particle as it is measured in its rest-frame. We will refrain from using the "relativistic", velocity dependent, mass.

The components of f^μ transform under Lorentz transformations in exactly the same way as the components of space-time coordinates. It is therefore a four-vector as well.

For a specific transformation from the rest frame of a particle to a frame where the particle moves with a velocity $\vec{v}$, we have

$$f^\mu = \Lambda^\mu_{\ \nu}(\vec{v}) f^\nu_{\text{rest}}. \tag{15.44}$$

where, we have found that,

$$\begin{aligned} \Lambda^0_{\ 0}(\vec{v}) = \gamma = \left(1 - \frac{\vec{v}^2}{c^2}\right)^{-\frac{1}{2}}, \quad \Lambda^i_{\ 0}(\vec{v}) = \Lambda^0_{\ i}(\vec{v}) = \gamma \frac{v^i}{c}. \\ \Lambda^i_{\ j}(\vec{v}) = \delta^i_j + \frac{v^i v^j}{v^2}(\gamma - 1) \end{aligned} \tag{15.45}$$

Therefore, the force on a moving particle is:

$$\vec{f} = \vec{F}_{\text{Newton}} + (\gamma - 1)\frac{\vec{v}\left(\vec{F}_{\text{Newton}} \cdot \vec{v}\right)}{v^2}, \tag{15.46}$$

and

$$f^0 = \gamma \frac{\vec{v} \cdot \vec{F}_{\text{Newton}}}{c} = \frac{\vec{v}}{c} \cdot \vec{f}. \tag{15.47}$$

In Newtonian mechanics, if the force $\vec{F}$ is given, we can compute the trajectory $\vec{x}(t)$ by solving the second order differential equation:

$$\frac{\mathrm{d}^2\vec{x}}{\mathrm{d}t^2} = \frac{\vec{F}(\vec{x}, t)}{m}. \tag{15.48}$$

Similarly, in special relativity, when the relativistic force f^μ is known, the differential equation 15.39 can, in principle, be solved to give the space-time coordinates as a function of the proper time τ:

$$x^\mu = x^\mu(\tau). \tag{15.49}$$

To calculate the trajectory, we then need to calculate the proper-time in terms of the time coordinate by inverting

$$x^0 = x^0(\tau) \rightsquigarrow \tau = \tau(x^0), \tag{15.50}$$

which we can use to cast the space components directly as functions of the time-coordinate.

We should not forget a second constrain that must be satisfied for our solutions $x^\mu(\tau)$, namely

$$\Omega \equiv g_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau} = 1. \tag{15.51}$$

We have for the derivative of Ω with respect to proper-time:

$$\frac{\mathrm{d}\Omega}{\mathrm{d}\tau} = 2g_{\mu\nu}\frac{\mathrm{d}^2x^\mu}{\mathrm{d}\tau^2}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau} = \frac{2}{mc^2}g_{\mu\nu}f^\mu\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}. \tag{15.52}$$

The rhs is a Lorentz invariant quantity. In a new frame,

$$g_{\mu\nu}f'^\mu\frac{\mathrm{d}x'^\nu}{\mathrm{d}\tau} = g_{\mu\nu}\left(\Lambda^\mu_{\ \rho}f^\rho\right)\frac{\left(\Lambda^\nu_{\ \sigma}\mathrm{d}x^\sigma\right)}{\mathrm{d}\tau} = \left(g_{\mu\nu}\Lambda^\mu_{\ \rho}\Lambda^\nu_{\ \sigma}\right)f^\rho\frac{\mathrm{d}x^\sigma}{\mathrm{d}\tau} = g_{\rho\sigma}f^\rho\frac{\mathrm{d}x^\sigma}{\mathrm{d}\tau}.$$

We are therefore allowed to compute $\dfrac{\mathrm{d}\Omega}{\mathrm{d}\tau}$ in any reference frame we wish. Let us choose the rest frame of the particle, where

$$x^\mu = (ct, \vec{0}), \quad f^\mu = (0, \vec{F}_{\text{Newton}}).$$

We obtain:

$$\frac{\mathrm{d}\Omega}{\mathrm{d}\tau} = \frac{2}{mc^2}\left(f^0\frac{\mathrm{d}x^0}{\mathrm{d}\tau} - \vec{f}\cdot\vec{x}\right) = 0. \tag{15.53}$$

Therefore, the quantity Ω is always a constant:

$$\Omega(\tau) = \text{constant}. \tag{15.54}$$

If for some initial value τ_0 we choose the constant to be one, we will have

$$\Omega(\tau) = \Omega(\tau_0) = 1, \qquad \forall\ \tau. \tag{15.55}$$

Exercise 15.2. Calculate the trajectory of a particle on which the four-vector force exerted is $f^\mu = 0$.

15.6 Energy and momentum

We can define a relativistic four-vector analogue of momentum as

$$p^\mu = mc\frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} \tag{15.56}$$

We have that

$$\mathrm{d}\tau = \left(c^2\,\mathrm{d}t^2 - \,\mathrm{d}\vec{x}^2\right)^{\frac{1}{2}} = c\,\mathrm{d}t\left[1 - \left(\frac{\mathrm{d}\vec{x}}{c\,\mathrm{d}t}\right)^2\right]^{\frac{1}{2}} = c\,\mathrm{d}t\left[1 - \frac{\vec{v}^2}{c^2}\right]^{\frac{1}{2}} = \frac{c\,\mathrm{d}t}{\gamma}. \tag{15.57}$$

Thus, for the time-component ($\mu = 0$) of the four-momentum we have

$$p^0 = mc\frac{\mathrm{d}x^0}{\mathrm{d}\tau} = m\gamma c. \tag{15.58}$$

For the space-components ($\mu = i = 1, 2, 3$) we have

$$p^i = mc\frac{\mathrm{d}x^i}{\mathrm{d}\tau} = m\gamma\frac{\mathrm{d}x^i}{\mathrm{d}t} = m\gamma v^i. \tag{15.59}$$

For small velocities, we can expand the factor γ as

$$\gamma = \left[1 - \frac{v^2}{c^2}\right]^{-\frac{1}{2}} \approx 1 + \frac{1}{2}\frac{v^2}{c^2} + \mathcal{O}\left(\frac{v^4}{c^4}\right). \tag{15.60}$$

Therefore, for small velocities the space-components of the four-momentum become the classical momentum,

$$p^i \approx mv^i + \ldots, \tag{15.61}$$

while the time-component becomes

$$p^0 \approx mc + \frac{1}{2c}mv^2 + \ldots \tag{15.62}$$

In the second term of the above expansion we recognize the kinetic energy $\frac{1}{2}mv^2$ of the particle. We then identify the relativistic energy of a particle with

$$E = cp^0 = m\gamma c^2. \tag{15.63}$$

Eliminating the velocity $\vec{v}$ from Eqs 15.59-15.63, we obtain the relation:

$$E = \sqrt{\vec{p}^2c^2 + m^2c^4} \tag{15.64}$$

15.7 The inverse of a Lorentz transformation

Recall the metric matrix

$$g_{\mu\nu} = \text{diag}(1, -1, -1, -1) \tag{15.65}$$

We define an inverse

$$g^{\mu\nu}: \quad g^{\mu\nu} g_{\nu\rho} = \delta^{\mu}_{\ \nu}, \tag{15.66}$$

where $\delta^{\mu}_{\ \nu}$ is the Kronecker delta. It is easy to verify that the inverse of the metric is the metric itself:

$$g^{\mu\nu} = g_{\mu\nu} = \text{diag}(1, -1, -1, -1). \tag{15.67}$$

Now consider a Lorentz transformation $\Lambda^{\mu}_{\ \nu}$, which satisfies the identity:

$$\Lambda^{\mu}_{\ \rho} \Lambda^{\nu}_{\ \sigma} g_{\mu\nu} = g_{\rho\sigma}. \tag{15.68}$$

We can prove that the matrix

$$\Lambda_{\mu}^{\ \nu} \equiv g_{\mu\rho} g^{\nu\sigma} \Lambda^{\rho}_{\ \sigma} \tag{15.69}$$

is the inverse of $\Lambda^{\mu}_{\ \nu}$. Indeed

$$\Lambda^{\mu}_{\ \lambda} \Lambda_{\mu}^{\ \nu} = g_{\mu\rho} g^{\nu\sigma} \Lambda^{\rho}_{\ \sigma} \Lambda^{\mu}_{\ \lambda} = g_{\sigma\lambda} g^{\nu\sigma} = \delta^{\lambda}_{\ \nu}. \tag{15.70}$$

If $\Lambda^{\mu}_{\ \nu}$ is a velocity $\vec{v}$ boost transformation of Eq. 15.45, then

$$\Lambda_0^{\ 0}(\vec{v}) = \gamma = \left(1 - \frac{\vec{v}^2}{c^2}\right)^{-\frac{1}{2}}, \quad \Lambda_i^{\ 0}(\vec{v}) = \Lambda_0^{\ i}(\vec{v}) = -\gamma \frac{v^i}{c}.$$
$$\Lambda_i^{\ j}(\vec{v}) = \delta_i^j + \frac{v^i v^j}{v^2}(\gamma - 1) \tag{15.71}$$

We, therefore have that the inverse of a boost is

$$\Lambda_{\mu}^{\ \nu}(\vec{v}) = \Lambda^{\mu}_{\ \nu}(-\vec{v}), \tag{15.72}$$

as we also expect physically.

15.8 Vectors and Tensors

It is now time to give officially a definition for vectors in special relativity. We call any set of four components which transform according to the rule:

$$V^\mu \to V'^\mu = \Lambda^\mu_{\ \nu} V^\nu \tag{15.73}$$

a *contravariant* vector. Contravariant vectors transform in the same way as space-time coordinates x^μ do under homogeneous Lorentz transformations.

Not all vectors transform as contravariant vectors. Consider the derivative $\frac{\partial}{\partial x^\mu}$. Under a Lorentz transformation, it transforms as:

$$\frac{\partial}{\partial x^\mu} \to \frac{\partial}{\partial x'^\mu} = \frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial}{\partial x^\rho}. \tag{15.74}$$

We have that

$$\left(\frac{\partial x^\rho}{\partial x'^\mu}\right)\left(\frac{\partial x'^\mu}{\partial x^\nu}\right) = \delta^\rho_{\ \nu} \rightsquigarrow \left(\frac{\partial x^\rho}{\partial x'^\mu}\right)\Lambda^\mu_{\ \nu} = \delta^\rho_{\ \nu}. \tag{15.75}$$

Therefore, $\left(\frac{\partial x^\rho}{\partial x'^\mu}\right)$ is the inverse of a Lorentz transformation $\Lambda^\mu_{\ \nu}$:

$$\frac{\partial x^\nu}{\partial x'^\mu} = \Lambda_\mu^{\ \nu}. \tag{15.76}$$

Substituting into Eq. 15.74, we find:

$$\frac{\partial}{\partial x^\mu} \to \frac{\partial}{\partial x'^\mu} = \Lambda_\mu^{\ \rho}\frac{\partial}{\partial x^\rho}. \tag{15.77}$$

We found that the derivative does not transform according to the Lorentz transformation but according to its inverse. All vectors which transform with the inverse Lorentz transformation:

$$U_\mu = \Lambda_\mu^{\ \nu} U_\nu, \tag{15.78}$$

are called covariant vectors.

For every contravariant vector U^μ there is a dual vector

$$U_\mu = g_{\mu\nu}U^\nu. \tag{15.79}$$

We can invert the above equation multiplying with $g^{\rho\mu}$,

$$g^{\rho\mu}U_\mu = g^{\rho\mu}g_{\mu\nu}U^\nu = \delta^\rho_{\ \nu}U^\nu = U^\rho. \tag{15.80}$$

The dual vector U_μ is a covariant vector. Indeed, under a Lorentz transformation we have

$$U_\mu \to U'_\mu = g_{\mu\nu}U'^\nu = g_{\mu\nu}\Lambda^\nu{}_\rho U^\rho = g_{\mu\nu}\Lambda^\nu{}_\rho g^{\rho\sigma}U_\sigma = \Lambda_\mu{}^\sigma U_\sigma. \tag{15.81}$$

The scalar product of a contravariant and a covariant vector

$$A \cdot B \equiv A^\mu B_\mu = A_\mu B^\mu = g_{\mu\nu}A^\mu B^\nu = g^{\mu\nu}A_\mu B_\nu \tag{15.82}$$

is invariant under Lorentz transformations. Indeed,

$$A \cdot B \to A' \cdot B' = A'^\mu B'_\mu = \Lambda^\mu{}_\rho A^\rho \Lambda_\mu{}^\sigma B_\sigma = \delta^\sigma_\rho A^\rho B_\sigma = A_\rho B_\rho = A \cdot B. \tag{15.83}$$

Let us define for short:

$$\partial_\mu \equiv \frac{\partial}{\partial x^\mu}, \tag{15.84}$$

and the dual contravariant vector:

$$\partial^\mu = \frac{\partial}{\partial x_\mu} = g^{\mu\nu}\partial_\nu. \tag{15.85}$$

The D' Alembert operator is the scalar product:

$$\Box \equiv \partial^2 \equiv \partial_\mu \partial^\mu = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2. \tag{15.86}$$

Due to it being a scalar product, the D'Alembert operator is invariant under Lorentz transformations.

Finally, we define a tensor with multiple "up" and/or "down" indices to be an object

$$T^{\mu_1\mu_2\cdots}_{\nu_1\nu_2\ldots} \tag{15.87}$$

which transfroms as:

$$T^{\mu_1\mu_2\cdots}_{\nu_1\nu_2\ldots} \to \Lambda^{\mu_1}_{\rho_1}\Lambda^{\mu_2}_{\rho_2}\ldots\Lambda_{\nu_1}{}^{\sigma_1}\Lambda_{\nu_2}{}^{\sigma_2}\ldots T^{\rho_1\rho_2\cdots}_{\sigma_1\sigma_2\ldots} \tag{15.88}$$