Lecture notes for Mechanics 1

Misha Rudnev

1 On principles. Introduction

If one studies natural phenomena, it is important to try to understand the underlying *principles.* These would ideally not only enable one to explain the range of familiar phenomena but may predict new phenomena or at least explain new phenomena when they are discovered. The method of principles was founded by Newton (1643–1727). Einstein (1879–1955) was a great master of the method of principles.

Principles of physics are not a matter of logical explanation. Their confirmation is experience only – they are based on reality, or experimental facts which so far have had no evidence to the contrary. To this end, "experiment" and "reality" are herein tautological. Practically, however, these are not the principles themselves – as there aren't so many – but their logical consequences that are observed in a vast variety of physical experiments and in general reality.

Here are the examples of physical principles. *The relativity principle* – there are no observational consequences of absolute motion, see footnote 6 below, is a meta-principle which establishes a class of subjects, or observers, who are to embark on the study of natural phenomena and get the same results of identical independently performed physical experiments. These are closely related to Newton's first law, the most philosophically mysterious one of three. See Section 3.2.

Another meta-principle: physics works by way of fundamental constants (such as, say the mass of electron, the Gravitational constant, etc.) fundamental quantities (such as, say mass, energy, momentum) and laws, which can be expressed as the relations between these fundamental quantities and constant.

Experience, or information that we possess is always limited, and so may be the scope of principles. Principles, discovered so far apply with limited precision only. Until now, the key trend in physics was expansion and extension of the principles' scope. When new phenomena are discovered, it may happen that old principles cannot account for them and have to be abandoned or extended. Abandonment would indicate that the principles were somewhat false (although they may have used logically consistent mathematics). An ideal physicist is always in quest for an experiment that would invalidate his favorite theory. But even if this happens and the theory goes busted, yet the principles can be extended to embrace a new, more mature theory – this is hopefully an indication that one is on the right track. This happened in the beginning of the XX century when in order to apply classical Newtonian mechanics to the microworld on sub-atomic length scales ($\lesssim 10^{-8}$cm) and, respectively, the fast world of speeds comparable with the speed of light in vacuum $c \approx 300,000$ km/s, its principles had to be revised to embrace quantum mechanics and relativity theory. Throughout the revision some ideas and models that had outlived themselves, such as luminiferous ether[1] had to be dropped. New fundamental constants (c,h – Plank's constant) had to be added. However, the main concepts of Classical mechanics (such as mass, energy, momentum, etc.) not only survived, but after being properly examined and extended ended up being understood more thoroughly, provided more evidence in support of the depth and validity of these concepts, as well as the method of principles. Both quantum and relativistic mechanics have Newtonian mechanics as its limit case (of large sizes and slow speeds). Within its range, Classical mechanics is most widely used, and if, in fact, one attempted to study the problem of, say, snooker ball movements using the full might of relativistic quantum mechanics, he would be hopelessly lost.

Many physicists have believed and many believe that some day, and soon, *all* fundamental principles of physics will have been discovered and understood in terms of *Grand Unification.* This has not happened so far. Even if it does happen, new principles will probably be required for progress in other natural sciences, dealing with more complex phenomena, such as chemistry and, above all, biology. However, today's vast array of experimental data in all natural sciences makes it very unlikely that some day the notions of, say, momentum and energy will have to be abandoned completely and replaced by totally different ones.

[0]Maths, University of Bristol, *m.rudnev@bris.ac.uk,* www.maths.bris.ac.uk/∼maxmr/mech1.html

[1]As physics studies natural phenomena, it is impossible to understand it without having some basic knowledge about them. A student, therefore, is expected to know and be interested in basic facts of physical reality: the Earth moves around the Sun along an elliptic orbit which is close to a circle, an airplane flies due to the lift force that arises as it moves through air and is due to the shape of the wing, etc. If some notions, like *luminiferous ether* above sound unfamiliar, Wikipedia provides a quick and reasonably reliable reference to these.

Physical theory is derived from the principles by means of mathematics. Physical expression is unthinkable without mathematics, which provides abstract models to physical objects: a moving particle becomes a point, a ray of light a straight line, etc. Physical principles become axioms in the mathematical theory which is then developed to derive their logical consequences as theorems. However, pure mathematics itself is an abstract and logically closed and consistent discipline: it may not happen that new evidence comes to life and renders a mathematical theorem which is true today false tomorrow (unless the proof was false from the beginning ... hopefully false proofs are quickly identified, usually by their authors). So all the theorems that follow from abstract mathematical modeling of empirical physical principles are 100% true mathematically, while physically they are only meaningful within the scope of validity of the principles involved. It shall be mentioned that any mathematical theory is limited by the variety of purely logical connections that exist between its concepts. Hence, a considerable extension of physical principles would usually stimulate the development of new mathematics that would be adequate to describe it. For instance, real variables' calculus does not suffice as adequate mathematical machinery for quantum mechanics. Similarly, today's theoretical physics requires and stimulates development of very advanced mathematical techniques.

However, many fundamental questions in pure maths may have no meaning in physics. Such is, for instance, the concept of an irrational number. The proofs that π is irrational (Lambert, 1761) and transcendental (von Lindemann, 1882) are regarded as important milestones in maths. In physics, however, it makes no sense to ask whether Plank's constant $h \approx 6.63 \times 10^{-34}$ joule/s is rational or irrational in a given system of units, because any measurement of its value inevitably involves error. This is not an equipment flaw, but a physical law: there are no precise measurements. Nor there is evidence as to whether rationality or irrationality of dimensionless constants, such as the *fine structure constant* $\alpha = \frac{e^2}{2hc\epsilon_0} \approx 1/137$, ($e \approx 1.6 \times 10^{-19}$ coulombs is the charge of the electron and ϵ_0 is dielectric permitivity of vacuum) is of any physical consequence. In the same vein, defining, e.g. density of water as the derivative

$$\rho = \frac{dm}{dV}$$

when dV is the volume of an "infinitesimally" small ball and dm the mass of water therein, one is, in fact, dealing with finite, however small, quantities dV and dm. Indeed, it is not at all clear what dm means if the diameter of the volume dV becomes smaller than the actual size of the water molecule. What's more, the molecules are in fact in the state of constant motion, and so to ensure that ρ does not depend on time in some possibly very difficult way, one has to ensure that dV is large enough, so the number of molecules contained therein is approximately constant. So dV shall realistically be rather large physically but small enough mathematically, so that one can take limits. The "physical" density is therefore

$$\rho = \frac{\Delta m}{\Delta V},$$

where the volume ΔV is small enough, but no too small. Fortunately, it happens that in the mathematical model dealing with the above macroscopic notion of water density, the functions involved are smooth enough, so that the "mathematical limit" $\frac{dm}{dV}$ equals approximately, but with extremely high accuracy, to the "physical limit" $\frac{\Delta m}{\Delta V}$. That is why the issue of scales and orders of magnitude is extremely important in physics.

In general, the question of an adequate mathematical model for a physical phenomenon is fundamental and often difficult, apart from purely mathematical difficulties within the model. On the other hand, physics can often content itself with mathematical statements that a pure mathematician would find not strict enough. As physics always has to deal with approximate values of its quantities, it often makes sense to use mathematical approximation and simply ignore all the mathematical subtlety that lies beyond this approximation.[2]

Another instance of occasional principal differences between physicist's and mathematician's viewpoints is the question of long-term stability of complex systems. An example of such is the Solar system, and the question

[2] For instance, the relativistic phenomenon of light aberration consists in the fact that a ray of light in the xy-plane forming an angle α with the x axis from the point of view of observer O who holds the flashlight will form a different angle α' from the point of view of observer O' who is moving uniformly along the x axis with velocity v. By using the Taylor expansions where terms containing the second and higher derivatives have been dropped, one can show that special relativity theory predicts that $|\alpha' - \alpha| \sim \frac{v}{c} \sin\alpha$ (which shows that O and O' indeed agree upon the direction of the x-axis where $\sin\alpha = 0$ but disagree on the direction of the y-axis, where $\sin\alpha = 1$). From a physicist's point of view, as long as v is reasonably small, the above formula can be treated as precise, simply because the relative error involved has the magnitude $\sim v^2/c^2$, which may be well beyond resolution of the angular measurements involved.

Relativity theory is not taught systematically in this course, but allusions to its principles and some phenomena are being made throughout.

is, roughly speaking, whether the influence of other planets, small with respect to the influence of the Sun, will have a serious long-term cumulative effect on the Earth's orbit, so as to make it unbounded. This would not happen had each planet interacted with the Sun only, in isolation from other planets, in which case their motion is described by Kepler's laws which are some 400 years old. But in reality the motions of each individual planet affects others, if only just a tiny bit. The question is: may these tiny effects accumulate over the years to cause a catastrophe? The natural physical parameter to measure the perturbation caused by all other planets is the ratio of their total mass to the mass of the Sun, which is $\sim 10^{-3}$. Mathematically, the problem falls into the realm of perturbation theory of Dynamical systems and is wide open. The fact is, very few actual problems enable a complete mathematical resolution. Most of the time, given a system of, say differential equations describing a particular mechanical system, mathematics fails to provide a mere formula that would just enable one to write down the solution.

Complex dynamical systems are known to be extremely unstable: their two arbitrarily close initial states may end up evolving over very long times by eventually utterly different scenarios. Even if the mathematical problem becomes solved, the Yes/No answer to the question of stability depends on the exact positions and velocities of the Earth, as well as all the other planets in the Solar system at a given moment of time. These positions, however, cannot ever be known exactly, and moreover, in the process of measurement, which involves some kind of interaction with the measuring device, will change. There are other serious obstacles: 10^{-3} still appears to be way too large for a value of a "mathematically valid" perturbation parameter for the mathematical theory to work. On the other hand, if the perturbation were small enough, the characteristic times over which a structural change may take place can be proved to be exponential in some inverse power of the perturbation parameter. These times would exceed the age of the Universe, and it is therefore unlikely that the mathematical model considering only non-relativistic gravity between the Sun and the planets as an isolated system can be adequate on such time scales.

Mechanics appeared as the earliest branch of physics. Mechanics studies motion and equilibrium of physical bodies. By motion its simplest form is meant: motion in mechanics is change of position relative to other bodies. Newton in his *Principia* (first published in 1687) was the first to formulate the system of principles of mechanics, and although he had many great predecessors, such as Archimedes (circa 287–212 b.c.), Kepler (1571–1630), Galileo (1564–1642), Huygens (1629–1695) and others, Newton is regarded as a founder of modern physics (and, as a matter of fact, a co-founder of differential-integrable calculus, which provides a natural mathematical language to express Newton's laws and their consequences). For 200 years after Newton, in spite of the industrial revolution of the XIX century, principles of Newtonian mechanics "worked" in all areas of human endeavor and needed not to be revised. The revisions came only in the beginning of the XX century and concerned atomic length scales and speeds comparable to the speed of light, which before the end of the XIX century had been simply out of reach. Mechanics of fast moving objects is called *relativistic* and its main principle is that no body can be accelerated to move faster than the speed of light in vacuum. That is, there is a limit speed, with which, in particular, any signal or any interaction can travel. In cosmic rays and modern accelerators particles move with speeds that are only by fractions of a millimeter per second smaller than c. Accelerators have been designed in accordance with principles of relativistic mechanics, and the fact that they work is a solid proof of its validity as a principle model of the world, at least within of out today's reach.

In addition to dealing with non-relativistic motions, Classical mechanics deals with *macroscopic* objects, namely those whose size is big enough to ignore the *uncertainty relation* discovered by Heisenberg (1901–1976). It states that if one makes a simultaneous measurement of any particle's position x and momentum p, the uncertainties δx and δp of measurement will always be such that

$$\delta x \cdot \delta p \gtrsim h,$$

the Plank constant. This is not a matter of how good the measuring equipment is, but *is* rather the fact that the two quantities, the position and momentum, cannot be in principle known with precision simultaneously, and making more precise the knowledge of one inevitably obscures the knowledge of the other. Thus, the central notion of Classical mechanics of a trajectory (where one is expected to know both the position and velocity at any time) is, in fact, an approximation. This approximation works well for bodies that are large enough, but for small bodies, whose size characteristic size x is small (while δx should not exceed the characteristic size) Classical mechanics is inapplicable.[3] Heisenberg's principle was necessary to explain the results of the whole

[3]Indeed, if one decides to talk about a trajectory of an electron in the hydrogen atom, then the error δx should be at most the

family of experiments whose outcomes did not make sense from the classical mechanics point of view. To embrace this principle, mathematical theory had to be extended to quantum mechanics, where a state of a particle is being characterised by a wave function – a vector in an infinite-dimensional Hilbert space – rather than a three-dimensional radius-vector $\boldsymbol{r} = (x, y, z)$. The mathematics involved goes beyond differential calculus, and instead of ordinary differential equations expressing the laws of classical mechanics, the laws of quantum mechanics are expressed vial partial differential equations. Quantum mechanics has classical mechanics as its limit case when one may regard h as zero. Considering rather small numerical value of h, there is no wonder that classical mechanics is so precise in its applications to real life, which include cars, airplanes, and missiles.

However, in both cases fundamental notions of Newtonian mechanics, such as energy, momentum, angular momentum, etc. not only were not abandoned, but after having found their clear extensions (which in the "big and slow" classical limit boil down to their classical definitions) have become foundation for relativistic and quantum mechanics.

2 Kinematics

Kinematics studies motion of particles regardless of its cause. The venue for kinematics is space-time, i.e. an *event*, an instant of reality, is characterised by where and when it has occurred. The main object for kinematics is a particle – a point in space which has no size, and whose radius-vector $\boldsymbol{r}$ is a function of time t, or the point moves along its *trajectory*. Kinematics is concerned with the most basic question – how does one describe a trajectory?

2.1 Space, time, and frames of reference

Motion in mechanics is understood as change of position of a mechanical body in space over time, relative to other bodies. A rail passenger opens a book in London, reads it until the train arrives in Bristol whereupon he closes it. The events of the book being open and closed happen at the same place – in his hands – from the passenger's point of view. On the other hand, the former event takes place in London and the latter one in Bristol. Hence, the statement that two events occur at the same place is meaningless until the frame of reference, relative to which the statement is being made, has been specified. And according to Galileo's relativity principle, there is no reason to believe that there is anything physically special about someone reading a book on a train London to Bristol versus doing this in London or in Bristol.

Construction of a frame of reference begins with specifying the reference point O, or the origin, together with three different directions from it. It is convenient to make the three directions mutually perpendicular. Mathematically, these directions, or coordinate axes (x, y, z) are identified by unit vectors $(\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k})$, and a position of any point in the resulting Euclidean three-dimensional vector space is now

$$\boldsymbol{r} = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}.$$

The fact that space is Euclidean, i.e. that it satisfies the axioms of Euclidean geometry, is another principle that cannot be proved, but so far on length scales from sub-atomic to Metagalaxy it has been verified with high precision. (General relativity generalises Euclidean geometry to Riemannian geometry where space acquires curvature. Riemannian geometry is nevertheless based on the same set of fundamental concepts as Euclidean geometry, such as length, angle between, ant parallel translation of vectors.)

The origin O of one coordinate system can be translated to any other point O' and a new coordinate system thereby obtained will be equally adequate. Besides, the frame of reference can be rotated around the origin, preserving the rigidity of mutual alignment of the three unit direction vectors, and there is no preferred direction for, say the x-axis. In addition, one has freedom in identifying the direction of the z-axis "up" or "down" after the directions of the x and y axes have been specified. If the direction of the z-axis is determined from the direction of the x, y axes by the right-hand-rule, then the coordinate system is called *right*, otherwise *left*. By

atom size, which is some 10^{-10}m. On the other hand, $p = mv$, with $m \approx 9.11 \times 10^{-31}$kg. The uncertainty principle than tells

$$\delta v \gtrsim \frac{6.64 \times 10^{-34}}{9.11 \times 10^{-31} \cdot 10^{-10}} \approx 7 \times 10^6 \text{ m/s},$$

which is in excess of the speed of the electron in the atom known to be about 10^6m/s. On the other hand, for the velocity of a ball of mass $1g$ the same calculation yields the uncertainty $\delta v \approx 10^{-20}$m/s, which is clearly negligible.

default, right-hand coordinate systems are used, but it is only because most of the people are right-handed, plus the kinematics of the Earth, Moon and Sun, wherein lie the origins of calendars, i.e. clocks and motions, is in a sense right-handed as well.

A rotation of a coordinate system whereupon the unit directions, or *basis* $(\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)$, turn into $(\boldsymbol{e}'_1, \boldsymbol{e}'_2, \boldsymbol{e}'_3)$ can be described as follows. Given any vector $\boldsymbol{a}$, whose coordinates equal (a_1, a_2, a_3) relative to the "old" basis $(\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)$, its coordinates (a'_1, a'_2, a'_3) with respect to the "new" basis $(\boldsymbol{e}'_1, \boldsymbol{e}'_2, \boldsymbol{e}'_3)$ will be given by the dot products $a \cdot \boldsymbol{e}'_1$, $a \cdot \boldsymbol{e}'_2, a \cdot \boldsymbol{e}'_3$, respectively, since all $\boldsymbol{e}$'s are unit vectors. Since $\boldsymbol{a} = a_1\boldsymbol{e}_1 + a_2\boldsymbol{e}_2 + a_3\boldsymbol{e}_3$, one gets:

$$\begin{array}{rcl} a'_1 & = & r_{11}a_1 + r_{12}a_2 + r_{13}a_3, \\ a'_2 & = & r_{21}a_1 + r_{22}a_2 + r_{23}a_3, \\ a'_2 & = & r_{31}a_1 + r_{32}a_2 + r_{33}a_3, \end{array} \tag{1}$$

where

$$r_{ij} = \boldsymbol{e}'_i \cdot \boldsymbol{e}_j = \cos\alpha_{ij}, \qquad i, j = 1, 2, 3,$$

where α_{ij} is the angle between the directions of the ith new (primed) and jth old (non-prime) axis. Using matrix notation

$$\begin{pmatrix} a'_1 \\ a'_2 \\ a'_3 \end{pmatrix} = R \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix},$$

where R is the 3×3 matrix whose entries are r_{ij}. Observe the following property of R. If one takes its transpose R^T, i.e replaces rows of R by its columns, the entries of R^T at the position ij are now the cosines of the angles between the directions of the ith old axis and jth new axis. This corresponds to "rotating backwards", i.e. expressing

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = R^T \begin{pmatrix} a'_1 \\ a'_2 \\ a'_3 \end{pmatrix}.$$

Combining the two expressions we see that

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = R^T R \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix},$$

and as this is true for any triple of real numbers (a_1, a_2, a_3) it means that $R^T R = \mathrm{Id}$, the identity matrix, whose entries are 1 on the main diagonal (when $i = j$) and 0 otherwise. This means that

$$R^{-1} = R^T,$$

and matrices with such properties are called *orthogonal*, in particular, their determinant must equal to 1.

Physically, the statement "consider the frame of reference" with origin O and coordinate directions $(\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k})$ means that one assumes the principal possibility of first and foremost defining straight lines. "Physical" straight lines are provided by rays of light. It is confirmed experimentally that rays of light in vacuum indeed are "straight", that is satisfy the axioms and theorems of Euclidean geometry with high precision. (In the atmosphere, due to changes in the refraction coefficient, which is the ratio $\frac{c_a}{c}$, $c_a \leq c$ being the speed of light in the atmosphere, light rays are no longer straight. But the way they curve can be described and calculated, using principles of Euclidean geometry.) Furthermore, one assumes the ability to measure the distance to any point in space, as well as the angle that the direction to that point forms with the coordinate axes – elementary trigonometry will then produce the point's coordinates.

Then there is a question of what length is. If there is a straight line and a "standard" rod, whose length is, say, one meter, then one can coordinatise this line using the rod. The French originated the meter in the 1790s as one/ten-millionth of the distance from the equator to the north pole along a meridian through Paris. It is realistically represented by the distance between two marks on an iron bar kept in Paris. The International Bureau of Weights and Measures, created in 1875, upgraded the bar to one made of 90 percent platinum/10 percent iridium alloy.

In 1960 the meter was redefined as 1,650,763.73 wavelengths of orange-red light, in a vacuum, produced by burning the element krypton (Kr-86). More recently (1984), the Geneva Conference on Weights and Measures has defined the meter as the distance light travels, in a vacuum, in 1/299,792,458 seconds with time measured by a cesium-133 atomic clock which emits pulses of radiation at very rapid, regular intervals. This takes us to the notion of time.

The last piece of conceptual equipment one needs to complete construction of a frame of reference is the clock. It is an empirical fact that has never been violated: stating that something occurred at place $\boldsymbol{r}$ and time t describes an event unambiguously. Mathematically, an *event* is a point with four "coordinates" (x, y, z, t) in the four-dimensional space-time continuum. Yet despite in the (x, y, z) space the notion of angle is well defined, one should not think that there is a particular angle between, say, the x and t axes.

Physically, a clock is any repeated periodic process that one uses to measure time. It is clear what "periodic" means mathematically, and physically it is desirable that the clock was highly "uniform" and reliable. In relativistic mechanics, which postulates the existence of the world constant c =speed of light in vacuum, as the maximum speed with which interaction or any signal can travel, a uniform clock can be defined unambiguously as follows. An observer as O places a mirror at some point A different from O so that a ray of light sent from O to A will be reflected back to A. As soon as the signal returns to O, the "experiment" is repeated. The sequence of events of departure/arrival of the signal are all equally separated in time. Any other physical process can be tested as periodic with respect to the above "light clock", and if it is periodic, the process can be used as a clock. In the past, most reliable clocks were astronomical, based on periodicity of Earth's movement with respect to distant stars. Today's most precise, i.e. "most periodic" clocks are the atomic cesium-133 ones.

The next question is that of how two events occurring at different space loci can be rendered as simultaneous. Classical physics regards time as absolute, and therefore the answer as to whether two events occur at the same time is straightforward and the same for every classical observer. But why should time be absolute if space is not? For a long time there was no evidence that simultaneity is devoid of absolute meaning, it was nt until the end of the XIX century that the industrial revolution enabled measuring devices enable to detect consequences of relativity of simultaneity experimentally.

The question of simultaneity is equivalent to the question of how to synchronise clocks. Indeed, if one wants to describe a trajectory $\boldsymbol{r}(t)$, one shall be able to know the time t at the location $\boldsymbol{r}$. So hypothetically one can imagine that there is an identical clock, namely a clock based ont eh same physical principle, say a cesium-133 clock whatever drives it, ticking at every point of space. Each clock measures *time intervals* in the same way, but all the clocks may be showing different times. If information could travel infinitely fast, it would be possible to send a signal from the origin to all clocks, instructing them to set the time equal to 1pm ten minutes after receiving the signal. The signal would reach all the clocks immediately, and ten minutes from then synchronisation would be a done deal.

But if there is a fundamental limit as to how quickly information can travel, in terms of c, then the above way is unacceptable. The only conceptual way to *postulate* simultaneity of two events is as follows. Suppose A, B are points on the x axis, symmetric with respect to the origin O, A being to the right. There is an instantaneous flash at O, and light from it reaches the points A and B simultaneously. Let us now consider another observer O' who moves in the positive direction of the x axis, relative to O in such a way that when the flash occurs O and O' coincide. Form the point of view of O', the point A is approaching, while the point B is receding. But the speed of light, the world's fundamental constant on which all the observers must agree, is still equal to c. So for O the signal will come to A first, and then to B: the two events are not simultaneous. This definition is not ad hoc: it involves an experimentally very well confirmed principle that there exists a fundamental physical constant equal to the *maximum* speed of sending information (which therefore must be the same for all observers, moving or not, in the same fashion as h or e are) and a logical construction.[4]

Similarly, the notion of size becomes relative. More precisely, suppose we have a horizontal rod moving along

[4] An interesting "paradox" that can be derived (this is a version of the so-called *pole and barn* paradox). Suppose, O' sees the following drama: A receives the signal first and shoots B dead before B receives anything. Consider another observer O'' who is moving in the negative x-direction. From his point of view the signal comes to B before it comes to A. Suppose, at the moment when B receives the signal, O'' sees B shoot A dead, before A gets the signal. How could then A shoot B dead if he was dead before having a chance to fire a shot? The answer is, of course – in order for the above drama to unravel, the bullet would have to travel with the speed at least $2c$, which is impossible. This construction implies that a possibility of "crossing the light barrier" would entail controversy in terms of the notions of cause and effect. On the other hand, one expects that the judgement on whether the event X caused the event Y must be the same for all observers. The existence of the maximum speed then implies that two events that have large spatial and small temporal separation cannot possibly have causal connection between each other.

the x axis. There is no difficulty measuring the length l_0 of the rod in the reference frame associated with the point O' in the middle of the rod, where the latter rests. However, regarding the observer at O, measuring the length of a rod would require marking the positions of the left and right ends of the moving rod *simultaneously.* As simultaneous for O and O' are different notions, the outcome l of length measurement by O will generally be different from l_0. There is only so much to the platinum-iridium meter!

Having discussed the issues of what a *frame of reference* is, it is now time to assume that we have it: we have separated the four-dimensional space-time into space and time and are now capable of determining with high accuracy spatial and temporal coordinates of events, using Euclidean geometry and synchronised clocks. Let us now consider a moving body. Suppose the size of the body, yet way within the "big" range for quantum-mechanical effects to be negligible, is on the other hand small enough, so that the body's position can be modeled by a single point in the abstract mathematical space $\mathbb{R}^3$. Such a body is called a *material point.* Clearly, whether or not a physical body can be regarded as a material point depends on a concrete problem considered: the Earth can be regarded as a point if one studies its motion around the Sum; one has to pay tribute to the non-zero size of the Earth considering its interaction with the Moon causing the tides, and finally, there is no meaning to thinking of the Earth as a point studying its axial rotation.

Note, however, that a more complicated physical body can always be regarded as a family of interacting material points.

2.2 Trajectory, velocity, acceleration, curvature.

We start out with some fixed frame of reference $(O, \boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}, t)$ and now study the motion of a *material point*, or *particle* P, i.e the dependence of its position $\vec{OP}$ with respect to the origin, as it changes in time. The vector-function

$$\vec{OP}(t) = \boldsymbol{r}(t) = x(t)\boldsymbol{i} + y(t)\boldsymbol{j} + z(t)\boldsymbol{k}$$

is called the *trajectory* of P. We often write just $\boldsymbol{r} = (x, y, z)$ as long as the vectors $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ have been fixed once and for all.

The first and second derivatives of the function $\boldsymbol{r}(t)$ are called the *velocity* and *acceleration*, respectively:

$$\boldsymbol{v}(t) = \frac{d\boldsymbol{r}(t)}{dt} = \dot{x}(t)\boldsymbol{i} + \dot{y}(t)\boldsymbol{j} + \dot{z}(t)\boldsymbol{k}, \qquad \boldsymbol{a}(t) = \frac{d\boldsymbol{v}(t)}{dt} = \frac{d^2\boldsymbol{r}(t)}{dt^2} = \ddot{x}(t)\boldsymbol{i} + \ddot{y}(t)\boldsymbol{j} + \ddot{z}(t)\boldsymbol{k}.$$

Time derivatives will be denoted by dots, rather than primes. Observe that the above formula looks simple enough due to the fact that the vectors $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ have been fixed, independent of time. Also observe that as $\boldsymbol{r}$ is a vector, then $\boldsymbol{v}$ and $\boldsymbol{a}$ are vectors as well, because $\boldsymbol{v} = \lim_{\Delta t \to 0} \frac{\Delta \boldsymbol{r}}{\Delta t}$ is obtained in the limit by multiplying a vector $\Delta \boldsymbol{r}$ by a real $\frac{1}{\Delta t}$, and $\boldsymbol{a}$ is obtained similarly from $\boldsymbol{v}$.

In view of this, one can define *relative velocity* and acceleration as follows. If $\boldsymbol{r}_1(t)$ is the trajectory of a particle P_1, and $\boldsymbol{r}_2(t)$ is the trajectory of a particle P_2, then the position of P_2 relative to P_1 is the vector

$$\boldsymbol{r}_{21}(t) = \boldsymbol{r}_2(t) - \boldsymbol{r}_1(t),$$

and relative velocity and acceleration are the quantities

$$\boldsymbol{v}_{21}(t) = \dot{\boldsymbol{r}}_{21}(t) = \dot{\boldsymbol{r}}_2(t) - \dot{\boldsymbol{r}}_1(t), \qquad \boldsymbol{a}_{21}(t) = \dot{\boldsymbol{v}}_{21}(t).$$

Geometrically, the velocity vector $\boldsymbol{v}$ is directed tangent to the trajectory at the point $\boldsymbol{r}(t)$.[5]

Similar to the trajectory $\boldsymbol{r}(t)$ one might plot the vector-function $\boldsymbol{v}(t)$ as well. The curve thereby obtained is called *hodograph.* The acceleration $\boldsymbol{a}$ plays the same role for the hodograph as $\boldsymbol{v}$ plays for the trajectory, in particular $\boldsymbol{a}$ is tangent to the hodograph.

The quantities $\boldsymbol{r}, \boldsymbol{v}, \boldsymbol{a}$ are vectors, and their addition is done by the parallelogram rule, as this is the case with radius-vectors in Euclidean space. Here is a subtle point, however. If one associates a different *frame of reference*

[5]Strictly speaking, $\boldsymbol{v} = \frac{dr}{dt}$ is a *definition* of a tangent vector to a curve, and only curves that are smooth enough possess tangent vectors. The notion of the tangent vector already becomes ambiguous for a broken line at those points where different segments meet. The coast of Britain is a continuous curve; however it is so jagged, or non-smooth, that is impossible to define a tangent line, except approximately, and, in fact, taking into account all the minuscule twists and turns of the coastal line would make its length grow to infinity.

with P_1, then the velocity $\boldsymbol{v}_2'$ of P_2 in that latter frame of reference is equal to $\boldsymbol{v}_{21}(t)$ only within the limits of classical mechanics, as in general the time t in the stationary frame whose origin is O and the time t' in the moving frame, whose origin is P_1 are different form one another, and by definition $\boldsymbol{v}' = \frac{d\boldsymbol{r}_{21}}{dt'}$, not with respect to dt. E.g. if both P_1 and P_2 move along the x-axis in opposite directions with speeds $c/2$, then $v_{21} = c$, and this does not contradict anything, because the particles move in different directions independently. On the other hand, invoking a formula, representing velocity addition law in special relativity yields $v' = \frac{4}{5}c$. So, if one associates an observer with the moving particle P_1, the latter sees the distance between P_2 and P_1 change at the rate $\frac{4}{5}c$, rather than c, according to the stationary observer at O. Needless to say, this is because $t \neq t'$.

2.2.1 Uniform motion and motion with constant acceleration

The simplest motion is the *uniform* motion when $\boldsymbol{a} = 0$, i.e. $\boldsymbol{v} = \boldsymbol{v}_0 = const.$ So, there are three independent first-order differential equations: $\frac{dx}{dt} = v_x$, $\frac{dy}{dt} = v_y$, $\frac{dz}{dt} = v_z$, with constant right-hand sides. The solutions are $x(t) = x_0 + v_x t$, $y(t) = y_0 + v_y t$, $z(t) = x_0 + v_x t$, or simply

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

where $\boldsymbol{r}_0 = (x_0, y_0, z_0)$ is the initial position Clearly, the trajectory is a straight line. If this is the case, one may wish to choose the x axis so that it is directed along $\boldsymbol{v}_0$, in which case the motion is simply $x(t) = x_0 + v_0 t$, while $y = z = 0$.

Nearly as simple is motion with constant acceleration $\boldsymbol{a} = const.$ In this case the differential equations are $\frac{d^2x}{dt^2} = a_x$, same independently for y and z, and the solution

$$\boldsymbol{r}(t) = \boldsymbol{r}_0 + \boldsymbol{v}_0 t + \frac{1}{2}t^2\boldsymbol{a},$$

where $\boldsymbol{v}_0$ is the initial velocity. One can always choose the origin, so that the plane defined by the vectors $\boldsymbol{v}_0, \boldsymbol{a}$ becomes the xy-plane. Furthermore, the y-axis can be chosen, so that $\boldsymbol{a}$ is collinear with it. So, the trajectory always lies in the plane $z = 0$, and is described by the system of two equations:

$$x = x_0 + v_x t, \qquad y = y_0 + v_y t + \frac{at^2}{2},$$

where a is the length of $\boldsymbol{a}$.

This is a parabola: without loss of generality let $x_0 = y_0 = 0$, then

$$t = x/v_x, \qquad y = v_y x/v_x + \frac{ax^2}{2v_x^2}.$$

The latter two formulas are all that one needs to deal with numerous free fall problems, where $a = \pm g$, the free fall acceleration, and the sign $\pm$ depends on whether the y axis is directed down or up. (This is part of kinematics, because one essentially does not need the second Newton's law $\boldsymbol{F} = m\boldsymbol{a}$ or gravity to account for it: it was Galileo who discovered that all bodies in vacuum fall with the same acceleration g. Observe that Newton's formula for the force $|\boldsymbol{F}| \sim m_1 m_2/r_{12}^2$ of gravitational attraction of two bodies with masses m_1, m_2 at a distance r_{12} does not explain the nature of gravity.)

2.2.2 Natural parameter, curvature, torsion

How does one describe a curve, *path* in three dimension? Imagine that there is a *test particle* moving along the curve, whatever causes the motion. Let us use the position, velocity, and acceleration of this particle to describe the curve. The ambiguity of such a description that as long as a test particle moves along a give path, its speed is irrelevant. So, it would be nice to have a "standard" particle, whose speed is always equal 1. This is the basis for using the "natural parameter" s to describe the inner geometric properties of the curve, rather than time t.

We have, and will use the notations v, a for the absolute values, or magnitudes $|\boldsymbol{v}|, |\boldsymbol{a}|$ (elsewhere these can be denoted as $\|\boldsymbol{v}\|, \|\boldsymbol{a}\|$) alias Euclidean lengths of the vectors $\boldsymbol{v}, \boldsymbol{a}$, respectively (the absolute value of the velocity is referred to as *speed*). E.g. $v = |\boldsymbol{v}| = \sqrt{v_x^2 + v_y^2 + v_z^2}$. The *distance* s_{12}, traveled by the material point between times t_1 and t_2 can be then computed by the formula

$$s_{12} = \int_{t_1}^{t_2} v(t)dt = \int_{t_1}^{t_2} \sqrt{\dot{x}^2(t) + \dot{y}^2(t) + \dot{z}^2(t)}dt.$$

Indeed, the distance traveled over an infinitesimal time dt is

$$ds = vdt,$$

and

$$s_{12} = \int_{t_1}^{t_2} ds.$$

Assuming that the trajectory "begins", say at $t = 0$, the distance traveled by the point is

$$s(t) = \int_0^t v(t)dt,$$

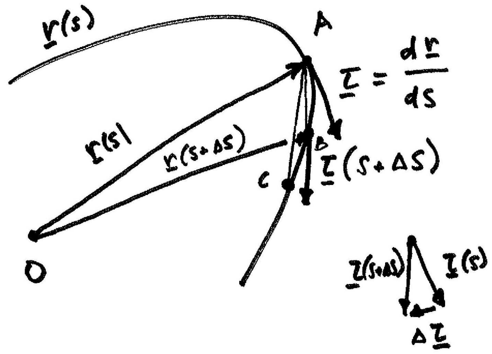
and as long as the particle keeps moving, this is clearly an increasing function of t. Therefore, t and s are in one-to-one correspondence and it is in principle possible to express unambiguously $t = t(s)$, and describe the trajectory as

$$\boldsymbol{r} = \boldsymbol{r}(t(s)),$$

as a function of s. Such a description is called *natural,* because it refers to a particular curve in space only in terms of the curve's *metric properties*, i.e. its length, regardless of any particle moving along this curve as trajectory. Equivalently, $t = s$ if a particle moving along the curve always has unit speed. Calling the derivative

$$\boldsymbol{\tau} = \frac{d\boldsymbol{r}}{ds},$$

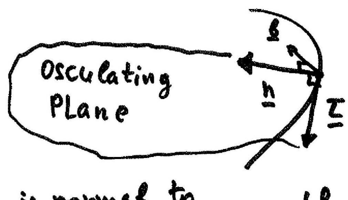
we observe that the magnitude of $\boldsymbol{\tau}$ is always 1. Indeed, $|d\boldsymbol{r}| = ds$: the absolute value of $d\boldsymbol{r}$ *is* the distance. So, $\boldsymbol{\tau}$ is a *unit* tangent vector to the trajectory: what changes along the trajectory is only the *direction* of $\boldsymbol{\tau}$. The figure illustrates this and the forthcoming concepts of the normal and binormal vectors.



curvature radius

curvature

$n = R\frac{d\tau}{ds}$, or $\frac{d\tau}{ds} = K n$



$b = [\tau \times n]$

b is normal to osculating plane, which is the limit plane of the triangle ABC, above

$\frac{db}{ds} = T n$

torsion

Plotting the dependence $\boldsymbol{\tau}(s)$ would yield a curve drawn on the unit sphere. A short chord to such a curve is almost perpendicular to the radius of the sphere, connecting its centre with the chord's endpoint, the angle limits to 90 degrees as the chord gets shorter. Therefore the derivative $\frac{d\boldsymbol{\tau}}{ds}$ is a vector which is perpendicular to $\boldsymbol{\tau}$. Similarly, and this will be used a lot,

$$\text{If } \boldsymbol{u}(t) \text{ is any vector-function, such that } |\boldsymbol{u}(t)| = const., \text{ then } \boldsymbol{u}(t) \cdot \frac{d\boldsymbol{u}(t)}{dt} = 0, \text{ for all } t.$$

The direction of the vector $\frac{d\boldsymbol{\tau}}{ds}$ will be determined as follows. The direction of $\boldsymbol{\tau}$ is determined as the limit direction of some sequence of chords $\Delta\boldsymbol{r}_1, \Delta\boldsymbol{r}_2, \Delta\boldsymbol{r}_3, \ldots$, to the trajectory that all begin at the same point $\boldsymbol{r}(t)$ and get shorter and shorter. In the figure $\Delta\boldsymbol{r}_1 = \vec{AC}$, $\Delta\boldsymbol{r}_2 = \vec{AB}$. Assuming that $\Delta\boldsymbol{r}_1$ and $\Delta\boldsymbol{r}_2$ are not parallel (i.e., the trajectory $\boldsymbol{r}(t)$ is not a straight line), let T_1 be the triangle determined by the vectors $\Delta\boldsymbol{r}_1$ and $\Delta\boldsymbol{r}_2$ (the triangle ABC in the figure). Similarly, let T_2 be the triangle determined by the vectors $\Delta\boldsymbol{r}_2$ and $\Delta\boldsymbol{r}_3$, and so on. Each triangle T_i determines a plane. The existence of the limit plane defined by the sequence of the triangles T_i s equivalent to the existence of the limit $\frac{d\boldsymbol{\tau}}{ds}$. The limit plane is called the *osculating plane* to the trajectory $\boldsymbol{r}(t)$ at a given point. The osculating plane clearly contains $\boldsymbol{\tau}$. The unit vector in the osculating plane, which is perpendicular to $\boldsymbol{\tau}$ and points inside the trajectory is now well defined. It is denoted as $\boldsymbol{n}$ and is called the *principal normal vector* to the trajectory.

If the osculating plane does not change along the curve, then it contains the curve itself, and the curve is called *plane*. In the latter case, one can always assume that the osculating plane is given as $z = 0$, and will only need two coordinates x, y to describe the curve. Otherwise, the curve is said to have a *twist*.

In any case, we have

$$\frac{d\boldsymbol{\tau}}{ds} = K\boldsymbol{n}, \tag{2}$$

for some real number K. This number is called *curvature* at a point P on the curve, and the quantity K^{-1} is called the *radius of curvature*. This is the definition of curvature.

Example – curvature of a circle. Computing curvature and torsion is not easy, due to a generally hard problem of determining the parameter change. $s(t)$ and $t(s)$. The idea, given a curve, is to mentally design a "test particle" following the curve in time in the easiest possible way, and then eliminate t in favour of the natural parameter s.

Consider the example when the trajectory $\boldsymbol{r}(t)$ is geometrically a circle of radius r. Of course, it is reasonable to try to describe the the circle via uniform circular motion along it. I.e., consider a particle, moving around the circle, so that its position on the circle is given by the angle α, and $\dot{\alpha} = \omega$, the angular velocity. The velocity vector $\boldsymbol{v}$ is tangent to the circle. Over the time dt the particle travels the distance $ds = rd\alpha = \omega r dt$ along the circle, therefore the speed $v = \omega r$ and

$$\boldsymbol{v} = \omega r \boldsymbol{\tau},$$

where $\boldsymbol{\tau}$ is the unit tangent vector to the circle. If $\omega = const.$, i.e. the *angular acceleration* $\dot{\omega}$ is zero, then the motion is periodic with period $T = \frac{2\pi}{\omega}$ and frequency $\nu = \frac{\omega}{2\pi}$ (ω is the angular velocity, reflecting how quickly the angle will change by 1 radian; the frequency ν shows how quickly the complete circle, i.e 2π radians will be traveled, so $\nu = \omega/2\pi$). The hodograph if a circle of radius ωr. The acceleration now is directed tangent to the hodograph, i.e., normal to $\boldsymbol{\tau}$ and equals ω times the "radius of the hodograph", i.e.

$$\boldsymbol{a} = \omega^2 r \boldsymbol{n} = \frac{v^2}{r}\boldsymbol{n},$$

where $\boldsymbol{n}$ is the principal normal to the circle. But

$$\boldsymbol{a} = \frac{d}{dt}(v\boldsymbol{\tau}) = v\frac{d\boldsymbol{\tau}}{dt} = v^2\frac{d\boldsymbol{\tau}}{ds},$$

therefore, eliminating $\boldsymbol{a}$ from the latter two formulae

$$\frac{d\boldsymbol{\tau}}{ds} = \frac{1}{r}\boldsymbol{n}.$$

Comparing with (2) we see that the radius of curvature of a circle equals the radius of the circle, in other words, the formula (2) generalises the notion of curvature from a circle to any three-dimensional curve.

Returning to the general case, it now makes sense to introduce another unit vector $\boldsymbol{b}$, which is normal to the osculating plane, and so that $(\boldsymbol{\tau}, \boldsymbol{n}, \boldsymbol{b})$ form a right triple, i.e their directions can be matched by the thumb, index, and middle fingers on the right (but not left!) hand. The vector $\boldsymbol{b}$ is called the *binormal* to the curve, and the triple $(\boldsymbol{\tau}, \boldsymbol{n}, \boldsymbol{b})$ of unit vectors is referred to as the Frenet (1816–1900) basis. The Frenet basis is attached to the curve and represents the "natural" moving frame of reference to be associated with a moving material point. The derivative $\frac{d\boldsymbol{b}}{ds}$ naturally shows how the direction of the osculating plane changes in time, i.e. how the curve twists, or to what extent it is not plane. Observing that

$$\boldsymbol{b} = \boldsymbol{\tau} \times \boldsymbol{n},$$

we have

$$\frac{d\boldsymbol{b}}{ds} = \frac{d}{ds}(\boldsymbol{\tau} \times \boldsymbol{n}) = \frac{d\boldsymbol{\tau}}{ds} \times \boldsymbol{n} + \boldsymbol{\tau} \times \frac{d\boldsymbol{n}}{ds}.$$

By (2) the first term is zero, because $\boldsymbol{a} \times \boldsymbol{a} = 0$ for any vector $\boldsymbol{a}$. Besides, since $\boldsymbol{b}$ is a unit vector, then $\frac{d\boldsymbol{b}}{ds}$ is perpendicular to $\boldsymbol{b}$, i.e. lies in the $(\boldsymbol{\tau}, \boldsymbol{n})$-plane (i.e. the osculating plane). But since now $\frac{d\boldsymbol{b}}{ds} = \boldsymbol{\tau} \times \frac{d\boldsymbol{n}}{ds}$, it must be perpendicular to $\boldsymbol{\tau}$, therefore we must have

$$\frac{d\boldsymbol{b}}{ds} = T\boldsymbol{n}, \tag{3}$$

for some real coefficient T. The latter coefficient T is called *torsion*, or *twist*, and can be both positive and negative. The two formulas (2) and (3) are referred to as Frenet formulae. If T is always zero, the curve is called plane. In this case, one can always choose coordinates, so that $z(t) = 0$, and the osculating plane is the xy-plane, where the curve lives. Torsion is then always zero, while curvature i smuch easier to compute than in the general case.

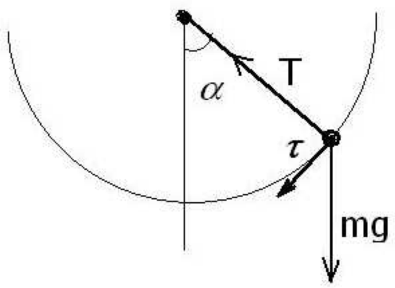
2.2.3 General formula for acceleration

Let us use (2) to obtain a general formula for acceleration. We have, using $ds = vdt$,

$$\boldsymbol{a} = \dot{\boldsymbol{v}} = \frac{d(v\boldsymbol{\tau})}{dt} = \frac{dv}{dt}\boldsymbol{\tau} + v\frac{d\boldsymbol{\tau}}{dt} = \frac{dv}{dt}\boldsymbol{\tau} + \frac{v^2}{r}\boldsymbol{n} \equiv \boldsymbol{a}_{\parallel} + \boldsymbol{a}_{\perp}. \tag{4}$$

Namely, acceleration always has two components – one $\boldsymbol{a}_{\parallel}$ in the direction of $\boldsymbol{\tau}$, which is called *linear* acceleration, accounting for the change of speed, and the other $\boldsymbol{a}_{\perp}$ in the normal direction $\boldsymbol{n}$, called *angular* acceleration, accounting for curvature of the trajectory.

Example – mathematical pendulum. Let us use this formula to write the equation of motion for the mathematical pendulum: a material point of mass m constrained to move on a circle of radius r under the influence of gravity by, say attaching it on a massless string of fixed length r, whose opposite end has been fixed. This example requires the Second law of Newton $\boldsymbol{F} = m\boldsymbol{a}$. Assuming it, let us use the formula (4) for the acceleration. The net force acting on the material point consists of gravity mg acting downwards and the constraint force $\boldsymbol{T}$ directed towards



the centre of the circle, i.e. along $\boldsymbol{n}$.

Let us look only at the tangent component of the acceleration. If α is the angle on the circle, with $\alpha = 0$ at the lowermost point, α increasing counterclockwise, then the component of the net force in the tangent direction equals $mg \sin \alpha$. Hence

$$m\frac{dv}{dt}\boldsymbol{\tau} = mg \sin \alpha \boldsymbol{\tau},$$

where $\boldsymbol{\tau}$ is tangent to the circle, pointing *clockwise*. Using now $v = -\frac{d\alpha}{dt}r$ (the minus sign is because α increases counterclockwise, while $\boldsymbol{\tau}$ points clockwise), we have

$$\ddot{\alpha} + \frac{g}{r}\sin \alpha = 0. \tag{5}$$

The general solution of this second order equation cannot be expressed via elementary functions, and requires the use of special Jacobi elliptic functions. However, assuming that α is small, i.e. the pendulum does *small oscillations*, one can approximate $\sin\alpha \approx \alpha$, in which case

$$\alpha(t) \approx A\cos(\sqrt{\frac{g}{r}}t + \phi),$$

where the constants A, ϕ are determined from initial conditions. This approximate solution is good enough only when $A \approx \sin A$, i.e. for small amplitudes A. If the amplitude increases, it will be later shown that the period of oscillations – which for small A does not depend on it and equals $2\pi\sqrt{\frac{r}{g}}$ – increases, and as A approaches π the period goes to infinity.

2.2.4 The number of degrees of freedom

Looking back at the above example, the position of the pendulum is naturally characterised by the angle α, rather than the two Cartesian coordinates x and y. Of course, knowing $\alpha(t)$ one can write down what $x(t)$ and $y(t)$ are. In general, if the evolution of a mechanical system can be fully described by the least number n of scalar functions $q_1(t), \ldots, q_n(t)$ of t, the number n is called the number of *degrees of freedom* of this system. The quantities $q_1(t), \ldots, q_n(t)$ are called *generalised coordinates.* E.g. a particle in three dimensions has three degrees of freedom, the generalised coordinates being x, y, z. A system of N particles has $3N$ degrees of freedom, the generalised coordinates being the union of the Cartesian coordinates $x_1, y_1, z_1, \ldots, x_N, y_N, z_N$ of each particle. A system may have constraints, i.e. functional relations between its Cartesian coordinates, and the number of degrees of freedom equals the number of Cartesian coordinates minus the number of constraints. E.g., for the pendulum, the Cartesian coordinates x, y are connected via a constraint $x^2 + y^2 = r^2$. Hence, the pendulum has one degree of freedom, and is best described in terms of the generalised coordinate α which is characteristic of the geometric shape of the trajectory – a circle.

Consider a *rigid body*, which can be viewed as a collection of particles, such that the distances between each pair of particles have been fixed once and for all. The number of degrees of freedom of a rigid body equals 6. Indeed, it suffices to choose three non-collinear points A, B, C inside the body, and knowing their movement in time will unambiguously determine the movement of any other particle in the body. There are three constraints on the points A, B, C: the distances AB, BC, AC are fixed and equal, day d_{AB}, d_{BC}, d_{AC}, respectively. So, the 9 Cartesian coordinates $x_A, y_A, z_A, x_B, y_B, z_B, x_C, y_C, z_C$ have three constraints imposed thereon:

$$\begin{array}{c}(x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2 = d_{AB}^2,\\ (x_C - x_B)^2 + (y_C - y_B)^2 + (z_C - z_B)^2 = d_{CB}^2,\\ (x_A - x_C)^2 + (y_A - y_C)^2 + (z_A - z_C)^2 = d_{AC}^2.\end{array}$$

Hence, one can expect to be able to decompose any free motion of a rigid body into the body moving as the whole (i.e. describing the motion of the *mass centre* in terms of its three coordinates x_C, y_C, z_C) and rotating with respect to some axis, passing through the mass centre, the rotation being characterised by a three-dimensional vector $\boldsymbol{\omega}$.

2.2.5 Appendix: some formulae for curvature

The definition of curvature by (2), in other words as

$$K = \left|\frac{d\boldsymbol{\tau}}{ds}\right|,$$

despite being the only intrinsic one, is not very practical if one wants to calculate it in some simple instances. The difficulty is that almost never one has the parameterisation of a curve in terms of the natural parameter s. Usually, the curve is parameterised as $\boldsymbol{r}(t)$. In this case, with the notations $\boldsymbol{v} = \dot{\boldsymbol{r}}$, $\boldsymbol{a} = \ddot{\boldsymbol{r}}$, and v for the absolute value of $\boldsymbol{v}$ it is possible to write a general formula for K as well. Indeed, as $ds = vdt$ we have

$$K = \frac{1}{v}\left|\frac{d\boldsymbol{\tau}}{dt}\right| = \frac{1}{v}\left|\frac{d}{dt}\frac{\boldsymbol{v}}{v}\right| = \frac{1}{v^3}\left|v\boldsymbol{a} - \frac{dv}{dt}\boldsymbol{v}\right|.$$

By the rule of differentiating dot product and the chain rule

$$\frac{dv}{dt} = \frac{d}{dt}\sqrt{\boldsymbol{v}\cdot\boldsymbol{v}} = \frac{\boldsymbol{v}\cdot\dot{\boldsymbol{v}}}{\sqrt{\boldsymbol{v}\cdot\boldsymbol{v}}} = \frac{\boldsymbol{v}\cdot\boldsymbol{a}}{v}.$$

Hence

$$K = \frac{1}{v^3}\left|v\boldsymbol{a} - \frac{(\boldsymbol{v}\cdot\boldsymbol{a})\boldsymbol{v}}{v}\right|. \tag{6}$$

Let us take a particular case when the curve is plain, so $y = f(x)$, $z = 0$. In this case one can identify x and t, writing the curve's equation as

$$\boldsymbol{r}(t) = t\boldsymbol{i} + f(t)\boldsymbol{j}, \quad \text{so} \quad \boldsymbol{v}(t) = \boldsymbol{i} + f'(t)\boldsymbol{j}, \quad \boldsymbol{a}(t) = f''(t)\boldsymbol{j}.$$

Substitution into (6), using $v = \sqrt{1+[f'(t)]^2}$ yields, after some algebra

$$K = \frac{1}{v^4}|-f'(t)f''(t)\boldsymbol{i} + f''(t)\boldsymbol{j}| = \frac{f''(t)}{v^4}\sqrt{1+[f'(t)]^2} = \frac{f''(x)}{(1+[f'(x)]^2)^{3/2}},$$

recalling $x = t$ – a formula known from calculus.

3 Dynamics

Dynamics is the part of mechanics dealing with motions of bodies under the action of forces. The force is in effect a measure of how the body interacts with other bodies. Laws of dynamics were discovered by Newton; they are empirical and cannot be proved.

3.1 Newton's laws

Newton starts his *Philosophiae Naturalis Principia Mathematica* (Mathematical Principles of Natural Philosophy, or *The Principia,* first published in 1687) with the following definitions, among others.

i. Quantity of matter is a measure of matter that arises from its density and volume jointly.

ii. Quantity of motion is a measure of motion that arises from the velocity and the quantity of matter jointly.

iii. Inherent force of matter is the power of resting by which every body, so far as it is able, perseveres at its state either of resting or moving uniformly straight forward.

iv. Impressed force is the action exerted on a body to change its state either of resting or of moving uniformly straight forward.

He then formulates his famous laws.

I. **Law 1** Every body preserves its state of being at rest or moving uniformly straight forward, except insofar as it is compelled to change its state by forces impressed.

II. **Law 2** A change in motion is proportional to the motive force impressed and takes place along the straight line in which that force is impressed.

III. **Law 3** To any action there is always an opposite and equal reaction; in other words, the actions of two bodies upon each other are always equal and always opposite in direction.

Note that Newton's concept of mass is somewhat obscure, because the definition of density has not been given. Besides, the "continuous" idea of density is macroscopic and does not apply to the world of elementary particles. Therefore, Newton's definition of mass has not quite lived up to these days.

Aristotle and his followers considered force as the cause of motion, and that force was necessary to sustain motion. Newton's first law elucidated the fact that such an idea of force was false. Instead, Newton essentially defines force as a measure of intensity of bodies' interaction which displays itself via the changes of their quantity of motion, or *momentum.*

3.2 First Law, inertial frames and Relativity principle

The First Newton's law, or the law of inertia was, in fact, discovered by Galileo. Galileo was probably the first to bring into natural sciences the abstract mathematical idea of empty space-time where bodies that do not interact with other bodies move freely. Or, at least he was the first to use this idea as a the starting point for building physical theory. A free material body is moving uniformly and rectilinearly, *by inertia.* Free bodies, or bodies that do not interact with other bodies are, if course, an idealisation and the practical question regarding the First law is whether this idealisation can be sustained with high precision, i.e. whether "almost" free bodies exist. Modern physics understands interaction, even if it is effected via omnipresent in classical mechanics strings, springs, etc., in terms of *force fields* created by other bodies. There are four types of interaction: electromagnetic, gravitational, strong and weak. The latter two act on length scales under 10^{-12}cm and hence are beyond classical mechanics' concern. The former two are long-range interactions: static electric and gravitational forces vanish slowly, as the inverse square of the distance between bodies; time-changing fields carried by electromagnetic and (presumably) gravitational waves vanish even slower, as one over the distance. However, the absence of electromagnetic fields can be easily verified by their different action on positive and negative charges. This is not the case with gravitational fields. However, the static gravitational field from distant objects of the Universe can be regarded as uniform, and one can always introduce a frame of reference, free falling in this field, similar to how the Earth's gravity is absent in the Mir station.

A slight conceptual generalisation of the notion of a free body it the notion of a closed, or isolated, system. This is a collection of bodies which interact only with each other, but not with the "outside world". I.e. together they represent a more complex free body, possessing a higher number of degrees of freedom, so that the corresponding generalised coordinates would describe the body's inner state.

Assuming that free bodies and isolated systems exist, the First law states that there exist frames of reference, where the motion of a free body will be uniform and rectilinear. Such a frame is called *inertial.* The content of the First law is – there exists at least one such frame. This is by now is an accumulation of a vast number of experimental facts. It is only via an experiment, or by comparison with other inertial frames that one can establish whether a particular frame of reference is inertial – by verifying that free bodies move uniformly along straight lines. Practically, inertiality of a particular frame is approximate. The snooker table can be regarded as an inertial frame if one uses it to describe motions of the balls on the table. On the other hand, distant stars, which can be considered as free bodies because of the vastness of cosmic distances, perform daily periodic movements with respect to someone watching them from the snooker table, so for studying stars a snooker table is not an inertial frame of reference, due to the rotation of the Earth. On the other hand, the Copernicus frame of reference, whose origin is located roughly at the centre of the Sun, or more precisely at the mass centre of the Solar system is "inertial enough" to study the stars (If one wants to get "more inertial", the origin can be chosen at the centre of the Galaxy.) Indeed, the Earth is moving curvilinearly with respect to the Copernicus or any star-associated system, hence with acceleration, and this explains why is not an inertial frame.

Most importantly, if one system K is inertial, then any system K' moving with respect to K uniformly and along a straight line is also inertial. This appears to be self-evident from the classical mechanics point of view, because it is tantamount to claiming that if a free body is moving uniformly along a straight line with respect to one inertial frame, it will also be moving and along a straight line with respect to another frame, which is itself moving uniformly and along the straight line with respect to the former frame. Indeed, suppose, a a free body is moving in an inertial frame K, where one must have the body's trajectory as $\boldsymbol{r}(t) = \boldsymbol{r}_0 + \boldsymbol{v}t$, with some constant $\boldsymbol{v}$. Let $\boldsymbol{r}'(t')$ be the trajectory of the same body, viewed in the system K'. In classical mechanics time is absolute: $t' = t$, and therefore

$$\boldsymbol{r}'(t) = \boldsymbol{r}(t) - \boldsymbol{V}t, \quad \text{so} \quad \boldsymbol{v}' = \dot{\boldsymbol{r}}'(t) = \boldsymbol{v} - \boldsymbol{V}, \quad \text{as} \quad t = t'. \tag{7}$$

This and the next formula are referred to as the*Galileo transformations*, or the *classical law of velocity addition.* However, is based on a *principle* which is only approximately correct if the velocities $\boldsymbol{v}, \boldsymbol{V}$ are small compared to c, namely that time is absolute: $t = t'$.

One can always assume that $\boldsymbol{V}$ is directed along the x-axis, then the coordinate transformations between the frames K and K' can be written simply as

$$x' = x - Vt, \quad y' = y, \quad z' = z, \quad t' = t. \tag{8}$$

Galileo went conceptually much further. He proposed what is now called the *relativity principle*, which postulates that there are no observational consequences of absolute motion: all laws of physics (including Newton's laws) must have the same form, regardless of whether frame of reference is used to express them.[6]

The idea behind Galileo's principle is as follows. Consider tho isolated systems – physics laboratories that move with respect to each other with constant speed and rectilinearly. Associate with them the two inertial frames of reference K and K' above. Then all laws of physics describing physical processes in both labs are exactly the same. The same is, for instance, the second Newton's law $\boldsymbol{F} = m\boldsymbol{a}$. Indeed, in the context of the two laboratories, suppose the force comes form some physical interaction, which depends on coordinates, relative velocities of interacting bodies, and, perhaps, time. If the same interaction occurs in both labs, then $\boldsymbol{F} = \boldsymbol{F}'$. But by (7) we have

$$\boldsymbol{a}' = \frac{d\boldsymbol{v}'}{dt'} = \frac{d\boldsymbol{v}'}{dt} = \frac{d}{dt}(\boldsymbol{v}(t) - \boldsymbol{V}) = \frac{d\boldsymbol{v}}{dt} - 0 = \boldsymbol{a}.$$

Besides, $m = m'$, because mass can be determined via some force interaction, say weighing. So, the second Newton's law $\boldsymbol{F} = m\boldsymbol{a}$ in K looks exactly the same: $\boldsymbol{F}' = m\boldsymbol{a}'$ in K'.

In more modern terms, Galileo's relativity principle postulates homogeneity of space and time. This means that physical properties of space-time do not depend on where and when physical phenomena take place: any physics experiment repeated in the two labs with the associated inertial frames K and K' under the same conditions shall yield exactly the same results no matter how far the two instances of the experiment are separated in space or time. I.e., there are no "privileged" points in space or time. (Of course, this applies to the abstract "empty" from interaction space-time only.) In the same vein, space is *isotropic*, i.e., all the directions in space are equivalent. E.g., in the absence of other force fields but gravity, a cannon that makes the same angle with the horizon shoots the same projectile at the same distance, no matter in which direction.

It follows from homogeneity of space-time and isotropy of space that in general, the transformation formulae between the inertial frames K and K' such as (8) must be expressed by linear functions of (x, y, z, t), whose coefficients depend only on the relative speed V between the two frames. Indeed, if the coefficients were allowed to depend on (x, y, z, t), the rule would depend on where K' was located relative to K, and this would contradict homogeneity. Similarly, their dependence on the direction of $\boldsymbol{V}$ would contradict the isotropy of space.

By the end of the XIX century physicists were confronted by a major problem. Maxwell's equations for electromagnetic field in vacuum were not invariant with respect to the Galileo transformations: change the variables in these equations according to (8) and the equations will look different. Instead of (8), Maxwell's equations were invariant with respect to the following changes:

$$x' = \gamma(x - Vt), \;\; y' = y, \;\; z' = z, \;\; t' = \gamma\left(t - \frac{V}{c^2}x\right), \qquad \text{where } \;\gamma = \frac{1}{\sqrt{1 - \frac{V^2}{c^2}}}. \tag{9}$$

These were called Lorentz transformations, and in the limit $c \to \infty$ they yield the Galileo transformations (8). This contradiction was thought to be accounted to non-inertiality of conventional Earth-associates frames of reference: the primordial inertial frame was to be associated with *luminiferious ether* – a hypothetical medium which carries electromagnetic waves similar to as the air carries sound. The idea, or the principle, however, could not sustain: the famous experiment by Michelson-Morley (Google it!) performed in 1887 proved it false.

Einstein's special relativity theory that resolved the contradiction was the triumph of the method of principles. Einstein postulated the relativity principle above all. If a frame of reference K' is moving uniformly and along the straight line relative to the inertial frame K, it is also inertial. (One can then assume that the direction of the coordinate axes in K and K' coincide and the origin O' of K' is moving along the x-axis with velocity V.) This means that any physical experiment repeated independently in the two labs associated K and K' would yield the same result. Whether this experiment consists in looking at butterflies flying or fish swimming, measuring the charge of an electron or Newton's gravitational constant, or ... the speed of light in vacuum c. To resolve the above contradiction with Maxwell's equations, Einstein had to assume that c is also a world's fundamental constant giving the *maximum speed of interaction.* This implies that $t = t'$ is untenable, because as it was discussed above

[6] Galileo wrote (the translation is not quite exact): "Seclude yourself with some friend of yours in a large room under a ship's desk. Bring with you all sorts of flies, butterflies, and other small flying insects. Fill a tank with water and set some fish free therein. Hang a bucket high up to the ceiling, so that the water from the bucket falls drop by drop into a small bottle with a narrow neck. Let the ship be standing still...". He then goes on saying that if the ship were, in fact, moving in calm seas, you and your friend under the deck would have no idea about it, just by looking at the insects flying, fish swimming, and water dripping vertically into the bottle. He also notes that if one jumps or throws a stone at a certain distance on the moving ship, it takes precisely the same effort as it does on the ground.

the notion of simultaneity becomes relative. Assuming the maximality of c, Lorentz transformations follow from the relativity principle (to be shown later), regardless of Maxwell's equations.

But the First Newton's law was not affected: trajectories of free bodies in inertial frames are straight lines, along which the free bodies are traveling with constant speeds. (Only straight lines can be generalised as paths chosen by light rays.) And given an inertial frame K, any other frame K' moving uniformly along the straight line with respect to K (e.g. K' can be associated with a moving free body) is inertial.

3.3 The Second and Third Laws, mass and momentum

According to Newton's framework, the Second law requires the notion of the quantity of motion, or *momentum* $\boldsymbol{p}$, which in turn requires the notion of mass. The formulation of the second law is then

$$\dot{\boldsymbol{p}} = \boldsymbol{F}, \tag{10}$$

and for a single particle in classical mechanics

$$\boldsymbol{p} = m\boldsymbol{v}, \quad \text{so} \quad \boldsymbol{F} = m\boldsymbol{a}.$$

But (10) is more general: in special relativity, a momentum of a particle moving with velocity $\boldsymbol{v}$ is

$$\boldsymbol{p} = \frac{m_0 \boldsymbol{v}}{\sqrt{1 - \frac{v^2}{c^2}}}, \tag{11}$$

where m_0 is *rest mass.* Acceleration is still $\dot{\boldsymbol{v}}$, but taking the time derivative $\dot{\boldsymbol{p}}$ is now more tricky, and it still defines force.

Let us discuss the notion of mass. Mass in mechanics is a measure of inertiality. Suppose, two material points have interacted, whereupon their velocities $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ have changed to $\boldsymbol{v}_1'$, $\boldsymbol{v}_2'$, by $\Delta\boldsymbol{v}_1$, $\Delta\boldsymbol{v}_2$, respectively. First off, if space is homogeneous and isotropic, the changes $\Delta\boldsymbol{v}_1$ and $\Delta\boldsymbol{v}_2$ must have opposite directions. Then one can write

$$m_1 \Delta\boldsymbol{v}_1 = -m_2 \Delta\boldsymbol{v}_2, \tag{12}$$

with some real coefficients m_1 and m_2. These are called the *inertial* masses of the material points. It follows that a body's mass is determined as a measure of its inertiality with respect to other bodies, therefore one should postulate that some body has a unit mass, which is a matter of choice of the system of units.

Amazingly, Newton discovered that the law of gravitational attraction is such that the force of gravitational attraction between two bodies is proportional to each's mass: if the force $\boldsymbol{F}_{21}$ is impressed by the first body onto the second one, and $\boldsymbol{r}_{12}$ is the radius vector from body 1 to body 2, then

$$\boldsymbol{F}_{21} = -G \frac{m_1 m_2}{r_{12}^2} \frac{\boldsymbol{r}_{12}}{r_{12}},$$

where the unit vector $-\frac{\boldsymbol{r}_{12}}{r_{12}}$ is there to indicate that this is an attracting force, and $G \approx 6.67300 \times 10^{-11} m^3 kg^{-1} s^{-2}$ is the Gravitational constant. This formula tells one nothing about the origin of gravity – Newton ascribed the formula to God. Einstein might have credited gravity to God as well, but his General relativity did explain the formula above, and why, in fact, the gravitational mass is so amazingly related to the inertial mass.

It is worth mentioning here that in relativistic mechanics the inertial mass is no longer a constant: instead of (12) one would have

$$m_{0,1} \left(\frac{\boldsymbol{v}_1'}{\sqrt{1 - \frac{(v_1')^2}{c^2}}} - \frac{\boldsymbol{v}_1}{\sqrt{1 - \frac{v_1^2}{c^2}}} \right) = -m_{0,2} \left(\frac{\boldsymbol{v}_2'}{\sqrt{1 - \frac{(v_2')^2}{c^2}}} - \frac{\boldsymbol{v}_2}{\sqrt{1 - \frac{v_2^2}{c^2}}} \right), \tag{13}$$

in other words the inertial relativistic mass depends on speed:

$$m = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}}, \tag{14}$$

where m_0, the rest mass, is the mass of a particle determined by an observer, with respect to which the particle does not move, e.g. if the observer is associated with the particle itself.

The relation (12) has curious consequences. Suppose, the interaction is taking place continuously, and over a time dt the velocities change by $d\boldsymbol{v}_1$, $d\boldsymbol{v}_2$. Then we have

$$m_1 d\boldsymbol{v}_1 = -m_2 d\boldsymbol{v}_2, \text{ so } m_1\dot{\boldsymbol{v}}_1 = -m_2\dot{\boldsymbol{v}}_2, \text{ so } \boldsymbol{P} = m_1\boldsymbol{v}_1 + m_2\boldsymbol{v}_2 = \boldsymbol{p}_1 + \boldsymbol{p}_2 = const. \tag{15}$$

So, viewed as the isolated system, the system of two particles has constant total momentum. This is a particular case of the law of the *momentum conservation.* Note that by (13) the law of conservation of momentum $\boldsymbol{P} = const.$ transcends to relativistic mechanics as well, and so represents a universal principle.

Defining, via the second Newton's law the forces $\boldsymbol{F}_{12} = m_1\boldsymbol{a}_1$ and $\boldsymbol{F}_{21} = m_2\boldsymbol{a}_2$ impressed by the second particle onto the first one and conversely, we have

$$\boldsymbol{F}_{12} = -\boldsymbol{F}_{21}, \tag{16}$$

which is a particular form of the Third law: action is equal in value and opposite in direction to reaction.

However, Newton's formulation of the third law appears to be more general: it states that (16) is valid for any system of particles, as long as the forces $\boldsymbol{F}_{12}$ and $\boldsymbol{F}_{21}$ can be correctly identified. Strictly speaking, however, it represents a conceptual difficulty. The Second law only defines a net force $\boldsymbol{F}$ acting on a particle, as the rate of change of a particle's momentum. In a practical mechanical model, involving different kinds of interaction, one tries to represent the net force as a superposition

$$\boldsymbol{F} = \boldsymbol{F}_1 + \boldsymbol{F}_2 + \ldots$$

of various forces. E.g. in the pendulum example considered above the net force was the superposition of the force of gravity and the tension force applied by the string. The fact that the two could be added as vectors is based on the physical assumption of *independence* of the two forces: the gravity mg is there regardless of the string; on the other hand, tension of the string (no wonder, one is much better off speaking of a massless string) can be realised in the absence of gravity by, say attaching to its both ends two charges of the same sign, which would repel each other.

So, the moral is that the Laws of Newton should be viewed as a system of principles, and one should not separate them from one another.

3.3.1 Hooke's Law and some other forces

Consider a simplest mechanical system: a body of mass m hanging vertically on a spring. One might try to hang different bodies and find out that, in fact, if their masses are not too great, the spring after the body has been hanged will become longer by $\Delta x = mg/k$, for some constant k, characterising the spring independent of the body that's being hanged. In terms of Newton's second law this means that a deformed spring impresses a force which is directed opposite to the deformation x and equals $-kx$. This is called Hooke' (1635–1703) law. The nature of the law is electromagnetic: the spring force is due to the change in the alignment of positive and negative charges inside the spring, where they were originally in equilibrium. The law is approximate and gets violated for large x. Wherever the spring force comes from, if the x axis is directed downwards, the equation of motion of the body hanging on the spring is

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

The solution $x(t)$ of this equation is the sum of the solution of the homogeneous equation

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

with the notation $\omega^2 = \frac{k}{m}$, and the particular solution. The latter is simply a constant $x = x_0 = mg/k$, and represents the equilibrium position of the body on spring. Hence $x(t)$ will represent oscillations along the equilibrium position, which can be taken for the origin:

$$x - x_0 = A\cos(\omega t + \phi),$$

where A is the amplitude and together ϕ is found from the initial conditions $x(0)$ and $\dot{x}(0)$. The period of oscillations is $T = \frac{2\pi}{\omega}$.

If besides one takes into account the fact that surrounding the air resists the motion by imposing a *drag force* which is proportional to the speed of the body and is directed opposite to it, there is an extra force $-\kappa\dot{x}$, with some *drag coefficient* κ. (This fact can be established empirically; more deeply it is a consequence of equations of gas/fluid dynamics and is once again approximate.)

Changing x to $x - x_0$ does not affect $\dot{x}$ or its higher derivatives. Having done this, mg in the equation of motion is gone, and the equation of motion, considering the drag force, is now

$$\ddot{x} + 2\gamma\dot{x} + \omega^2 x = 0,$$

where $\gamma = \frac{1}{2}\kappa/m$. (The coefficient 2 is a matter of convenience.) The characteristic polynomial has roots $-\gamma \pm i\sqrt{\omega^2 - \gamma^2}$, which are complex, provided that $\gamma < \omega$, or the system is not *overdamped.* The solution is now

$$x - x_0 = Ae^{-\gamma t}\cos(\sqrt{\omega^2 - \gamma^2}\, t + \phi),$$

representing damped oscillations. Their period is now $T = \frac{2\pi}{\sqrt{\omega^2-\gamma^2}}$, getting longer with the growth of γ. The amplitude is vanishing exponentially. Namely, the ratio of the values $x - x_0$ at the times t_1 and t_2, giving the two the successive maxima (t_1 is such that $t_1\sqrt{\omega^2 - \gamma^2} + \phi = 2\pi k$, t_2 is such that $t_2\sqrt{\omega^2 - \gamma^2} + \phi = 2\pi(k+1)$, for some integer k) equals $e^{\gamma T}$. The exponent $\gamma T = 2\pi\frac{\gamma}{\sqrt{\omega^2-\gamma^2}}$ is called the *logarithmic decrement*, showing how many times the log of the amplitude decreases over a single period of oscillations.

In general, there are two main types of dynamics problem. One is to calculate forces, given the accelerations within a system of material bodies. This is fairly simple, as the relation $\boldsymbol{F} = m\boldsymbol{a}$ is a linear equation. The other type of problems is to determine accelerations within a particular mechanical system. This is more difficult, and may be regarded as the central problem of mechanics. Consideration of such a problem should start out with the identification of independent forces acting on each body in the system. Such forces can be gravity, deformation forces, drag forces, etc. In addition, a system may be constrained, like the pendulum, to move not in the whole space but along some regions of space only. Constraints impose reaction forces (a constrained body “feels” the constraints), and the problem of identifying the acceleration becomes inseparable from calculating the reaction forces.

In addition, there are friction forces. Friction has electromagnetic nature, but mechanical friction laws have been dealt with long before it became understood. Consider a massive body on a surface, and try to move it. This requires some fitness level. If, say the body is a massive rectangular block on a horizontal surface, there is no way to shift it horizontally for anyone who is not capable of exerting a force which is some fraction of the block’s weight. This is very different form drag forces: a single man can (in principle) pull a massive barge through water, provided that he does not have an ambition of moving it too rapidly. In other words, drag forces are proportional to speed and are therefore small if the speed is small. On the other hand, friction is roughly proportional to the reaction of the surface, whereupon the body is positioned. If the surface is horizontal and the body is not moving, then the reaction $N = Mg$, and in order to shift the body one needs to exert a horizontal force in excess of μMg, for some $\mu > 0$. The coefficient μ can drop after the body has been set in motion. Thus, one has to distinguish between *static* and *dynamical* friction. The easiest model that one assumes is as follows. By the *friction coefficient* one always means the dynamic friction coefficient, which is determined as the fraction of the weight of the body, equal to the value of a horizontal force it takes to sustain the body in uniform motion along a horizontal surface. If a body is on any surface, one defines the *dynamic friction force* equal in value to μN, where N is the value of the reaction force that the body experiences from the surface. The dynamic friction force is directed opposite to the body’s velocity along the surface.

It the body is at rest with respect to the surface, the force $\boldsymbol{F}_{\|}$ acting along the surface may not suffice to move the body. Then, by the second Newton’s law one has $\boldsymbol{F}_{sf} + \boldsymbol{F}_{\|} = 0$, where $\boldsymbol{F}_{sf}$ denotes the *static friction force.* One can also write $F_{sf} = \mu_s N$, where μ_s is the static friction coefficient. Assuming that bodies do not get stuck to surfaces, one always has

$$\mu_s \leq \mu,$$

where μ is the dynamic friction coefficient. Laws of friction were studied in detail by Coulomb (1736–1806).

E.g. consider a car moving with constant speed v along a curve of radius R. What is the maximum speed $v_{\max}$ that the car may have before it starts skidding if the friction coefficient between the tyres’ rubber and pavement is μ? The car is moving along the circle, so is not moving in the direction normal to the circle. On the other hand, by (4) its acceleration is directed towards the centre of the curve. The only force that acts on the car in the direction towards the centre is the static friction force, so one has

$$mv^2/R = \mu_s mg$$

The motion will sustain as long as $v = \sqrt{\mu_s g R}$, and the maximum static friction is equal to dynamic friction, i.e. when $\mu_s = \mu$. Hence the maximum speed $v_{\max} = \sqrt{\mu g R}$.

3.3.2 Non-inertial frames: centrifugal forces

In the above example, a passenger in the car feels the "centrifugal force" than pushes him outside the centre of the circle. The passenger is moving with acceleration and therefore does not represent an inertial frame of reference. It is often useful, however to write the second Newton's law in such a non-inertial frame of reference. The acceleration $\boldsymbol{a}'$ of the body i the frame associated with itself is clearly zero.

By the Galileo transforms (7) it is no longer true that $\boldsymbol{a}' = \boldsymbol{a}$, because $\boldsymbol{V}$ is not a constant. What is true, however is that in general $\boldsymbol{a}' = \boldsymbol{a} - \dot{\boldsymbol{V}}$. As $\boldsymbol{V}$ is the velocity of one frame with respect to the other, $\boldsymbol{V} = \boldsymbol{v}$, the body's velocity. In other words, in the non-inertial frame associated with the moving body one can introduce an artificial *correspondence force*, which equals $-m\boldsymbol{a}$, where $\boldsymbol{a}$ is the body's acceleration in the inertial frame. The main equation in the previous example, with respect the non-inertial frame associated with the car can be written as

$$\boldsymbol{F}_c + \boldsymbol{F}_{sf} = 0,$$

where $\boldsymbol{F}_c$ is the correspondence centrifugal force that the body feels acting on it off the centre and equal mv^2/R.

Sometimes this idea does bring a certain simplification. Consider the following example. A person of mass M is sitting on a chair of mass m, and the coefficient of friction between their trousers' material and the chair is μ. The chair can be moved without friction along the floor (say, the floor is ice). What horizontal force F would suffice to pull the chair from under the person? First, suppose there has been no success, as F is small. It has, however, set the whole system in motion, with acceleration $a = F/(M+m)$. In the non-inertial frame associated with the person, the acceleration $\boldsymbol{a}' = 0$: this is due to the static friction force, which compensates the "correspondence force" $Ma = F\frac{M}{m+M}$. Hence the person will remain on the chair as long as

$$\mu_s M g = F\frac{M}{m+M}$$

But as one has $\mu_s \le \mu$, remaining on the chair will become impossible if $F \ge \mu(M+m)g$.

3.4 Conservation of momentum, mass centre

The second Newton's law $\boldsymbol{F} = m\boldsymbol{a}$ clearly implies that if $\boldsymbol{F} = 0$, $\boldsymbol{a} = 0$, so the velocity $\boldsymbol{v} = const.$ or the momentum $\boldsymbol{p} = const.$ (This does not imply the First law, because the Second law is true only in inertial frames, and the First law postulates that they exist.) In addition, if the projection of a force $\boldsymbol{F}$ on a specific direction is zero, then the projection of the momentum on that direction is constant. (This is used, for instance, in problems about projectiles, where the only force, gravity, is vertical, so the horizontal velocity is preserved.)

Let us generalize the law of conservation of momentum to a system of N particles. Suppose, the ith particle is acted upon by forces $\boldsymbol{F}_{ij}$ via the interactions with other particles $j \neq i$, and all interactions are independent. In addition, suppose, the system is in some external force field $\boldsymbol{F}_e$, which acts on each particle individually via the force $\boldsymbol{F}_{ii}$.

By the Second law, for the ith particle of mass m_i, with the radius-vector $\boldsymbol{r}_i$ and momentum $\boldsymbol{p}_i$:

$$\dot{\boldsymbol{p}}_i = \sum_{j=1}^{N} \boldsymbol{F}_{ij}, \qquad i = 1, \ldots, N.$$

Let us sum these equations over all i (let us not write the upper limits in the sums, as they are irrelevant):

$$\sum_i \dot{\boldsymbol{p}}_i = \sum_{i,j} \boldsymbol{F}_{ij} = \sum_i \boldsymbol{F}_{ii} + \sum_{i \neq j} \boldsymbol{F}_{ij}.$$

Namely, in the double sum $\sum_{i,j} \boldsymbol{F}_{ij}$ we have separated the "diagonal term" with $i = j$. By the Third law, for $i \neq j$, we have $\boldsymbol{F}_{ij} = -\boldsymbol{F}_{ji}$, and therefore the second sum above is zero. Thus

$$\dot{\boldsymbol{P}} = \frac{d}{dt} \sum_i \boldsymbol{p}_i = \sum_i \boldsymbol{F}_{ii} = \boldsymbol{F}_e. \tag{17}$$

In particular, if $\boldsymbol{F}_e = 0$, then $\boldsymbol{P} = const$. This is the general formulation of the law of conservation of momentum: the total momentum of an isolated system is preserved. Or is preserved its projection on some direction, such that the projection of the net external force on this direction is zero.

Observe that in classical mechanics

$$\boldsymbol{P} = \sum_i m_i \boldsymbol{v}_i = M\dot{\boldsymbol{R}}_{mc},$$

where $M = \sum_i m_i$ is the total mass, and

$$\boldsymbol{R}_{mc} = \frac{\sum_i m_i \boldsymbol{r}_i}{M} \tag{18}$$

is called the position of the system's *mass centre*. From this point of view, one can say that the mass centre of an isolated system is moving with constant velocity

$$\boldsymbol{V}_{mc} = \dot{\boldsymbol{R}}_{mc} = \frac{\sum_i m_i \boldsymbol{v}_i}{\sum_i m_i}. \tag{19}$$

If a system is impressed upon by external forces, then by (17) its mass centre moves under the action of the net external force as a single particle of mass M. This statement is sometimes called the *theorem of motion of the mass centre*:

$$M\frac{d\boldsymbol{V}}{dt} = \boldsymbol{F}_e,$$

where $\boldsymbol{V}$ is the velocity of the mass centre, and $\boldsymbol{F}_e$ the external force.

Besides, the mass centre is by its definition an additive quantity: the mass centre of two systems with total masses M_1 and M_2, and mass centres at $\boldsymbol{R}_{mc1}$, $\boldsymbol{R}_{mc2}$ coincides with the mass centre of two particles of masses M_1 and M_2 positioned at $\boldsymbol{R}_{mc1}$ and $\boldsymbol{R}_{mc2}$. It suffices to show it for 4 particles; the case when N particles are partitioned into n groups is identical. We have

$$\boldsymbol{R} = \frac{m_1\boldsymbol{r}_1 + m_2\boldsymbol{r}_2 + m_3\boldsymbol{r}_3 + m_4\boldsymbol{r}_4}{m_1 + m_2 + m_3 + m_4} = \frac{(m_1 + m_2)\frac{m_1\boldsymbol{r}_1 + m_2\boldsymbol{r}_2}{m_1+m_2} + (m_3 + m_4)\frac{m_3\boldsymbol{r}_3 + m_4\boldsymbol{r}_4}{m_3+m_4}}{(m_1 + m_2) + (m_3 + m_4)} = \frac{M_1\boldsymbol{R}_1 + M_2\boldsymbol{R}_2}{M_1 + M_2},$$

where $M_1, \boldsymbol{R}_1$ characterise the mass centre of the first and second particles viewed as a single group, while $M_2, \boldsymbol{R}_2$ characterise the mass centre of the third and fourth particles viewed as the other group.

It is convenient to associate with any closed system an inertial *mass centre frame*, whose origin is positioned at $\boldsymbol{R}_{mc}$, moving with the constant velocity $\boldsymbol{V}_{mc}$. Indeed, in the absence of external forces the mass centre is moving with constant velocity, and then the mass center frame is inertial, where the mass centre, i.e. the system "as whole" rests.

As an application, consider a boat of length L and mass M which is uniformly distributed along the length. The boat is floating in water. A person of mass m walks slowly from one end of the boat to the other, whereupon the boat moves by a distance X. To find X, one can regard the boat and the person as an isolated system – as the person walks slowly, the boat will move slowly, and the drag force impelled by the water can be disregarded. The mass centre of the system therefore should not have moved. The mass centre of the boat is in the middle of the boat. Taking it as the origin, by (18), the mass centre of the system when the person is on he left/right end of the boat, is at the distance $\frac{m}{M+m}\frac{L}{2}$ to the left/right of the middle of the boat. As the system's mass centre shall not have moved, throughout the person walking from one end to the other the boat must move in the opposite direction, eventually by the distance $\frac{m}{M+m}L$.

Note than, in fact the assumption of uniformity of mass distribution throughout the boat was superfluous: the above argument goes through wherever the boat's mass centre is.

As another particular case consider the so-called *law of conservation of mass.* Mechanics deals with interactions between bodies, such that the total mass of the bodies involved is constant. This is not true, say in nuclear fusion, when part Δm of the mass of light nuclei that collide to form heavier nuclei gets transformed into (huge, equal to Δmc^2) energy that gets released. The law of conservation of mass is, in fact, a particular case of the law of conservation of energy. While the former is valid only within the limits of classical mechanics, the latter is one of the most fundamental principles of all physics.

Consider two particles of masses m_1 and m_2 that interact and collide, whereupon they form a single particle of mass m. The self-evident fact that $m = m_1 + m_2$ can in fact be derived from the law of conservation of momentum and Galileo's principle. Let $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ be the velocities of the particles before the collision and $\boldsymbol{v}$ be the velocity of the composite particle after the collision.

By the law of conservation of momentum, as there are no external forces involved,

$$m_1\boldsymbol{v}_1 + m_2\boldsymbol{v}_2 = m\boldsymbol{v}. \tag{20}$$

Let some other inertial frame K' move with the velocity $\boldsymbol{V}$ with respect to the origin.

According to Galileo transforms, the velocities of the particles in the inertial frame K' are

$$\boldsymbol{v}_1' = \boldsymbol{v}_1 - \boldsymbol{V}, \qquad \boldsymbol{v}_2' = \boldsymbol{v}_2 - \boldsymbol{V}, \qquad \boldsymbol{v}' = \boldsymbol{v} - \boldsymbol{V}.$$

The law of conservation of momentum in K' becomes

$$m_1(\boldsymbol{v}_1 - \boldsymbol{V}) + m_2(\boldsymbol{v}_2 - \boldsymbol{V}) = m(\boldsymbol{v} - \boldsymbol{V}).$$

Comparing with (20) yields

$$(m_1 + m_2)\boldsymbol{V} = m\boldsymbol{V},$$

and as this is the case for any $\boldsymbol{V}$, one must have $m = m_1 + m_2$.

Of course, this is a particular case of the above discussion about the mass centre. Suppose, N particles within an isolated system over some time get glued into one, due to internal interactions only. The mass centre momentum $M\boldsymbol{V}_{cm}$ does not change throughout the interaction, yet after all the particles have become one, $\boldsymbol{V}$ is the velocity of the composite particle. Hence, $M = \sum m_i$ is its mass.

3.4.1 Two-body problem and effective mass

Consider two material points with masses m_1 and m_2 which interact with each other. We have by the Second and Third laws:

$$m_1\ddot{\boldsymbol{r}}_1 = \boldsymbol{F}_{12}, \qquad m_2\ddot{\boldsymbol{r}}_2 = -\boldsymbol{F}_{12}. \tag{21}$$

Dividing the first equation by m_1, the second by m_2 and then subtracting the second equation from the first one yields

$$\ddot{\boldsymbol{r}}_1 - \ddot{\boldsymbol{r}}_2 = \left(\frac{1}{m_1} + \frac{1}{m_2}\right)\boldsymbol{F}_{12},$$

or denoting $\boldsymbol{r} = \boldsymbol{r}_1 - \boldsymbol{r}_2$

$$\mu\ddot{\boldsymbol{r}} = \boldsymbol{F}_{12}, \qquad \mu = \frac{m_1 m_2}{m_1 + m_2}. \tag{22}$$

If, as it often happens, the force of interaction $\boldsymbol{F}_{12}$ depends only on the relative position of the particles (let us now drop the $_{12}$ subscript) the resulting equation

$$\mu\ddot{\boldsymbol{r}} = \boldsymbol{F}(\boldsymbol{r}) \tag{23}$$

is independent and self-contained, describing the motion of a virtual particle with "effective mass" μ. Of course, this single equation is not equivalent to the two equations (21), but if one adds to it the fact that the mass centre, located at

$$\boldsymbol{R} = \frac{m_1\boldsymbol{r}_1 + m_2\boldsymbol{r}_2}{m_1 + m_2}$$

moves with constant velocity, as there are no external forces, then together with $\dot{\boldsymbol{R}} = \boldsymbol{V} = \mathit{const}$ the two equations are equivalent to (21). What is important mathematically is the fact that the equation (23) has the same form in the mass centre frame, because the difference $\boldsymbol{r} = \boldsymbol{r}_1 - \boldsymbol{r}_2$ is invariant with respect to Galileo transforms (7). Thus, the system's motion has been broken up into the independent motion of the mass centre, or the system as whole and the relative motion of the particles with respect to each other, which is described by (23). The reason it is important is that in general a system of two particles has six degrees of freedom, so is generally described by a system of six (!) second-order differential equations. But the above argument shows that in this system of equations, by means of passing to the mass centre frame of reference, the three equations corresponding to the motion of the mass centre are trivial. So, rather than solving six equations, one has only to solve three equations in (23). We shall later see that as long as

$$\boldsymbol{F}(\boldsymbol{r}) = f(r)\frac{\boldsymbol{r}}{r}, \tag{24}$$

i.e, when the interaction force (such as gravity, Coulomb, etc.) depends only on the distance between the interacting bodies and acts along the line connecting them, the equations (23) can be solved completely, by using polar coordinates. The force in this case is called *central.*

There is no physical meaning here in the effective mass notion – this is only a mathematical convenience. As an example, consider $m_1 = m$, mass of the Earth and $m_2 = M$, mass of the Sun; assume that the Earth is rotating around the Sun, i.e. the distance $r = const$. The force

$$\boldsymbol{F} = -G\frac{mM}{r^2}\frac{\boldsymbol{r}}{r}$$

is gravity, and the minus sign shows that is it directed opposite to the vector starting at the Sun and ending at the Earth (both are treated as material points).

Then, from (22) we have

$$\ddot{\boldsymbol{r}} = -G\frac{M+m}{r^3}\boldsymbol{r} = -G\frac{M+m}{r^2}\boldsymbol{n},$$

where $\boldsymbol{n}$ is the principal normal vector to the trajectory of the Earth with respect to the Sun (see the section on Kinematics).

It follows that the relative motion of the Earth with respect to the Sun is uniform, with the angular velocity ω defined by (4) as

$$G\frac{M+m}{r^2} = \omega^2 r,$$

namely

$$\omega = \sqrt{G\frac{M+m}{r}} \approx \sqrt{G\frac{M}{r}}.$$

If, however, the mass of the Earth were equal to the mass of the Sun and the radius of the orbit were the same, the Earth would have to move $\sqrt{2}$ times faster, so a year would be $\sqrt{2}$ times shorter. Note that so far we have, in fact, assumed that the distance between the Sun and the Earth is constant; in reality this is true only approximately, and a more sophisticated analysis would tell that the orbit is, in fact, an ellipse.

3.4.2 Appendix. Jet propulsion

In all the above examples a mechanical system consisted of individual particles whose masses were fixed. However nothing prevents one from considering a "continuous" exchange if mass between different bodies that constitute a mechanical system. Such phenomena are characteristic of dynamics involving liquid or gas. The simplest example is jet propulsion.

Consider a closed system "rocket payload plus fuel". Fuel burns in the jet engine and is ejected to propel the rocket. By the mass of the rocket $m(t)$ we mean a variable quantity representing the mass of the rocket itself plus the mass of fuel inside of it at a given time. Suppose, over a time dt the mass of the rocket has changed by $dm < 0$. I.e., dm kilos of fuel have been burned into gas and ejected, with some velocity $\boldsymbol{v}_g$. (The law of conservation of mass is highly accurate for chemical reactions underlying combustion, due to their relatively small energy yield). Then the velocity of the rocket $\boldsymbol{v}$ gets an increment $d\boldsymbol{v}$, so that the net change of momentum is zero:

$$(m + dm)(\boldsymbol{v} + d\boldsymbol{v}) + dm\boldsymbol{v}_g - m\boldsymbol{v} = 0.$$

Neglecting higher order terms, and introducing the relative velocity $\boldsymbol{v}_r = \boldsymbol{v}_g - \boldsymbol{v}$ at which the gas is being ejected from the engine, we have

$$m\frac{d\boldsymbol{v}}{dt} = \frac{dm}{dt}\boldsymbol{v}_r. \tag{25}$$

This looks like the Second law, with the right-hand side where $\boldsymbol{F}_r = \boldsymbol{v}_r \frac{dm}{dt}$ which may be regarded as the jet propulsion force, the product of the rate at which the fuel burns and the speed at which it is ejected. The mass m, however, is now also a function of time.

If there is an additional external force $\boldsymbol{F}_e$ acting on the rocket, then clearly

$$m\frac{d\boldsymbol{v}}{dt} = \boldsymbol{F}_r + \boldsymbol{F}_e.$$

If there are no external forces, the rocket moves along a line, and the engine works steadily, i.e. $v_r = const$, then variables in the differential equation (25) separate:

$$dv = v_r \frac{dm}{m},$$

so

$$\frac{v}{v_r} = \log\left(\frac{m_0}{m}\right), \qquad \text{or} \qquad m = m_0 e^{-\frac{v}{v_r}}. \tag{26}$$

where m_0 was the start mass of the rocket when $v = 0$. The latter formula bears the name of Tsiolkovsky (1859–1935). It shows that in order for the rocket to reach the speed Nv_r, its mass shall decrease e^N times from the original mass m_0. I.e. fuel to be burned has to take the huge $1 - e^{-N}$ proportion of the total launch mass.

In today's rockets v_r is some two kilometers per second, and it seems unlikely that combustion-based engines could account for v_r greater than 4 km/s. In order for a rocket to overcome the Earth's gravity and be able to reach, say the Moon, it needs to achieve $v = 11.2$ km/s, while to leave the limits of the Solar system the minimum speed is 16.7 km/s. Thus to achieve the former goal, payload may constitute no more than $e^{-11.2/4} \approx 1/17$ of the brutto start weight of the rocket. To achieve the latter goal, the share of payload drops to approximately 1/64.

If one adds to it that fuel is needed to return, then for a round-trip to a planet whose mass, for simplicity, equals the mass of the Earth one needs to have $m/m_0 \approx \frac{1}{17^2}$. In addition, fuel is needed to break, at least to decrease the rocket's velocity by several km/s to be captured into the target's gravitational field. This adds another factor of 10 to the denominator. Navigate, have something for emergency, etc. and in fact, for a return trip one would need a reasonably realistic ratio $m_0/m \approx 3600$, which means 3.6 tons of fuel per kilo of payload, and astronauts better be thinner than jockeys. Nonetheless, such a mass ratio is still practically achievable. But for interstellar missions chemical combustion-based engines seem to be utterly hopeless. (Relativistic corrections to Tsiolkovsky's formula are easy to take into account using (13) but unfortunately they only make things worse. If one wants the rocket to move with the speed comparable with the speed of light, the ratio m_0/m grows even greater than predicted by the Tsiolkovsky formula.) In order to reach the nearest star and complete a return trip over a lifetime, traveling, say at 1/4 of the speed of light would require the ratio m_0/m to be some 10^{3257}. So, for each gram of payload one would need some 10^{3251} tons of fuel, while the mass of the observable Universe is "only" some 10^{56}g. What makes today's jet propulsion engines unacceptable for such missions is the relatively small value of the gas jet ejection speed v_r, some kilometers per second only, which is the consequence of the relatively small energy yield of chemical reactions underlying the engine's design. Even today largely hypothetical atomic hydrogen-fueled engines would yield only $v_r \approx 10$km/s – still way too small to enable interstellar travel but possibly quite comfortable for moving around the Solar system. (A hydrogen engine would require combustion temperatures of up to 5000 degrees Celsius, which poses a great challenge to engineers.)

4 Methods of integration of Newton's equation

Solving, or *integrating* Newton's equations

$$\dot{\boldsymbol{p}}_i = \boldsymbol{F}_i, \qquad i = 1, \ldots, N$$

for a system of N particles is the principal objective of dynamics. In general this poses a great challenge, as above we have a system of $3N$ second-order differential equations, which are usually all coupled and nonlinear. Complete integration of Newton's equations is possible only in a very narrow class of so-called classical *integrable systems*. E.g., a body sliding down an incline with angle α in a stationary gravitational field is an integrable system (of one degree of freedom): its acceleration along the incline is constant and equals $g \sin\alpha$, which certainly enables one to write down the trajectory $\boldsymbol{r}(t)$. We have also seen that momentum conservation, leading to the idea of mass centre enables one to reduce the number of equations that have to be dealt with.

More complex systems require more sophisticated methods of integration (solving). These methods should first and foremost take into account the system's geometric features, using the right set of generalised coordinates. E.g. in the pendulum example we were able to arrive in the differential equation (5) after having taken into account its circular geometry which expressed itself in the fact that the natural variable to characterise the pendulum's state is the angle α between the pendulum and the vertical.

The concepts to be considered further, with the final goal being integration of Newton's equations, are force impulse, work and energy, and angular momentum.

4.1 Force impulse

If a force acts on a material point, then the momentum gained by the point depends not only on the magnitude of the force, but also on the duration of the interaction. Suppose the force $\boldsymbol{F}$ in Newton's $\dot{\boldsymbol{p}} = \boldsymbol{F}$ has acted between times t_1 and t_2. Multiplying both sides of the above equation by dt and integrating in time between t_1 and t_2 yields

$$\boldsymbol{p}(t_2) - \boldsymbol{p}(t_1) = \int_{t_1}^{t_2} \boldsymbol{F} dt.$$

The right-hand side is called the *impulse of force* $\boldsymbol{F}$. The above statement reads that the change of a body's momentum equals the impulse of the net force. For a system of some number N particles we use (17) to get the same statement:

$$\boldsymbol{P}(t_2) - \boldsymbol{P}(t_1) = \int_{t_1}^{t_2} \boldsymbol{F}_e dt, \tag{27}$$

where $\boldsymbol{F}_e$ is the net external force acing on the system and $\boldsymbol{P}$ is total momentum, or the mass centre momentum. The easiest example is when the force is constant and has magnitude F. Then the change of the momentum equals in absolute value $F(t_2 - t_1)$.

In general, external forces would most often depend on the positions of the particles in the system. In this case the force impulse

$$\int_{t_1}^{t_2} \boldsymbol{F}_e dt = \int_{t_1}^{t_2} \boldsymbol{F}_e(\boldsymbol{r}_1(t), \ldots, \boldsymbol{r}_N(t)) dt,$$

which is not easy to find. For this reason the force impulse is rarely used to integrate Newton's equations of motion. Still, it is useful to explain at least qualitatively some physical phenomena.

One can make an interesting experiment to this end. Hang a small dumbbell on one thread, and tie another thread to the bottom of the weight. If one pulls gently at the second thread, increasing the force, the upper thread will snap. Indeed, let T_0 be the limit tension that a thread can stand before snapping. If one pulls gently at the lower thread, the system can be always viewed as at rest, and therefore the tension of the upper thread must be equal to the weight of the dumbbell plus the force T_2 at which one pulls at the lower thread. Thus the upper thread will break when $T_2 + Mg = T_0$.

If one pulls at the lower thread abruptly however, the lower thread will break instead. Indeed, before the upper thread will snap it has to be stretched by some critical increment of length. In order to do so, the dumbbell must be set in motion. The issue is that by (27) the heavier the body, the greater force impulse is required to set it in motion. Setting the heavy dumbbell in motion requires a considerable force impulse, and if a force acts over a very short time even though the force itself may be of considerable magnitude, the dumbbell practically would not move, as its momentum, being proportional to M, will change only little from zero. This means that upon a very short time of interaction T_1 is still approximately equals to Mg (the upper thread has barely stretched). On the other hand T_2 shall compensate the force at which one pulls, and will snap if this force exceeds T_0, provided that it has been applied quickly enough.

The ideas of work, energy and angular momentum discussed further are more complex, especially when they apply to system with more than one degree of freedom (as the notion of angular momentum always does). We shall start out by introducing the notion of work and potential for one degree of freedom systems.

4.2 Work and potential energy in the case of one degree of freedom

Consider a particle of mass m constrained to move along the straight line. This is a one degree of freedom mechanical system: the state of the system is described by the particle's position x and velocity $v = \dot{x}$ at any time.

If one knows $x(0) = x_0$ and $\dot{x}(0) = v_0$, and all the forces acting on the particle, then solving Newton's equation

$$m\ddot{x} = F,$$

with the above initial conditions will determine the state of the system, i.e. the position $x(t)$ of the particle, for any $t \in \mathbb{R}$: both in the past and future.

The force F may in general be the function of $x, \dot{x}$, and t. However, a very important and typical case is when F depends on x only. This corresponds to a physical reality of an external *stationary force field* such as, say gravity. The intensity of the force field, i.e. the measure of its effect on mechanical bodies depends only on the position of these bodies. Such are, for instance, the gravitational force field, created by the Sun that governs planetary motions in the Solar system or the electrostatic fields of atomic nuclei that govern the electrons in an atom (although the electrons live according to law of quantum, rather than classical mechanics). Such is the gravitational field of the Earth close to the ground can be approximately regarded as constant, which is the trivial dependence on hight x.

Mathematically, in the above one-dimensional model we will further assume $F = F(x)$. In this case we have $m\frac{dv}{dt} = F(x)$ and multiplying it by $dx = vdt$ we get

$$mv\,dv = F(x)dx \qquad \text{or} \qquad m\frac{dv}{dx} = F(x). \tag{28}$$

Suppose, over some time interval between t_1 and t_2 the particle moves from position x_1 to position x_2, and its velocity changes from v_1 to v_2 accordingly. The differential equation (28) does not contain t explicitly at all, and one can look for its solution in the form $v(x)$, i.e as the dependence of the particle's velocity on its position. Integrating (28), where the variables x, v have been separated, we have

$$\frac{mv_2^2}{2} - \frac{mv_1^2}{2} = \int_{x_1}^{x_2} F(x)dx, \tag{29}$$

and if v_2, x_2 are regarded as variables, and v_1, x_1 initial conditions, then this alone defines the dependence $v(x)$.

The quantity

$$K(v) = \frac{mv^2}{2}$$

is called the particle's *Kinetic energy*, and the most important thing about it is that $K \geq 0$ and equals zero only if $v = 0$, i.e., when the particle is a rest (classical mechanics does not consider massless particles such as photons; in Relativistic mechanics massless particles *must* move with the speed of light.) Note, however, that the particle which is *instantly* at rest will *stay* at rest only if there are no forces acting on it.

The right-hand side of (29) is called *work* done by the force F. Observe that as long as the net force F can be partitioned into a sum of several particular forces acting in a particular problem, it makes sense to talk about the work done by each particular force: it simply follows from the additivity property of an integral

$$\int (F_1 + F_2 + \ldots + F_N)dx = \int F_1 dx + \int F_2 dx + \ldots + \int F_N dx.$$

So (29) can be read as follows: the change in a particle's kinetic energy equals work done by external forces.

An important notion related to work is *power*

$$W = \frac{dA}{dt},$$

which is work done by a particular force per unit time. Hence, to compute power, one must fix x_1 in the right-hand side of (29) and consider x_2 as a variable in the integral for work: let us assume that x_1 corresponds to $t = 0$ and x_2 to a variable time t. Then

$$W = \frac{d}{dt}\int_{x_1}^{x_2(t)} F(x)dx = \frac{d}{dt}\int_0^t F(x(t'))v(t')dt' = F(x(t))v(t). \tag{30}$$

In other words, power equals force, acting on the particle at a given moment times the particle's velocity.

Solving the differential equation (28) is tantamount to finding the integral for work. Let us denote

$$U(x) = -\int F(x)dx, \qquad \text{so} \qquad F(x) = -\frac{dU}{dx}. \tag{31}$$

The antiderivative $U(x)$ is called the *potential energy*, of the particle, corresponding to the force F. Being an indefinite integral, it is defined up to an additive constant. However, the constant plays no dynamical role: the relation (29) can be now rewritten as

$$K_2 - K_1 = -(U_2 - U_1), \qquad \text{or} \qquad E = K(v) + U(x) = const. \tag{32}$$

This formula represents the *law of conservation of energy*: total energy of a particle, which is the sum of its kinetic and potential energy, is constant throughout the particle's motion.

Note that the introduction of potential energy in (31) became possible because the force $F = F(x)$ was assumed to be a function of x only. Such forces are called *conservative*, because they enable the concept of potential energy that leads to energy conservation. Comparing (32) and (29) we see that work done by a conservative force equals minus the change in the potential energy, which depends only on the starting point x_1 and x_2, but not on *how* the particle got from x_1 to x_2. E.g., if an absolutely elastic ball has been dropped from a hight x_1 and let to jump, its speed at the height x_2 is always the same, no matter how many times the ball has been jumping up and down in the mean time. The notions of conservative force and potential will be soon generalised to forces acting in space, rather than along straight line.

The expression (32) provides full description of the particle's trajectory in the form $v = v(x)$, i.e. tells one what the particle's velocity will be when it visits a particular point in space. One can plot the curves $K(v)+U(x) = const$ in the (x, v)-plane: to do this no advanced maths is needed as long as $U(x)$ has been found. Observe that as $K \sim v^2$, these curves will always be symmetric with respect to the x-axis. They are called *phase curves*.

So far, we have been moving towards revealing more and more information about the solution $x(t)$ of Newton's equation. Our ultimate wish is to get $x(t)$, i.e. to plot the solution's graph in the (t, x)-coordinate plane. We can already do this in the (x, E)-plane: the solution $x(t)$ is represented there by a horizontal line $E = const$. Observe that given E, the particle can possibly find itself only at those places x, where $E \geq U(x)$, for otherwise would imply that the particle's kinetic energy becomes negative – an impossibility.

Furthermore, in the (x, v)-plane the solution $x(t)$ will be represented by a phase curve $v = \pm\frac{2}{m}\sqrt{E - U(x)}$. The shape of this curve contains further information about $x(t)$. Now it is time to find $x(t)$ explicitly.

Mathematically, the main merit of (32) in comparison to (28) is that one act of integration has already been performed, and in order to solve the Newton second-order equation one has to integrate only two times. The value of the total energy E can be found from initial conditions – it depends on x_0 and v_0 only. But now (32) represents the first order differential equation

$$\frac{dx}{dt} = \pm\sqrt{\frac{2}{m}(E - U(x))}. \tag{33}$$

The presence of $\pm$ in front of the square root brings no extra difficulty: it merely accounts for the particle moving in the positive or negative direction along the x-axis. Suppose, $v_0 > 0$, then until the particle stops (if ever) or, in fact, as long as there is a reason to believe that $x(t)$ will be an everywhere differentiable function (so the velocity will change continuously) one should choose the plus sign. In the latter equation the variables have separated again, and we can continue as follows:

$$dt = \sqrt{\frac{m}{2}}\frac{dx}{\sqrt{E(x_0, v_0) - U(x)}}.$$

We have written $E = E(x_0, v_0)$ to emphasise how the initial conditions will enter the final solution explicitly. Integration now yields

$$t = \sqrt{\frac{m}{2}}\int_{x_0}^{x}\frac{dy}{\sqrt{E(x_0, v_0) - U(y)}}. \tag{34}$$

The potential U being either given or having been computed from the force F by (31), we may now be able to find the integral in the right-hand side to have the trajectory $x(t)$ of the particle given implicitly as $t = t(x)$. Finding the inverse function will finally secure the trajectory $x(t)$, with all the initial conditions incorporated.

We have therefore described a method of solving Newton's equation for a one degree of freedom system no matter what $F(x)$. The merit of this is that second order differential equations are in general more difficult than first order equations, but we have reduced the second order equation $m\ddot{x} = F(x)$ to the first order equation (33).

4.2.1 Example: mass on a spring

Consider a mass oscillating on a spring, with no other forces than $F(x) = -kx$. Newton's equation $m\ddot{x} = -kx$ is easy to solve anyway, the standard technique for second-order linear equations yields

$$x(t) = a\sin(\omega t + \phi),$$

where $\omega = \sqrt{\frac{k}{m}}$ and the constants a, ϕ can be found by solving algebraic equations $x(0) = x_0$, $\dot{x}(0) = v_0$. Let us go down the more general "Energy way" instead. (It was the simplicity of the force $F = -kx$ that has enabled us to just write down the above solution, but the second-order equation technique would fail if, for instance, instead of a linear force $F = -kx$ one had the non-linear $F = -k\sin x$ to be considered soon as well.)

First, we have $U(x) = \frac{kx^2}{2} + C$ for the potential energy. The choice of C is irrelevant, and it only makes sense to choose $C = 0$. We now need to take the initial conditions into account to find the system's energy. Suppose, at the time $t = 0$ the spring is not deformed, but the body acquires (after a kick of sorts) an initial velocity $v_0 > 0$. Then the energy $E = \frac{mv_0^2}{2} = const.$

Now, (34) becomes

$$t = \sqrt{\frac{m}{2}} \int_0^x \frac{dy}{\sqrt{E - \frac{ky^2}{2}}} = \sqrt{\frac{m}{2E}} \int_0^x \frac{dy}{\sqrt{1 - \frac{k}{2E}y^2}} = \sqrt{\frac{m}{k}} \left(\arcsin \sqrt{\frac{k}{2E}} x \right).$$

Therefore, as $E = \frac{mv_0^2}{2}$,

$$x(t) = \sqrt{\frac{mv_0^2}{k}} \sin(\omega t),$$

where $\omega = \sqrt{\frac{k}{m}}$. Let us verify that the factor $a = \sqrt{\frac{2E}{k}} = \sqrt{\frac{mv_0^2}{k}}$ makes sense. It gives the amplitude of the spring's oscillations, when x reaches its minimum or maximum. In both cases $\dot{x}$ will be zero, so all the system's energy will be represented by the spring's potential energy. I.e $\frac{ka^2}{2} = E$, which gives a as it is.

To illustrate the advantage of the energy method, consider now a non-linear spring, where $F = -k\sin x$. Newton's equation is then

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

which is the same as the equation (5) for the pendulum, where $\omega^2 = \frac{g}{r}$. And as the force is 2π-periodic in x and integrates into zero over $[-\pi, \pi]$, we can consider x modulo 2π, i.e. as an angle as well. So, in fact, we *are* now dealing with the pendulum, having identified k with g and m with the length r, see (5).

The potential energy now is $U(x) = -k\cos x$, and let us consider a particular case when the system's total energy E equals k, the maximum of the potential, corresponding to $x = \pi$. In the case of the pendulum, when $x = \alpha$ is the angle formed by the pendulum with the vertical, this corresponds to the pendulum standing upright (suppose that instead of a string, the weight of the pendulum is suspended on a massless spoke of length r).

We have

$$t = \sqrt{\frac{m}{2k}} \int_0^x \frac{dy}{\sqrt{1 + \cos y}}.$$

Using the identity $1 + \cos y = 2\cos^2 y/2$, we have, using the fact that $\int \sec y dy = \ln|\tan(y/2 + \pi/4)|$ (derive it by multiplying the numerator and denominator by $\cos y$, then use the substitution $u = \sin y$):

$$t = \sqrt{\frac{m}{4k}} \int_0^x \frac{dy}{\cos(y/2)} = \sqrt{\frac{m}{k}} \int_0^{x/2} \frac{dy}{\cos y} = \sqrt{\frac{m}{k}} \ln|\tan(\pi + x)/4|.$$

Hence,

$$x(t) = 4\arctan e^{\omega t} - \pi.$$

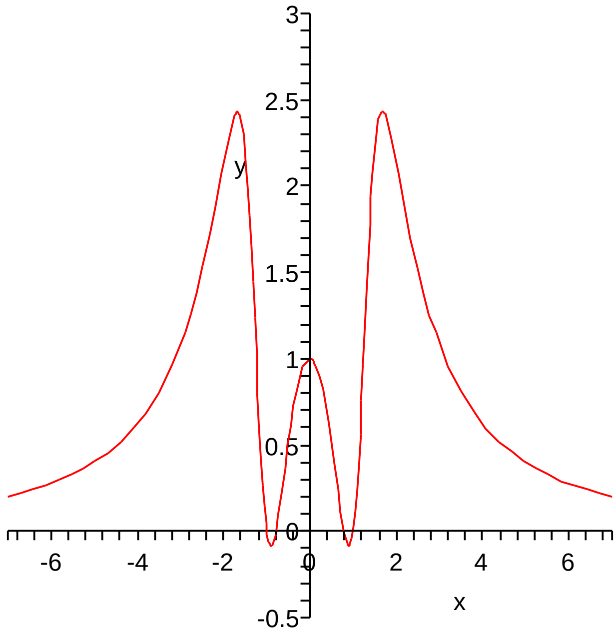
Indeed, if $t = 0$, we have $x(0) = 4\arctan 1 - \pi = 0$. If t increases, $\arctan e^{\omega t}$ limits to $\pi/2$, so as $t \to \infty$, $x(t)$ asymptotically approaches π. This shows that if a spoke pendulum at rest is given the right velocity, it will be approaching an upright position over infinite time, without ever being able to go over it. Similarly, as $t \to -\infty$ $x(t)$ goes asymptotically to $-\pi$, which in terms of the pendulum means approaching the upright position "from the other side". Therefore, one can imagine that at time $t = -\infty$, the pendulum was standing upright in unstable equilibrium, with $x = -\pi$. Then it was given an infinitesimal push and started falling, at time $t = 0$ passed the bottom point $x = 0$, and by the time $t = +\infty$ will swing back to the upright position $x = \pi$.

4.3 Potential energy and domains of possible motions

Knowing the potential energy function $U(x)$ enables one to say a lot about the qualitative features of the particle's motion, without computing the integral (34). Indeed, completely solving Newton's equation $m\ddot{x} = F(x)$ takes two integrations; introducing $U(x)$ has involved one. The qualitative analysis of the solutions simply follows from comparing the function $U(x)$ with different values of total energy, which is considered as given. It is solely based on the fact that

$$\frac{mv^2}{2} = E - U(x) \geq 0.$$

First of all, given E, one can talk about the *domain of possible motions*. Namely given E, the particle cannot find itself at the position x where $E < U(x)$, due to the fact that the kinetic energy (and, in fact, mass) cannot be negative. Consider a particular example of the potential $U(x)$ given by the following figure.



One can think of this figure as a plot in coordinates (E, x). Then the trajectory of a particle, depending on E and x_0, will be given by a single horizontal line segment $E = const$, so that this segment always lies above the potential energy curve $U(x)$. The height of the segment over the curve $U(x)$ will determine the velocity v. Indeed,

$$v = \pm\sqrt{2(E - U(x))/m}. \tag{35}$$

This line segment, depending on E and x_0 will be either finite or infinite, and hence one can talk about the particle's finite, which in this case would mean periodic, as well as infinite motions.

Suppose the total energy of the particle is $-.1 < E < 1$ and at $t = 0$ the particle is positioned at x_0 equal to plus or minus 1. Then the motion of the particle $x(t)$ will be finite and periodic: the particle will never be able to leave the *potential well* in which it is confined. Depending on a particular value of E in the above range, $x(t)$ will oscillate back and force between the values x_m and x_M, which are defined by the intersection of the horizontal segment $E = const$ with the potential energy curve. E.g. if $E = .9$ and the particle starts at $x_0 = 1$ (with some velocity v_0 to ensure $E = .9$) then approximately $x_m = .4$ and $x_M = 1.2$. The points x_m and x_M are called *turning points:* at these points $E = U$, so the velocity is zero, and the particle cannot penetrate beyond these points, or its kinetic energy would become negative.

According to (34), the period of oscillations between the turning points x_m and x_M is

$$T = 2\sqrt{\frac{m}{2}}\int_{x_m}^{x_M} \frac{dx}{\sqrt{E - U(x)}},$$

the factor 2 is there because the integral itself is how long it would take to go from x_m to x_M. It will take the same time to go back – this is due to the fact that $K \sim v^2$, or equivalently because the integral from x_M to x_m equals minus the integral from x_m to x_M, or equivalently to the presence of the $\pm$ sign in (33): when v is negative, the minus sign in front of the square root in (35) has to be chosen.

A totally valid analogy for understanding these concepts would be imagining a ball that can roll freely up and down a one-dimensional hill with profile $U(x)$ in a constant gravity field. (Indeed, the potential energy in the uniform gravity field is proportional to the altitude x). Then if the ball at $t = 0$ were placed at x_m and let go, it would roll up and down the wall of the well, form x_m to x_M and backwards. Similarly, if $E = .5$ and $x_0 = x_m \approx 4.4$, the ball would roll right with increasing speed. At $x \to +\infty$ its velocity would asymptotically reach the value $v = \sqrt{2E/m}$, as $U(x)$ goes to zero as $x \to +\infty$. This is an example of infinite motion. However, it is semi-infinite only: as long as $E < 2.5$, the ball (particle) would not be able to find itself left of x_m. (In quantum mechanics this is not the case: a quantum particle can "tunnel" through a potential wall, the probability of tunneling decreasing exponentially with the width and height of the wall.)

However, if the energy of the particle $E > 2.5$, the motion is double-infinite. Indeed, for any x, $K(x) = E - U(x)$ will be bounded from below, so the velocity will be bounded form below and retain the sign, and any x will be reached in positive/negative time.

Finally, the most interesting scenario is when E equals one of the local minima or maxima of the potential (in the figure such critical values are approximately $-.1$, 1, and approximately 2.5). If x_* is a local minimum/maximum of the potential, then the force $F(x_*) = -U'(x_*)$ acting on the particle at that point is zero. Hence, a particle which is at rest at x_* at $t = 0$ would stay there for ever, by the First law. In other words, for these *critical* values $E_* = E(x_*)$ of E, the Newton equation enables solutions $x(t) = x_*$, for all t. Such x_* are called equilibria. An equilibrium is *stable* if it corresponds to a local minimum of $U(x)$, because a slight increase of energy from E_* would result in finite motion only. (Alternatively, if there is small dissipation of energy from the system, the particle, after having been perturbed from equilibrium would eventually come to rest at the equilibrium point.) If however x_* is a local maximum of U, it corresponds to an unstable equilibrium: a ball positioned right on the very tip of a hill would fall off upon receiving the slightest push giving it extra energy.

Observe that if x_* is a critical point of $U(x)$, then we can Taylor-expand

$$U(x) = U(x_*) + \frac{1}{2}U''(x_*)(x - x_*)^2 + \dots.$$

Indeed, as x_* is a critical point, the first derivative U' there is zero. Without loss of generality, the constant $U(x_*)$ can be rendered zero. Then,

$$U(x) \approx \frac{1}{2}U''(x_*)(x - x_*)^2, \tag{36}$$

which is the potential energy of a spring, stretched by $z = x - x_*$, with the spring constant $U''(x_*)$. Therefore, for the energies E slightly greater than $E_* = U(x_*)$, the motion

$$x(t) - x_* \approx a \sin(\omega t + \phi), \tag{37}$$

where the amplitude $a = x_* - x_m$, x_m being the left turning point, is determined by $E - E_*$, while ϕ is determined by the initial condition $x(0)$. Besides, $\omega = \sqrt{\frac{U''(x_*)}{m}}$. Mathematically this expresses the most fundamental idea of the Taylor approximation in calculus: in this case any smooth function $U(x)$ near its critical point x_* can be viewed as a parabola with coefficient $\frac{1}{2}U''(x*)$, as long as the latter is nonzero, the relative error of the approximation going to zero as x gets closer x_*.

If x_* corresponds to the local maximum of U however, and suppose $U''(x_*) < 0$, then denoting again $k = |U''(x_*)|$, for small deviations $z = x - x_*$ we have an approximate differential equation

$$m\ddot{z} - kz = 0.$$

Indeed, $\ddot{z} = \ddot{x}$ and $F = -dU/dx = -dU/dz = -U''(x_*)z$. Two linearly independent solutions of this equations are

$$x(t) - x_* \sim e^{\mp\omega t}, \tag{38}$$

where now $\omega = \sqrt{\frac{|U''(x_*)|}{m}}$. These solutions asymptotically approach zero: one in positive and the other in negative time. In terms of the above figure this means that if x_0 equals, for instance, 6 and $v_0 < 0$, so that and the total energy equals the maximum of U, which is approximately 2.5, then the ball will be rolling left up the hill whose tip is at $x = 2$, getting closer and closer to the top, reaching it ad infinitum. Similarly, if $x_0 = -.8$, $v_0 > 0$ and such that $E = 1$, the ball will be rolling right up the hill at $x = 0$ over infinite time, without ever getting over the top.

Similarly the motion with energy $E = 1$, starting at $x_M \approx 1.2$ and further moving towards the left is no longer periodic, or its "period"

$$T \sim \int_0^{x_M} \frac{dx}{\sqrt{1-U(x)}} = \infty.$$

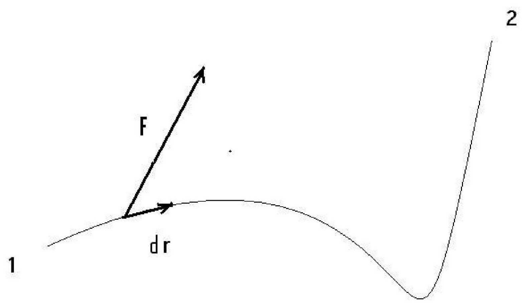
Indeed, near $x = 0$ we have $U(x) = 1 - \frac{1}{2}kx^2 + \dots$. Therefore the denominator in the above formula, as $x \to 0_+$ becomes proportional to $1/x$, and the antiderivative $\ln x$ is infinite at $x = 0$. This is just another expression of the above mentioned fact that localised near an unstable equilibrium, linearly independent solutions of the Newton equation have exponential form $e^{\mp\omega t}$.

4.4 Work and Energy in general

Let us now extend the concepts of work and potential energy beyond systems of one degree of freedom. We once again start out with the Second law, only now it is in the vector form:

$$m\frac{d\boldsymbol{v}}{dt} = \boldsymbol{F},$$

and we are therefore trying to solve three differential equations at the same time even if we deal with a single particle in 3d. Suppose, the particle is moving between the points 1 and 2 along the trajectory $\boldsymbol{r}(t)$, see fig. (In fact, the same argument works for $\boldsymbol{r} \in \mathbb{R}^n$ for any n, which makes sense, because $\boldsymbol{r}$ may generally denote the aggregate of the coordinates used to describe a mechanical system, which may well contain more than just a single particle.)



Let $d\boldsymbol{r}$ be an infinitesimal change of the particle's radius-vector along the trajectory. Take the dot product of both sides of the Newton equation with $d\boldsymbol{r}$ and observe that $d\boldsymbol{r} = \boldsymbol{v}dt$. One gets

$$m\boldsymbol{v} \cdot d\boldsymbol{v} = dA, \qquad \text{where } \; dA = \boldsymbol{F} \cdot d\boldsymbol{r}. \tag{39}$$

The quantity dA is called infinitesimal, or elementary work, done by the force $\boldsymbol{F}$ over the position change $d\boldsymbol{r}$, clearly

$$dA = |\boldsymbol{F}||d\boldsymbol{r}|\cos\phi,$$

where ϕ is the angle between the directions of the vectors $\boldsymbol{F}$ and $d\boldsymbol{r}$. In the Newton formula $\boldsymbol{F}$ is the net force, acting on a particle. However, if $\boldsymbol{F}$ is a superposition of forces

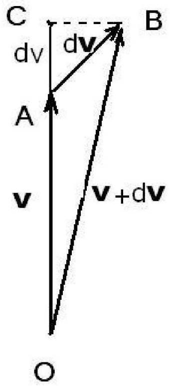
$$\boldsymbol{F} = \boldsymbol{F}_1 + \boldsymbol{F}_2 + \dots,$$

one can naturally talk about elementary work, done independently by each individual force. Observe that if a force and the velocity $\boldsymbol{v}$ are perpendicular to each other, then $dA = 0$.

We will now be integrating equation (39) which means summing the elementary bits (39) over a huge numebr of such, sums becoming integrals in the limit. The left-hand side of (39) enables one to introduce the notion of kinetic energy

$$K(v) = \frac{mv^2}{2},$$

similar to (28), where $v = |\boldsymbol{v}| = \sqrt{\boldsymbol{v}\cdot\boldsymbol{v}}$ is speed. In order to arrive in K we claim that $\boldsymbol{v}\cdot d\boldsymbol{v} = vdv$, and in fact, the same relation is true for any vector. Note: $\boldsymbol{v}\cdot d\boldsymbol{v}$ is the dot product of the velocity $\boldsymbol{v}$ and its increment $d\boldsymbol{v}$. But vdv is the product of two scalars: the speed v and the increment dv of speed. One way to verify this is geometric, see the following figure,



where is boils down to the vector identity $\vec{OA}\cdot\vec{AB} = |OA||AC|$. Equivalently, analytically

$$m\boldsymbol{v}\cdot d\boldsymbol{v} = m(v_x dv_x + v_y dv_y + v_z dv_z) = m\,d\left(\frac{v_x^2+v_y^2+v_z^2}{2}\right) = dK. \tag{40}$$

Observe that the above differential formulae represent the limit cases as $\Delta\boldsymbol{r}\to 0$ of "finite" formulae

$$\Delta K = mv\Delta v + \ldots,\quad \Delta\boldsymbol{r} = \boldsymbol{v}\Delta t + \ldots,\quad \Delta A = \boldsymbol{F}\cdot\Delta\boldsymbol{r} + \ldots,$$

where $\ldots$ denote higher order terms in small quantities Δv, Δt, $\Delta\boldsymbol{r}$. The relative contribution of the omitted terms will go to zero as Δv, Δt, $\Delta\boldsymbol{r}$ go to zero.

We now integrate (39) between the particle's states 1 and 2, characterised by positions and velocities $\boldsymbol{r}_1,\boldsymbol{v}_1$ and $\boldsymbol{r}_2,\boldsymbol{v}_2$, occurring at times t_1 and t_2, respectively. Mathematically, we represent the trajectory $\boldsymbol{r}(t)$ as a broken line consisting of straight segments $\Delta\boldsymbol{r}_i$, $i=1,\ldots,N$ whose length goes to zero and number N to infinity. Summing the left-hand side is easy: we get just a definite integral in a single variable v:

$$\int_{v_1}^{v_2} mv\,dv = \int_{v_1}^{v_2} dK(v) = \frac{mv_2^2}{2} - \frac{mv_1^2}{2} = K_2 - K_1,$$

similar to the left-hand side of (29). The value depends only on the initial speed v_1 and final speed v_2, no matter how v changed form v_1 to v_2. Indeed, let us mentally partition the trajectory on the "elementary work" figure into a huge number of small bits $\Delta\boldsymbol{r}_i$. Then

$$\int_{v_1}^{v_2} dK(v) \approx \sum_i \Delta K(i) = K_2 - K_1,$$

because in the latter sum everything, but the first and the last term cancels telescopically.

Integration of the right-hand side of (39) means summing elementary works $dA\approx\Delta A$ on each bit $\Delta\boldsymbol{r}_i$ of the trajectory $\boldsymbol{r}(t)$. Such a sum would generally depend on the force $\boldsymbol{F}$, *as well as* the trajectory $\boldsymbol{r}(t)$. Denoting γ_{12} the piece of the trajectory $\boldsymbol{r}(t)$ between the endpoints $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ one gets

$$A_{12} = \int_{\gamma_{12}} dA = \int_{\gamma_{12}} \boldsymbol{F}\cdot d\boldsymbol{r} = \lim_{N\to\infty}\sum_{i=1}^{N}\Delta A_i = \lim_{N\to\infty}\sum_{i=1}^{N}\boldsymbol{F}_i\cdot\Delta\boldsymbol{r}_i, \tag{41}$$

where the vector $\Delta\boldsymbol{r}_1$ begins at $\boldsymbol{r}_1$ and the vector $\Delta\boldsymbol{r}_N$ ends at $\boldsymbol{r}_2$, and for $i=1,\ldots,N$, the force $\boldsymbol{F}_i$ is applied at the starting point of each $\Delta\boldsymbol{r}_i$.

Thus, the analog of (32) is now

$$K_2 - K_1 = A_{12}:\qquad \frac{mv_2^2}{2} - \frac{mv_1^2}{2} = \int_{\gamma_{12}} \boldsymbol{F}\cdot d\boldsymbol{r}. \tag{42}$$

The change in kinetic energy equals the net work of all the forces. Observe that $v^2 = \boldsymbol{v}\cdot\boldsymbol{v}$.

Thus work in more than one dimension defined via (41) represents a so-called *path* integral. It depends in general not only by the endpoints 1 and 2, but the path γ_{12} connecting them. Given $\boldsymbol{r}(t)$, and hence $\boldsymbol{v}(t)$ by differentiation, computing work is an easy task. First off, $d\boldsymbol{r} = \boldsymbol{v}(t)dt$. The force $\boldsymbol{F}$ may be the function of $\boldsymbol{r}$, $\boldsymbol{v}$, and even t, but knowing $\boldsymbol{r}(t)$ enables, by substituting all the time dependencies into $\boldsymbol{F}$ express it as a vector-function $\boldsymbol{F}(t)$ only. Suppose the endpoints correspond to the times t_1 and t_2. Then

$$A_{12} = \int_{t_1}^{t_2} \boldsymbol{F}(t)\cdot\boldsymbol{v}(t)dt,$$

the usual integral of a scalar function. In the same fashion, power

$$W = \frac{dA}{dt} = \boldsymbol{F}(t)\cdot\boldsymbol{v}(t).$$

The real question though is how to use the notion of work in order to get $\boldsymbol{r}(t)$, which is unknown. For this purpose, in the case of one degree of freedom, we could introduce potential energy $U(x)$, associating it via $F = -U'$ with any force $F = F(x)$, depending only on x. Can this be done in the case of more than one degree of freedom? Yes and no, depending on the force. We would now like to claim $\boldsymbol{F}(\boldsymbol{r}) = -\nabla U(\boldsymbol{r})$ for some scalar function $U(\boldsymbol{r})$.

The major difference in comparison with the case of one degree of freedom is that there, as long as the force is the function of the position x only, one has the Newton-Leibniz formula, which says that the integral $\int_{x_1}^{x_2} F(x)dx$ depends only on x_1 and x_2. In higher dimensions, this is no longer the case: there are many examples of $\boldsymbol{F}(\boldsymbol{r})$, for which $\boldsymbol{F}(\boldsymbol{r}) = -\nabla U(\boldsymbol{r})$ is simply impossible. Simple intuition suggests that even in two degrees of freedom the force $\boldsymbol{F}(x,y)$, having *two* components (F_x, F_y), each of which is a function of x and y is quite unlikely to be the minus gradient of just *one* function $U(x,y)$: there are many more pairs of functions $(F_x(x,y), F_y(x,y))$ than single functions $U(x,y)$.

So, potential energy can not be introduced only for any force, and the work done by a force between the initial and final states 1 and 2 may depend on not just these states but the path γ_{12} connecting them.

E.g., suppose the particle is moving in the xy plane and the endpoints 1 and 2 are both on the x-axis, with coordinates $(0,0)$ and $(1,0)$; suppose $\boldsymbol{F} = \frac{1}{y^2+1}\boldsymbol{i}$ is directed along the x-axis and gets smaller away from it. Then if the path γ_{12} is the segment of the x-axis connecting the endpoints, obviously one has $A_{12} = 1$. On the other hand, if the path γ first goes up along the y-axis from $y = 0$ to $y = N$, then proceeds horizontally from $(0, N)$ to $(1, N)$, and then down from $(1, N)$ to $(1, 0)$, the work of $\boldsymbol{F}$ along the going up and down sections is zero, because there $\boldsymbol{F}$ and $d\boldsymbol{r}$ are perpendicular. On the horizontal section the work equals $\frac{1}{1+N^2}$. Thus the work done by $\boldsymbol{F}$ in this case, despite $\boldsymbol{F} = \boldsymbol{F}(\boldsymbol{r})$ only and is independent of t and $\boldsymbol{v}$, does depend on how the particle gets from point 1 to point 2.

But Nature has arranged that the key physical interaction forces *do* allow for potential energy. A force $\boldsymbol{F} = \boldsymbol{F}(\boldsymbol{r})$ is called *potential* if there exists a scalar function $U(\boldsymbol{r})$ such that $\boldsymbol{F} = -\nabla U(\boldsymbol{r})$. I.e. if

$$\boldsymbol{F}(\boldsymbol{r}) = F_x(\boldsymbol{r})\boldsymbol{i} + F_y(\boldsymbol{r})\boldsymbol{j} + F_z(\boldsymbol{r})\boldsymbol{k}, \qquad \text{then } F_x = -\frac{\partial U}{\partial x},\ F_y = -\frac{\partial U}{\partial y},\ F_z = -\frac{\partial U}{\partial z}.$$

If this is the case, the elementary work

$$-dA = \nabla U(\boldsymbol{r})\cdot d\boldsymbol{r} = \frac{\partial U}{\partial x}dx + \frac{\partial U}{\partial y}dy + \frac{\partial U}{\partial z}dz = dU.$$

Alternatively,

$$-\Delta A = \nabla U(\boldsymbol{r})\cdot\Delta\boldsymbol{r} + \ldots = \Delta U + \ldots,$$

where the relative contribution of the terms $\ldots$ vanishes as $\Delta\boldsymbol{r}\to 0$. In other words, if the force is potential, one has

$$A_{12} = \int_{\gamma_{12}} \boldsymbol{F}\cdot d\boldsymbol{r} = \int_{\gamma_{12}} dU(\boldsymbol{r}) = U(\boldsymbol{r}_1) - U(\boldsymbol{r}_2),$$

regardless of the path γ_{12} connecting $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$. What follows from (42) immediately then is the *law of conservation of mechanical energy*, generalising (32):

$$E = K(v) + U(\boldsymbol{r}) = const, \tag{43}$$

for systems subject to the action of potential forces only potential forces only. If there are other non-potential forces denoted as $\boldsymbol{F}_{np}$ (they can depend on $\boldsymbol{r}, \boldsymbol{v}$ and t) one can only claim that

$$\Delta E = \Delta[K(v) + U(\boldsymbol{r})] = \int_{\gamma_{12}} \boldsymbol{F}_{np} \cdot d\boldsymbol{r} = A_{np}, \tag{44}$$

the term in the right-hand side being work done by non-potential forces. *The change of total energy equals the work done by non-potential forces.* The latter ones are friction, drag, etc., and are commonly physically associated with energy dissipating away via heat. (Slightly more generally, see below, potential and non-potential forces are referred to as conservative and dissipative, respectively – those that conserve energy and those that don't ... the distinction between potential and conservative forces are way beyond this course.)

Fortunately, the fundamental physical interaction forces are either potential or *gyroscopic*. The latter are forces that always act perpendicular to the velocity and therefore produce no work. An important example of gyroscopic forces is the Lorentz force acting on charged particles moving in magnetic field.

Let us find potentials U, corresponding to some forces. E.g., if we have a constant downward force $\boldsymbol{F} = -mg\boldsymbol{k}$, then $U = mgz$. Indeed, $-\nabla U = 0\boldsymbol{i} + 0\boldsymbol{j} - mg\boldsymbol{k}$. Another important example of a potential force is a *central force*

$$\boldsymbol{F}(\boldsymbol{r}) = f(r)\hat{\boldsymbol{r}}, \qquad \text{where } r = |\boldsymbol{r}|, \;\; \hat{\boldsymbol{r}} = \frac{\boldsymbol{r}}{r}, \tag{45}$$

acting at a given point along a line connecting the point with the origin. In this case

$$U = U(r) = -\int f(r)dr, \tag{46}$$

i.e. all one needs to do to find U is take the antiderivative of $f(r)$ as a function of one variable. To verify this, use the chain rule: take $U = U(r) = U(\sqrt{\boldsymbol{r} \cdot \boldsymbol{r}})$ and compute its partial derivatives. E.g.

$$\frac{\partial U}{\partial x} = U'(r)\frac{\partial\sqrt{x^2+y^2+z^2}}{\partial x} = U'(r)\frac{x}{r},$$

similar for $\frac{\partial U}{\partial y}$, $\frac{\partial U}{\partial z}$. Thus

$$\nabla U = U'(r)\,\frac{x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}}{r} = U'(r)\hat{\boldsymbol{r}}.$$

Another way of verifying (46) is simply noticing that similar to (40), we have $\boldsymbol{r} \cdot d\boldsymbol{r} = rdr$ – this merely is a vector identity, see the figure preceding (40). Hence,

$$dA = f(r)\frac{\boldsymbol{r} \cdot d\boldsymbol{r}}{r} = f(r)dr, \;\; \text{and } U(r) = -\int f(r)dr,$$

as we are now dealing with a function of a single variable r, whose integral between r_1 and r_2 is well defined.

For instance, for the gravity force

$$\boldsymbol{F} = -G\frac{Mm}{r^2}\hat{\boldsymbol{r}}$$

of attraction acting on a body of mass m and radius-vector $\boldsymbol{r}$ by a body of mass M, positioned at the origin (the minus sign shows that the force is directed towards rather than away from the origin), we find

$$U(r) = -G\frac{Mm}{r} + C.$$

The constant is insignificant and can be assigned zero value. Hence, if the body with the mass m is being brought by gravity form infinity, where $U = 0$, its potential energy decreases with the decrease of r, and the kinetic energy grows. Conversely, if M pertains to the Earth and m to a rocket, and one asks what launch velocity v_0 the rocket needs to be able to overcome the attraction of the Earth, the answer, by the law of conservation of energy, is

$$\frac{mv_0^2}{2} - G\frac{Mm}{r_0} \geq 0,$$

where r_0 is the radius of the Earth. Indeed, the left-hand side is the rocket's energy at launch. The right-hand side is its energy far away from the Earth, so that $U = 0$ assuming the rocket would have exhausted all its initial momentum to fight against gravity.

Mathematically, one can determine whether a given force, or *force field* – this is the term used to denominate a force $\boldsymbol{F} = \boldsymbol{F}(\boldsymbol{r})$ that depends on the position only – is potential is by performing integration. E.g. let $\boldsymbol{F} = y\boldsymbol{i} + x\boldsymbol{j}$. It is potential. Indeed, one must have $y = -\frac{\partial U}{\partial x}$, so $U(x, y) = -xy + C(y)$. On the other hand, one must have

$$x = -\frac{\partial U}{\partial y} = \frac{\partial(xy - C(y))}{\partial y}, \quad \text{so } C'(y) = 0, \;\; C = const, \;\; U = -xy + C.$$

However, the force field $\boldsymbol{F} = y\boldsymbol{i} - x\boldsymbol{j}$ is not potential. Indeed, the same reasoning leads to the conclusion that $U(x, y) = -xy + C(y)$ and $\frac{\partial U}{\partial y} = x$, which is inconsistent, as C cannot depend on x.

Another way of establishing that a given force field $\boldsymbol{F}$ is not potential is producing two different paths with the same endpoints, so that the work of $\boldsymbol{F}$ along these paths is not the same.

Observe that claiming for a force field $\boldsymbol{F}(\boldsymbol{r})$ that its work along any path depends only on the endpoints of the path is equivalent to a claim that the work of $\boldsymbol{F}$ along any closed path is zero. Indeed, let γ_1 and γ_2 be two different curves (paths) with the same endpoints P, Q, let γ be a closed path which consists in going from P to Q along γ_1 and then returning to P along γ_2. An integral along a closed path is often denoted as $\oint$, and the work of $\boldsymbol{F}$ along the close path γ equals

$$\oint_\gamma \boldsymbol{F} \cdot d\boldsymbol{r} = \int_{\gamma_1} \boldsymbol{F} \cdot d\boldsymbol{r} - \int_{\gamma_2} \boldsymbol{F} \cdot d\boldsymbol{r} = 0.$$

To generalise the notion of potential forces, a force field is called *conservative* if its work along any closed path is zero, or equivalently for any points P, Q the work done along a path connecting P and Q does not depend on the particular path. If a force field is potential, it is conservative. The converse is true in the Euclidean space of any dimension (so within this course *conservative* and*potential* is the same thing) but is not necessarily true if the domain of particle's possible motions has more complicated geometry.[7]

Non-conservative forces are friction, drag, etc. A force is called *dissipative* if it results in the loss of kinetic energy. Thus the dynamic friction force, as well as drag force are dissipative. Static friction is not dissipative, as obviously bodies can be accelerated due to static friction.

4.4.1 Additivity of energy

Kinetic energy is an additive quantity: the kinetic energy of a system of particles equals the sum of kinetic energies of individual particles. This follows as the integration of the Newton equation performed in the preceding section can be done for each particle individually. However, unlike the momentum, which is proportional to $\boldsymbol{v}$, the kinetic energy is proportional to v^2, and so the work done by internal forces inside a system of particles is usually not zero and does result in the change of the system's kinetic energy. The reason is that as far as the work done by the jth force over ith particle $dA_{ji} = \boldsymbol{F}_j \cdot d\boldsymbol{r}_i$ in concerned, having $\boldsymbol{F}_{j_1} = -\boldsymbol{F}_{j_2}$ for a pair of forces (action equals reaction) does not mean that the net work done by the two is zero, due to different $d\boldsymbol{r}_i$ that each individual force is being dot-multiplied by. E.g., suppose two electrons are repelled from each other from rest by the electric force. Their speeds, and hence the kinetic energy of the system increase, while the total momentum remains zero.

In a system of N particles, suppose $\boldsymbol{v}_i'$ is the ith particle's velocity in a frame K' moving with the velocity $\boldsymbol{V}$ with respect to the origin, then $\boldsymbol{v}_i = \boldsymbol{v}_i' + \boldsymbol{V}$. So

$$K_i = \frac{m_i}{2}(v_i'^2 + 2\boldsymbol{v}_i' \cdot \boldsymbol{V} + V^2).$$

[7] E.g. in the Euclidean plane take $\boldsymbol{F} = \frac{y}{r^2}\boldsymbol{i} - \frac{x}{r^2}\boldsymbol{j}$, where $r^2 = x^2 + y^2$. It is easy to see that the work done by $\boldsymbol{F}$ along a path connecting two points P and Q equals the angle (in radians) POQ. Thus is a closed path from P to P does not contain the origin O, the work along such a path is always zero, while it is $\pm 2\pi$ each time the path goes around the origin. So the vector field is not potential. Let us now cut the Euclidean plane along the non-negative x-axis or any other line beginning at the origin and going to infinity. Then paths going around the origin are now forbidden, so the vector field, restricted to the new domain ($\mathbb{R}^2$ minus the cut) is conservative, but it is still not potential.

If K' is associated with the mass centre of the system, then $\sum_i m_i \boldsymbol{v}'_i = 0$, and so the total kinetic energy

$$K = \sum_i K_i = \sum_i \frac{m_i {v'_i}^2}{2} + \frac{V^2}{2}\sum_i m_i = K' + \frac{MV^2}{2}.$$

This statement is called the König theorem: the kinetic energy of a system of particles equals their kinetic energy in the mass centre frame (where the system rests as whole) plus the kinetic energy of the mass centre as a single particle of mass $M = \sum m_i$ moving with the velocity $\boldsymbol{V}$. Recall that

$$\boldsymbol{V} = \frac{\sum_i m_i \boldsymbol{v}_i}{M}.$$

4.4.2 Collisions of particles

Collisions (or scattering) of particles represent obviously important applications of the mechanical theory partially developed above. Collisions theory gets quite involved if the particles are allowed to move beyond a single direction, and especially if they are allowed to have non-point sizes, when it becomes meaningful to allow them to rotate. The general idea of dealing with collisions is avoiding the analysis of complex and short-lived electromagnetic processes occurring at the moment of the collision *per se* and rather using the laws of conservation of momentum and energy (if applicable) to describe the system in terms of its state "before" and "after" the collision.

E.g. suppose a particle of mass m_1 is moving with the velocity v_1 along a line and hits another particle of mass m_2, whereupon the particles continue to move together. The total momentum of the system is preserved, while the energy is definitely not preserved. To see this we need a calculation, which becomes shortest if done in the mass centre frame. Regardless of the collision, the mass centre moves with the velocity $v = m_1 v_1/(m_1 + m_2)$, due to the conservation of momentum. By the König theorem above the total kinetic energy of the system is the sum of kinetic energies of the particles relative to the mass centre, plus the kinetic energy of the mass centre itself. The latter stays the same: the mass centre keeps moving with the same velocity v. However, after the collision the particles no longer move relative to the mass centre. But before the collision the first and second particle were moving with respect to the mass centre with the velocities, respectively

$$v'_1 = v_1 - v \quad \text{and} \quad -v.$$

Hence, the loss of kinetic energy (verify the calculation!) is

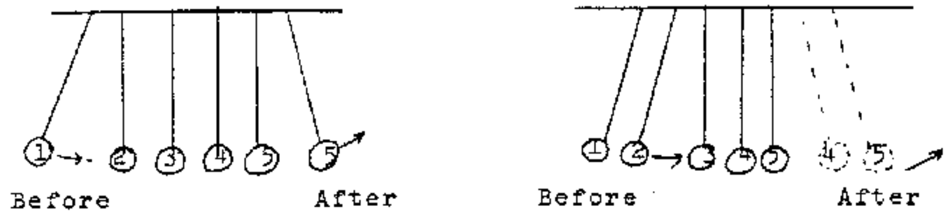
$$\Delta K = \frac{m_1(v_1 - v)^2}{2} + \frac{m_2 v^2}{2} = \frac{\mu v_1^2}{2}, \quad \text{where} \quad \mu = \frac{m_1 m_2}{m_1 + m_2}.$$

Note that if the particles have the same mass m, $\mu = m/2$, so half of the kinetic energy gets lost. If $m_2 \gg m_1$ then $\mu \approx m_1$, so almost all energy is lost.

Such a collision, namely when after the collision all the kinetic energy relative to the mass centre is lost, is called absolutely non-elastic. *Mechanical* energy then is definitely not conserved. The reason is that the kinetic energy of the particles at the moment when they collide gets transferred into their *internal* energy, resulting in the increase of their temperature, which represents the inner average motions of the molecules forming the particles. Analysing temperature is beyond the expertise of Classical mechanics. If one takes the change in the internal energy before and after the collision into account, the *total physical* energy of the system is, of course, preserved. The law of conservation of total energy is perhaps *the* fundamental law of physics. Einstein's special relativity shows that energy and momentum are, in fact, components of a single four-dimensional vector.

A collision of two mechanical bodies is called *elastic* if mechanical energy before and after the collision is preserved. (In Classical mechanics elastic collisions are an idealisation, yet in relativistic particles' mechanics they are omnipresent, as it does not make sense to talk about a particle's temperature, and all the particle's physical energy is contained in its velocity and mass). E.g. consider an elastic collision of two particles with equal masses m and velocities $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. In the mass center frame moving with some velocity $\boldsymbol{v}$, where the velocities are denoted by $\boldsymbol{u}$'s, the particles before the collision move with velocities $\boldsymbol{u}_1 = \boldsymbol{v}_1 - \boldsymbol{v}$ and $\boldsymbol{u}_2 = \boldsymbol{v}_2 - \boldsymbol{v}$, so that $\boldsymbol{u}_1 = -\boldsymbol{u}_2$. Thus the problem is therefore one-dimensional, and the vector notations for $\boldsymbol{u}$'s will further be dropped. After the collision the total momentum in the mass centre frame is still zero, so the "new" velocities u'_1 and u'_2 still satisfy $u'_1 = -u'_2$. In addition, total energy is preserved, so ${u'_1}^2 + {u'_2}^2 = u_1^2 + u_2^2$. There are two possibilities then: either

$u_1' = u_1$ and $u_2' = u_2$ or $u_1' = -u_1$ and $u_2' = -u_2$. Physically, for obvious reasons (the left particle remains on the left of the right particle) the former possibility is unacceptable. Therefore, after an elastic collision of two particles of equal mass, they simply exchange their velocities: $\boldsymbol{v}_1' = \boldsymbol{v}_2,\ \boldsymbol{v}_2' = \boldsymbol{v}_1$. (The other way of seeing it is that two particles of the same mass after the elastic collision "move through each other but exchange identities". Here is a standard cartoon illustrating elastic collisions.



4.4.3 Potential energy of interaction

Considering systems of particles, it is a common physical reality that interaction (e.g. gravitational, electrostatic) forces between particles are such that they depend only on the distances between particles and act along the lines connecting pairs of particles. If this is the case, it is possible to introduce the potential energy of interaction and therefore use conservation of energy to describe such systems of particles.

To see this, consider a system of N particles, and suppose the jth particle imposes the force $\boldsymbol{F}_{ij}$ onto the ith particle (thus $i \neq j$), so that

$$\boldsymbol{F}_{ij} = F_{ij}(r_{ij})\hat{\boldsymbol{r}}_{ij},$$

where $r_{ij} > 0$ is the distance between the particles, and $\hat{\boldsymbol{r}}_{ij}$ is the unit vector in the direction of $\boldsymbol{r}_{ij} = \boldsymbol{r}_i - \boldsymbol{r}_j$, i.e. pointing from the jth particle in the direction of the ith one. So, $\hat{\boldsymbol{r}}_{ij} = \frac{\boldsymbol{r}_{ij}}{r_{ij}}$. If $K_i = \frac{m_i v_i^2}{2}$, by (39, 40) we have, for each particle

$$dK_i = \sum_{j \neq i} \boldsymbol{F}_{ij} \cdot d\boldsymbol{r}_i, \qquad i = 1, \dots, N.$$

We now sum over all i, introducing the total kinetic energy $K = \sum_i \frac{m_i v_i^2}{2}$ and use the third Newton's law stating that $\boldsymbol{F}_{ij} = -\boldsymbol{F}_{ji}$:

$$dK = \sum_{j>i} \boldsymbol{F}_{ij} \cdot d(\boldsymbol{r}_i - \boldsymbol{r}_j) = \sum_{j>i} F_{ij}(r_{ij}) \frac{\boldsymbol{r}_{ij} \cdot d\boldsymbol{r}_{ij}}{r_{ij}} = \sum_{j>i} F_{ij}(r_{ij}) dr_{ij}.$$

Indeed, the last equality follows from the vector identity $\boldsymbol{r}_{ij} \cdot d\boldsymbol{r}_{ij} = r_{ij} dr_{ij}$ – in the past we've had exactly the same identity for $\boldsymbol{v}$, see (39, 40) as well as for $\boldsymbol{r}$ dealing with central forces, see (46).

Now the variables have separated, each term in the left-hand side depends on v_i only, each term in the right-hand side on r_{ij} only. Therefore, defining

$$U_{ij}(r_{ij}) = -\int F_{ij}(r_{ij}) dr_{ij}, \qquad U = \sum_{j>i} U_{ij},$$

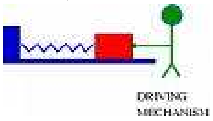
after integrating term by term independently, we conclude that

$$E = K + U = const.$$

E.g. suppose three electrons are initially positioned at rest at three corners of the equilateral triangle with side r_0 and then let fly under the electric repulsion force. In this case $F_{ij} = k/r_{ij}^2$, for some k proportional to e^2. (This is a repulsion force, so it is indeed directed along $\boldsymbol{r}_i - \boldsymbol{r}_j$, not opposite; a good somewhat heuristic rule here is that attraction forces bear a minus sign, while repulsion forces do not.) Hence $U_{ij} = k/r_{ij}$, and $E = 3k/r_0$, as there are three pairs of interacting electrons. After the electrons fly far enough, $r_{ij} \to \infty$, so in the limit $U = 0$, and the whole of E has been recycled into the electrons' kinetic energy, which is equal $3m\frac{v^2}{2}$ – there are three electrons. So, the limit speed of each electron at infinity will be $v_\infty = \sqrt{\frac{2k}{mr_0}}$.

4.5 Harmonic oscillations

An important example when the Newton equation can be fully solved in quadratures is the *harmonic oscillator*. Consider, for instance, a mass m attached to a spring with the spring constant k on a horizontal table.



On this picture, in addition to the spring there is an external driving force to be considered further, but for now assume it is zero. However, suppose the motion occurs in a medium (air or liquid), so there is a dissipative damping force, proportional to the speed with the coefficient κ and acting in the direction against it.

4.5.1 Undriven oscillator

In the lack of external driving force the Second law is

$$m\ddot{x} = -kx - \kappa\dot{x}, \tag{47}$$

where x is the deformation of the spring from equilibrium. Equivalently

$$\ddot{x} + 2\gamma\dot{x} + \omega^2 x = 0, \tag{48}$$

where $\gamma = \kappa/2m$, $\omega = \sqrt{k/m}$.

This is a homogeneous linear ODE, and we seek its solution as $x(t) = Ce^{\lambda t}$, for some λ and any C.[8] From linearity and homogeneity of the equation we can, in fact, seek complex, rather than just real λ. This would result naturally in complex x. Then, by linearity of the equation, if the complex $x(t) = \Re x(t) + i\Im x(t)$ satisfies the equation, then both its real and imaginary parts $\Re x(t)$ and $\Im x(t)$ are real solutions of the equation.

Substitution of the above $x(t)$ into the equation yields

$$\lambda^2 + 2\gamma\lambda + \omega^2 = 0, \quad \text{so} \quad \lambda = -\gamma \pm i\sqrt{\omega^2 - \gamma^2}.$$

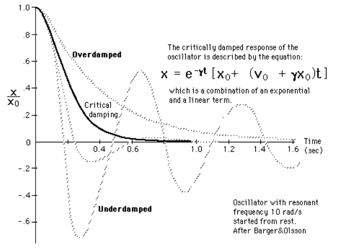
We therefore get two linearly independent solutions for differen choices of the $\pm$ sign, unless $\omega = \gamma$. In the latter case, one solution $x(t) = C_1 e^{-\gamma t}$ has been found. The other solution (see the footnote) can only have the form $C_2 t e^{-\gamma t}$, and it is easy to verify that it satisfies the equation (48) if and only if $\gamma = \omega$. (Verify this and do not neglect the Leibniz formula for the second derivative of the product: $(uv)'' = uv'' + 2u'v' + vu''$.)

So, denoting $\tilde{\omega} = \sqrt{|\omega^2 - \gamma^2|}$ the solutions in general are

$$x(t) = \begin{cases} C_1 e^{-(\gamma-\tilde{\omega})t} + C_2 e^{-(\gamma+\tilde{\omega})t}, & \text{if } \omega < \gamma, \\ (C_1 + C_2 t)e^{-\gamma t}, & \text{if } \omega = \gamma, \\ C_1 e^{-(\gamma-i\tilde{\omega})t} + C_2 e^{-(\gamma+i\tilde{\omega})t}, & \text{if } \omega > \gamma. \end{cases} \tag{49}$$

The case $\gamma > \omega$ is called *overdamped*, the case $\gamma = \omega$ *critical* and the case $\gamma < \omega$ *underdamped*, or normal. The constants above are determined by the initial conditions $x(0) = x_0$ and $\dot{x}(0) = v_0$, and can be both zero, which corresponds to no motion at all.

[8] The reason it is sought in this form is that the coefficients in the equation are time-independent. Thus if $x(t)$ is a solution of this equation, then for any T, $x(t+T)$ must also be a solution (because differentiating $x(t)$ with respect to either t or $t' = t+T$ yields the same result.) The exponential function has the desired property: $x(t+T) = Ce^{\lambda T}e^{\lambda t} = C'e^{\lambda t}$. And so does a more general solution $x(t) = C_1 e^{\lambda t} + C_2 t e^{\lambda t}$, which kicks in in the critical damping case below, but no other.



In the underdamped case to be discussed further in more detail the real solution can be extracted as

$$x(t) = e^{-\gamma t}[A\cos\tilde{\omega}t + B\sin\tilde{\omega}t] = Ce^{-\gamma t}\cos(\tilde{\omega}t - \phi), \tag{50}$$

where

$$C = \sqrt{A^2 + B^2}, \qquad \phi = \arctan(B/A).$$

Observe that the later is just a version of Pythagoras' theorem, after writing

$$A\cos\tilde{\omega}t + B\sin\tilde{\omega}t = C\left(\frac{A}{C}\cos\tilde{\omega}t + \frac{B}{C}\sin\tilde{\omega}t\right)$$

and using the formula for $\cos(\alpha - \phi)$, with $\alpha = \tilde{\omega}t$ and ϕ as above.

Note that whatever the relation between γ and ω, as long as $\gamma > 0$ damping makes $x(t)$ vanish exponentially. If $\gamma = 0$, then (50) is still valid, with $\tilde{\omega} = \omega$, and represents *simple harmonic motion* with frequency ω. In general, (50) is referred to as *damped oscillations*, with frequency $\tilde{\omega}$ and *logarithmic decrement* γ. The motivation for the latter terminology is that the time-separation between the nearby maxima of the solution in (50) equals $T = 2\pi/\tilde{\omega}$. Over this time the pre-factor $e^{-\gamma t}$ becomes $e^{-\gamma T}$ smaller. In other words, the ratio

$$\log\frac{x(t)}{x(t+T)} = \gamma.$$

4.5.2 Driven harmonic oscillator

Consider now the *forced*, or driven oscillator, namely in the Newton equation (47) let us add some extra time-dependent driving force $F = F(t)$ (the man in the above cartoon gets to work). An interesting case is when the force is constant or periodic. Let $f = F/m$, the ODE we are dealing with is now

$$\ddot{x} + 2\gamma\dot{x} + \omega^2 x = f(t), \tag{51}$$

and its general solution is the sum of the *transient* solution from (49), solving the homogeneous equation (48) and *any* particular solution $x_p(t)$ solving the whole equation (51)[9].

Let us further consider only the case of (damped) oscillations when $\omega > \gamma \geq 0$. So, the goal is to find a single particular $x_p(t)$ solution to (51). Observe that as

$$x(t) = x_p(t) + Ce^{-\gamma t}\cos(\tilde{\omega}t - \phi), \tag{52}$$

for $\gamma > 0$ the initial conditions which determine C and ϕ will have almost no effect on $x(t)$, for large t, due to the exponential decay $e^{-\gamma t}$. Thus $x_p(t) \approx x(t)$ for large t is called *steady state solution.*

Furthermore, note that if $f(t) = f_1(t) + f_2(t) + \ldots$, then, in fact,

$$x_p(t) = x_{p,1}(t) + x_{p,2}(t) + \ldots,$$

where each $x_{p,i}(t)$ is obtained as if the term $f_i(t)$ were the only one in the right-hand side. This is the reflection of the linearity of the equation which makes the forces' action *independent.*

[9]Indeed, by linearity of (51), if $x(t)$ is *any* solution of it and $x_p(t)$ *some* solution, the difference $x(t) - x_p(t)$ must satisfy the homogeneous equation.

The simplest $f(t)$ is the constant force $f = const$. This corresponds to the spring whose Newton's equation was (47) hanging vertically rather than horizontally, when gravity comes into play, in which case $f = g$. The particular solution then is simply

$$x_p(t) = \frac{g}{\omega^2} = \frac{mg}{k} = x_0.$$

Physically, it corresponds to the equilibrium position of the mass hanging vertically on the spring, stretched by x_0, with $kx_0 = mg$. In other words, the only impact of gravity is that the (damped) oscillations (51) now pertain to the deviation from equilibrium $x(t) - x_0$, rather than from the unstretched spring.

Now let us consider a periodic $f(t)$, i.e $f(t) = f(t+T)$ for some minimum period $T > 0$. A periodic function with the period T is given by a Fourier series with frequencies $n\Omega$, where $\Omega = 2\pi/T$. However, as we have shown, each term in the series will have an independent effect, and therefore it suffices to consider only

$$f(t) = f_0 \cos \Omega t \quad \text{or} \quad f(t) = f_0 \sin \Omega t.$$

To account for both sine and cosine, it is convenient to deal with

$$f(t) = f_0 e^{i\Omega t}$$

instead. Then, after the complex $x_p(t)$ has been found, its real part $\Re x_p(t)$ will describe the response to the cosine forcing and its imaginary part $\Im x_p(t)$ – to the sine forcing.

Let us try the solution $x_p(t) = Ke^{i\Omega t}$ for some complex K. Substituting it into (51) and canceling the exponential (never zero!) out, we have:

$$K(-\Omega^2 + 2i\gamma\Omega + \omega^2) = f_0.$$

This works, unless simultaneously $\gamma = 0$ and $\Omega = \omega$. The latter scenario is called *full resonance* and if it is the case, our ansatz $x_p(t) = Ke^{i\Omega t}$ was wrong. Let us deal with this situation first, so $\gamma = 0$ and $\Omega = \omega$; now try $x_p(t) = Kte^{i\omega t}$. Substitution into

$$\ddot{x} + \omega^2 x = f_0 e^{i\omega t}$$

yields, after canceling the exponential

$$K(-t\omega^2 + 2i\omega + t\omega^2) = f_0,$$

so $K = -i\frac{f_0}{2\omega}$. Thus, the oscillator's resonant response to

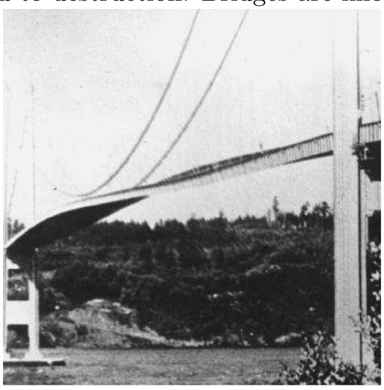
$$f = f_0(\cos \omega t + i \sin \omega t)$$

is

$$-it\frac{f_0}{2\omega}\cos \omega t + \frac{f_0}{2\omega} t \sin \omega t.$$

In particular, if $f = f_0 \cos \omega t$, the response is $x_p(t) = \frac{f_0}{2\omega} t \sin \omega t$; if $f = f_0 \sin \omega t$, the response is $x_p(t) = -\frac{f_0}{2\omega} t \cos \omega t$.

In both cases, it is proportional to t and is therefore unbounded as $t \to \infty$. Physically, unbounded solutions lead to destruction. Bridges are known to be sensitive to resonances.



For the future, suppose no full resonance occurs, for which it suffices to have either $\gamma > 0$, or $\Omega \neq \omega$. Then we have the complex solution

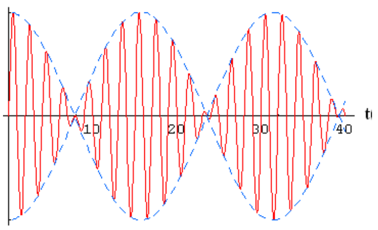
$$x_p(t) = \frac{f_0}{(-\Omega^2 + \omega^2) + 2i\gamma\Omega} e^{i\Omega t}$$

It makes sense to write the pre-exponential factor K as $K = Ae^{i\psi}$, where A is the modulus of the complex number K and ψ the phase. Observe that for any complex $z \neq 0$, the modulus of $1/z$ is one over the modulus of z, while the phase of $1/z$ is minus the phase of z. Thus

$$\begin{aligned} K &= \frac{f_0}{\sqrt{(\omega^2-\Omega^2)^2+4\Omega^2\gamma^2}} e^{-i \arctan \frac{2\gamma\Omega}{\omega^2-\Omega^2}}, \\ x_p(t) &= \frac{f_0}{\sqrt{(\omega^2-\Omega^2)^2+4\Omega^2\gamma^2}} e^{i\left(\Omega t - \arctan \frac{2\gamma\Omega}{\omega^2-\Omega^2}\right)}. \end{aligned} \tag{53}$$

Observe that the response to the cosine driving force will contain the real part of the complex exponential, i.e. the cosine of $\Omega t - \arctan \frac{2\gamma\Omega}{\omega^2-\Omega^2}$; the response to the sine driving force will be the sine of the above.

By (52) in the full solution the transient solution can be dropped, as long as $\gamma > 0$. Otherwise, if $\gamma = 0$ and $\Omega \neq \omega$, $x(t)$ represents the superposition of two sine curves, one with frequency ω and the other with frequency Ω. A typical countenance for such $x(t)$ is as sequence of "beats", whose maximum amplitude grows as $\Omega \to \omega$. The envelope of the beats has frequency $\frac{1}{2}|\omega - \Omega|$, providing "amplitude modulation" of "fast" oscillations with frequency $\frac{1}{2}(\omega + \Omega)$. This follows basically from the formula $\cos \omega t + \cos \Omega t = 2 \cos \frac{|\omega-\Omega|t}{2} \cos \frac{|\omega+\Omega|t}{2}$.



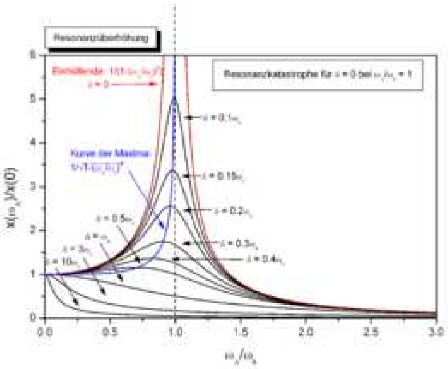
Assuming $\gamma \neq 0$, it makes sense to ask for which Ω does the amplitude K of the steady state solution attain its maximum. To do this is tantamount to minimizing the square of the denominator $G(y) = (y - \omega^2)^2 + 4y\gamma^2$, with respect to $y = \Omega^2$. We have

$$G'(y) = 2(y - \omega^2) + 4\gamma^2,$$

and zero if $y = \omega^2 - 2\gamma^2$, which, considering that $y > 0$, occurs only if $\omega > \sqrt{2}\gamma$. Therefore, the steady state solution will have maximum amplitude provided that

$$\Omega^2 = \omega^2 - 2\gamma^2.$$

If $\omega^2 < 2\gamma^2$, then the above function $G(y)$ is increasing for $y > 0$, and therefore the amplitude K of the steady state solution will be a decreasing function of Ω. Different relations $K(\Omega)$, for different values of γ can be qualitatively seen on the following diagram (never mind the German).



4.6 Angular momentum

4.6.1 Constants of motion

The mathematical aspects of mechanics grow increasingly harder as the number of degrees of freedom of a mechanical system increases. As far as systems of one degree of freedom were concerned, in Section 4.2 it was shown that under rather general circumstances one can introduce the notion of potential and then use energy conservation to completely solve (in other words, integrate) the Newton equations of motion of a "one-dimensional" particle. Further in Section 4.4 it turned out that when the number of degrees of freedom exceeds one, the class of forces when the law of conservation of mechanical energy would apply had to be narrowed to potential forces only. Still, the key physical interaction forces are potential.

In the case of one degree of freedom, the mere fact of energy conservation resulted in equation (33), which is the first-order differential equation. A single first-order differential equation is fairly easy to solve, or integrate.

When the number of degrees of freedom equals, for instance two, the analogue of (33) is

$$\dot{x}^2 + \dot{y}^2 = \frac{2}{m}(E - U(x,y)) \qquad \Leftrightarrow \qquad E = \frac{m}{2}(\dot{x}^2 + \dot{y}^2) + U(x,y) = const. \tag{54}$$

This is no longer a single ordinary differential equation. Yet its advantage over the Newton equations is that (54) contains only the first time-derivatives of $\boldsymbol{r} = (x, y)$. On the other hand, it is a single relation, which obviously does not suffice to obtain the two unknown functions $x(t)$ and $y(t)$.

In order to solve the Newton equations, i.e. find the unknown functions $x(t), y(t)$ satisfying given initial conditions, one may try to obtain another relation in the form

$$F(\dot{x}, \dot{y}, x, y) = C = const. \tag{55}$$

Such a quantity F, which is in general a function of position and velocity (and sometimes even time) is called a *constant*, or *integral of motion*. Heuristically, the integral of motion F (a particular case is the energy E) is a function that depends on $\boldsymbol{r}, \dot{\boldsymbol{r}}$, yet being evaluated along the trajectory $\boldsymbol{r}(t)$ of a particle, it always returns the same value equal to C. (Geometrically this means that the trajectory viewed as a phase curve lies within a single level set of the function F). In particular, knowing the initial conditions $\boldsymbol{r}, \dot{\boldsymbol{r}}$ at $t = 0$ would suffice to fix the value of F once and for all.

If the number of degrees of freedom equals two, having another constant of motion (55) independent of energy E would give another relation binding $\dot{x}, \dot{y}$ and x, y. Together with (54) they would constitute now a system of two *first-order* ODEs, which in general shall be easier to solve than the Newton equations, which comprise a system of two *second-order* ODEs.

Today, the method of constants of motions is the most common approach to studying the mathematical aspects of mechanics. In order to identify constants of motion, one may have to use coordinate systems other than the Cartesian one and are dictated by the system's geometry. We shall further introduce the concept of *angular momentum*, as one of the most important application of the method of constants of motion and completely integrate the problem of motion of a single particle in central force field by using polar coordinates.

4.6.2 The rate of change of angular momentum equals torque.

Let us start out in full generality. Consider a particle moving in three dimensions, whose evolution is governed by the Newton equation

$$m\ddot{\boldsymbol{r}} = \boldsymbol{F}.$$

Let us take the cross, i.e. vector product of both sides with $\boldsymbol{r}$, multiplying on the left (recall that in general $\boldsymbol{a} \times \boldsymbol{b} = -\boldsymbol{b} \times \boldsymbol{a}$ for two vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^3$. We get

$$m(\boldsymbol{r} \times \ddot{\boldsymbol{r}}) = \boldsymbol{r} \times \boldsymbol{F}.$$

The right-hand side is called *torque* generated by force $\boldsymbol{F}$ (in this case $\boldsymbol{F}$ is the net force acting on the particle, so we have the net torque, but if $\boldsymbol{F} = \sum \boldsymbol{F}_i$ is a sum of individual forces, it makes sense to define torque generated by each force as $\boldsymbol{r} \times \boldsymbol{F}_i$.) The left-hand side equals

$$\frac{d}{dt}(m\boldsymbol{r} \times \dot{\boldsymbol{r}}) \equiv \frac{d}{dt}\boldsymbol{L}, \tag{56}$$

thus defining the quantity

$$\boldsymbol{L} = m\boldsymbol{r} \times \dot{\boldsymbol{r}},$$

called the particle's *angular momentum*. Indeed,

$$\dot{\boldsymbol{L}} = m\frac{d}{dt}(\boldsymbol{r} \times \dot{\boldsymbol{r}}) = m(\dot{\boldsymbol{r}} \times \dot{\boldsymbol{r}}) + m\boldsymbol{r} \times \ddot{\boldsymbol{r}} = 0 + m\boldsymbol{r} \times \ddot{\boldsymbol{r}},$$

because the cross product of any vector with itself equals zero. The quantity $\boldsymbol{L}$ is trivially zero when the vectors $\boldsymbol{r}$ and $\dot{\boldsymbol{r}}$ are collinear. A heuristic way of thinking of $\boldsymbol{L}$ is that it gives a measure of how curved the particle's trajectory is or, for a system of particles, how the system rotates. Observe that the quantity $\boldsymbol{L}$ is additive: the total angular momentum of a system of particles equals the sum of individual angular momenta of each separate particle.

The formula (56) is just a paraphrase of the Newton law. However, it gives it an interesting twist, explaining, for instance, how a lever works: one can balance a weight M by a weight m on a lever, provided that $Ml = mL$, where l and L are the lengths of the arms of the lever, whereto the masses M and m are attached, respectively. Indeed, in order that the lever be in the state of equilibrium, its total angular momentum shall always remain zero, and thus the total torque, with respect to the origin which is at the *fulcrum*, or the pivot point, of the lever. Note that the directions of the (horizontal) radius-vectors toward the masses M and m are opposite, and therefore the torques obtained by taking cross-products of these vectors with downward vertical vectors Mg and mg also act in opposite directions.

Without consequences in this exposition, let us note that in fact, $\boldsymbol{L}$ being a cross product, its direction is determined by the right hand rule if the coordinate frame used is right (as is conventionally the case) or the left-hand rule if the coordinate frame axes orientation is left.

4.6.3 Law of conservation of angular momentum

The most interesting case regarding the angular momentum vector is when torque is zero. This occurs not only when $\boldsymbol{F} = 0$, but more generally when the vectors $\boldsymbol{F}$ and $\boldsymbol{r}$ are collinear. In particular, this happens when the force is *central*, i.e. its value depends only on the distance to a fixed origin and acts along the line connecting the particle and the origin, see the defining formula (45). Such forces are common in nature – gravity, electrostatics – and moreover they arise in two-particles' interactions, see Section 3.4.1.

If torque is zero, as will be assumed from now on, then one gets the *law of conservation of angular momentum*

$$\boldsymbol{L} = const,$$

and as $\boldsymbol{L}$ is a vector, this means there are *three* constants of motion:

$$L_x = m(y\dot{z} - z\dot{y}), \qquad L_y = m(z\dot{x} - x\dot{z}), \qquad L_z = m(x\dot{y} - y\dot{x}).$$

If one deals with a system of particles, all of the above is true with respect to the net angular momentum. Physically, an isolated system of many interacting particles, such as liquid or gas, is often *isotropic*, i.e. there is no particular reason for this system to behave any different after having been rotated as whole by any angle. This implies that the net torque in this system is zero, and thus the net angular momentum is conserved.

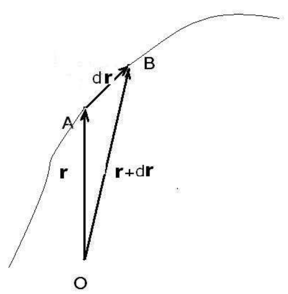
Returning to the case for a single particle moving in three dimensions, suppose its angular momentum is constant, which as has been mentioned, is the case when the force acting on the particle is central, which will be treated in some detail. $\boldsymbol{L}$ is a vector, hence if it is preserved, both its direction and magnitude are preserved. If $L = 0$, then this means that the vectors $\boldsymbol{r}$ and $\boldsymbol{v}$ are always collinear, and this is possible only if the particle is moving along a straight line passing through the origin. If the magnitude L of $\boldsymbol{L}$ is not zero, then it makes sense to speak of the direction of $\boldsymbol{L}$, which shall always be the same. But $\boldsymbol{L}$, as a cross product, is directed perpendicular to the plane, spanned by the vectors $\boldsymbol{r}$ and $\boldsymbol{v}$. This means that the vectors $\boldsymbol{r}$ and $\boldsymbol{v}$ always lie in the same plane. I.e. the motion of the particle is, in fact, two-dimensional, rather than three-dimensional!

Without loss of generality then one can direct coordinate axes in such a way that the particle's trajectory lies in the xy-plane. Then conservation of angular momentum means that

$$L = m(x\dot{y} - y\dot{x}) = const. \tag{57}$$

Let us look at the geometric meaning of the above quantity. We have

$$const. = \boldsymbol{r} \times \frac{d\boldsymbol{r}}{dt} = \boldsymbol{r} \times \boldsymbol{v}.$$



Geometrically, see the above figure, as the particle moves along a trajectory, one can talk about the area swept by its radius vector. In the figure, as the particle has moved from A to B over the time dt, its radius-vector $\boldsymbol{r}$ has swept the triangle OAB, whose area equals in absolute value $\frac{1}{2}|\boldsymbol{r} \times d\boldsymbol{r}| = \frac{1}{2}|\boldsymbol{r} \times \boldsymbol{v}|dt$. By the law of conservation of momentum, $|\boldsymbol{r} \times \boldsymbol{v}| = const.$ What follows is that as long as the angular momentum is preserved, the particle's radius vector will sweep equal areas over equal time intervals, the area swept over a unit time being equal to $L/2m$. This fact was discovered experimentally by Kepler (first published in his book *Astronomia nova* in 1609) and is today referred to as Kepler's second law of planetary motion. Indeed, with reasonable approximation precision planets can be thought of as moving independently in the central gravitational field of the stationary Sun (the relative error of such an approximation can be measured by the dimensionless quantity which is the net mass of the planets divided by the mass of the Sun, which is about .001). The First Kepler law, stating that the planets' orbits are ellipses has a more subtle explanation, owing to the fact that the force of gravitational attraction is proportional to the inverse square of the distance.

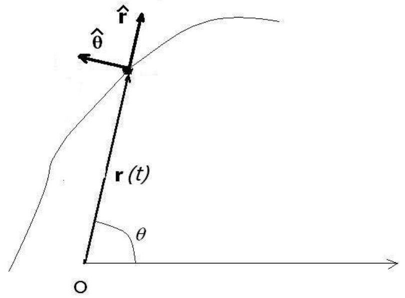
We will further consider the particle of mass m moving in the central force field, restating (45):

$$\boldsymbol{F}(\boldsymbol{r}) = f(r)\hat{\boldsymbol{r}}, \qquad \text{where } r = |\boldsymbol{r}|, \;\; \hat{\boldsymbol{r}} = \frac{\boldsymbol{r}}{r}, \tag{58}$$

By the law of conservation of angular momentum, we can assume that $\boldsymbol{r} = (x, y)$ is two-dimensional rather than three-dimensional.

4.7 Polar coordinates

In order to study Newton's equations with the right-hand side (58), it makes sense to introduce a coordinate system where $\hat{\boldsymbol{r}}$ would be a coordinate axis vector. Such a system, however, will be attached to the moving particle: indeed the quantity $\hat{\boldsymbol{r}}$ depends on time t. Let us call $\hat{\boldsymbol{\theta}}$ a unit vector perpendicular to $\hat{\boldsymbol{r}}$, so that $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{\theta}}$ are a right pair, just like the Cartesian unit vectors $\boldsymbol{i}$ and $\boldsymbol{j}$. Only, contrary to $\boldsymbol{i}$ and $\boldsymbol{j}$, the directions of $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{\theta}}$ change in time.



Let us look at the Cartesian coordinates of the vectors $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{\theta}}$. If θ is the angle that the particle's position forms with the x-axis, then

$$\hat{\boldsymbol{r}} = \cos\theta \boldsymbol{i} + \sin\theta \boldsymbol{j}, \qquad \hat{\boldsymbol{\theta}} = -\sin\theta \boldsymbol{i} + \cos\theta \boldsymbol{j} \tag{59}$$

The pair (r, θ), where $r \geq 0$ is the distance from the origin, and $\theta \in [0, 2\pi)$ (and is not defined at the origin) are called *polar coordinates* of a particle. Knowing these two numbers enables unambiguously defines the particle's position (strictly speaking, at the origin θ is not defined). Cartesian coordinates (x, y) of the particle are expressed via polar coordinates as

$$x = r\cos\theta, \qquad y = r\sin\theta. \tag{60}$$

Evolution of the particle will be understood fully if we can obtain the functions $r(t), \theta(t)$ as the particle moves.

In order to obtain the differential equations for the functions $r(t), \theta(t)$ we would like to use the conservation of *angular momentum and energy.* Both are constant, as the force is central, i.e it allows for the potential energy $U(r) = -\int f(r)dr$.

Let us look at the value L of the angular momentum: as a vector $\boldsymbol{L}$ is directed perpendicular to the xy plane where the particle moves.

We let us project the velocity $\boldsymbol{v}$ on the directions along the radius vector and perpendicular to it:

$$\boldsymbol{v} = \dot{r}\hat{\boldsymbol{r}} + r\dot{\theta}\hat{\boldsymbol{\theta}}. \tag{61}$$

This is just kinematics: the second term is linear velocity along the circle of the fixed radius r, the first one reflects the change of r itself. Since $\hat{\boldsymbol{r}}$ and $\boldsymbol{r}$ are collinear, than the first term dos not contribute to angular momentum, in other words,

$$L = mr^2\dot{\theta}. \tag{62}$$

To be more precise, L here is allowed to be positive or negative, depending on whether the motion is counterclockwise or clockwise, so L is, in fact, the *signed* value of the angular momentum vector $\boldsymbol{L}$.

The same formula can be obtained using (60) and the product and Chain rule to express $\dot{x}$ and $\dot{y}$. We have, by (57):

$$L = m(x\dot{y} - y\dot{x}) = m[r\cos\theta(\dot{r}\sin\theta + r\dot{\theta}\cos\theta) - r\sin\theta(\dot{r}\cos\theta - r\dot{\theta}\sin\theta)] = mr^2\dot{\theta}.$$

The quantity L shall be thought of as given – having the initial conditions $\boldsymbol{r}(0)$ and $\boldsymbol{v}(0)$ gives $\boldsymbol{L} = m\boldsymbol{r}(0) \times \boldsymbol{v}(0)$, and L is its absolute value. Hence, let us express

$$\dot{\theta} = \frac{L}{mr^2}. \tag{63}$$

We see that the angle θ evolves either counterclockwise, or clockwise, or is constant zero, depending on whether the constant angular momentum vector $\boldsymbol{L}$ points up or down along the z axis, or is zero. The quantity $\dot{\theta}$ is often denoted as ω, the angular velocity.

Let us also get the expression for energy: we know by (46) that a central force is potential, so the total energy

$$E = \frac{mv^2}{2} + U(r), \qquad U(r) = -\int f(r)dr$$

is constant. To express $v^2 = \boldsymbol{v} \cdot \boldsymbol{v}$ in polar coordinates we just use (61): the latter decomposes $\boldsymbol{v}$ as a sum of two mutually orthogonal vectors, so

$$v^2 = \dot{r}^2 + (r\omega)^2. \tag{64}$$

So, energy conservation is now

$$E = \frac{m}{2}(\dot{r}^2 + (r\omega)^2) + U(r) = const, \tag{65}$$

and just as the angular momentum value L, the value of energy E can be taken as given, because it can be determined by the initial conditions. Now, we can exclude $\omega = \dot{\theta}$ by (63) and obtain

$$E = \frac{m\dot{r}^2}{2} + \left(\frac{L^2}{2mr^2} + U(r)\right) = const. \tag{66}$$

This equation no longer contains θ and can be solved for $r(t)$ by the methods that have been designed for one degree of freedom systems in Section 4.2. Namely, the function $r(t)$ arises if one deals with a standard one-degree of freedom system, whose potential, rather than being just $U(r)$, equals

$$V(r) = U(r) + \frac{L^2}{2mr^2}.$$

$V(r)$ is called *effective potential* and is the sum of the actual potential $U(r)$ and another *centrifugal term* $\frac{L^2}{2mr^2}$ which reflects the presence of rotational motion, as long as $L \neq 0$. Observe that if $L \neq 0$, $V(0) = \infty$ and therefore for all t one has $r > 0$. If $L = 0$, then as we have mentioned before, the particle moves along the straight line passing through the origin, and the variable r can be simply replaced by x, i.e. one is dealing with one-dimensional particle moving along the x-axis.

Differentiating (66) with respect to time, we get

$$\dot{r}\left(m\ddot{r} - (f(r) + \frac{L^2}{mr^3})\right) = 0. \tag{67}$$

I.e. either $r(t) = const.$ or

$$m\ddot{r} = \tilde{f}(r), \qquad \text{with } \tilde{f}(r) = f(r) + \frac{L^2}{mr^3}. \tag{68}$$

The quantity $\tilde{f}$ is called *effective force* and equals the sum of the actual force $f(r)$ and the additional centrifugal term

$$\frac{L^2}{mr^3} = m\omega^2 r,$$

by (63).

The two equations (67) and (63) represent the sought for equations of motion, and their most important feature is that they are not coupled. Given the initial conditions, one determines the constants L, E, then solves (67) for $r(t)$, and then solves (63) for $\theta(t)$.

Observe that the case when r is such that $f(r) + \frac{L^2}{mr^3} = 0$, i.e the "real" force $f(r)$ is compensated by the centrifugal force $\frac{L^2}{mr^3}$ corresponds to solutions $r(t) = r_C = const.$, which means moving along a circle of fixed radius r_C with constant angular velocity given by (63). In other words, this means that in (**??**) the case $\dot{r}(t) = 0$ may only occur when the effective force is zero, i.e (**??**) is equivalent to (67).

If one is on the circular orbit, then on it $\omega_C = \dot{\theta} = \frac{L}{mr_C^2}$, so the period of revolution around the orbit is

$$T_C = 2\pi\frac{mr_C^2}{L}.$$

4.7.1 Direct derivation of the equations of motion (optional)

Alternatively, here is the direct derivation of the main equations (67) and (63) from Newton. Clearly, the particle's radius vector

$$\boldsymbol{r} = r\hat{\boldsymbol{r}}. \tag{69}$$

We would now like to differentiate the above expression twice, noticing that when we do so, both r and $\hat{\boldsymbol{r}}$ are functions of t. The result will be made equal to $\boldsymbol{F}$ in (58), by the Second Newton law.

By the Leibniz formula for the second derivative of the product $(fg)'' = f''g + 2f'g' + fg''$ we have

$$\ddot{\boldsymbol{r}} = \ddot{r}\hat{\boldsymbol{r}} + 2\dot{r}\dot{\hat{\boldsymbol{r}}} + r\ddot{\hat{\boldsymbol{r}}}. \tag{70}$$

Furthermore, by (59) and Chain rule

$$\dot{\hat{\boldsymbol{r}}} = \dot{\theta}(-\sin\theta\boldsymbol{i} + \cos\theta\boldsymbol{j}) = \dot{\theta}\hat{\boldsymbol{\theta}},$$

so

$$\ddot{\hat{\boldsymbol{r}}} = \ddot{\theta}\hat{\boldsymbol{\theta}} + \dot{\theta}\dot{\hat{\boldsymbol{\theta}}} = \ddot{\theta}\hat{\boldsymbol{\theta}} - \dot{\theta}^2\hat{\boldsymbol{r}},$$

since by (59) and Chain rule

$$\dot{\hat{\boldsymbol{\theta}}} = \dot{\theta}(-\cos\theta\boldsymbol{i} - \sin\theta\boldsymbol{j}) = -\dot{\theta}\hat{\boldsymbol{r}}.$$

(We have seen many times that a time derivative of a fixed length vector is perpendicular to this vector.)

Substituting the above expressions for $\dot{\hat{\boldsymbol{r}}}$ and $\ddot{\hat{\boldsymbol{r}}}$ into (69) we get

$$\ddot{\boldsymbol{r}} = (\ddot{r} - r\dot{\theta}^2)\hat{\boldsymbol{r}} + (2\dot{r}\dot{\theta} + r\ddot{\theta})\hat{\boldsymbol{\theta}} \tag{71}$$

This is the formula for acceleration expressed in polar coordinates. Now, substituting it into $\boldsymbol{F} = m\ddot{\boldsymbol{r}}$, with $\boldsymbol{F}$ given by (58) we observe that the force vector has a zero $\hat{\boldsymbol{\theta}}$ component. I.e., equalising the $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{\theta}}$ components of the vectors $m\ddot{\boldsymbol{r}}$ and $\boldsymbol{F}$ we obtain two scalar equations:

$$\begin{cases} m(\ddot{r} - r\dot{\theta}^2) & = & f(r), \\ 2\dot{r}\dot{\theta} + r\ddot{\theta} & = & 0. \end{cases} \tag{72}$$

The second equation enables us to exclude θ from the first one. Indeed, assuming that $r \neq 0$ (which can never happen as long as the angular momentum is not zero) we can multiply the second equation by mr and observe that it then becomes

$$0 = 2mr\dot{r}\dot{\theta} + r^2\ddot{\theta} = \frac{d}{dt}(mr^2\dot{\theta}).$$

Hence, $mr^2\dot{\theta} = const$. We have thus merely arrived into the absolute value of the angular momentum: $const = L$. Now eliminating $\dot{\theta}$ from the first equation (71) reproduces (67).

4.7.2 Brief discussion of orbits

The equation (67), or rather its energy representation (66) can be solved for $r(t)$ via the energy method of Section 4.2. After the function $r(t)$ has been found, one goes back to (63) and integrates it:

$$\theta(t) = \theta(0) + \frac{L}{m}\int_0^t \frac{d\tau}{r^2(\tau)} \tag{73}$$

The pair of functions $(r(t), \theta(t))$ defines the particle trajectory, or *orbit.* A particular orbit is defined via the initial conditions $r(0), \theta(0)$, and the two constants of motion E and L, which depend on the initial position and initial velocity of the particle.

Let us now focus on finding $r(t)$. The presence of a centrifugal term in the effective force makes solving this equation explicitly for an arbitrary $f(r)$ quite difficult, so the equation is usually approached by energy methods. Looking back at (33) – (34) all one needs to do is replace x there with r and the potential $U(x)$ by the effective potential $V(r)$. Hence, the general solution $t(r)$, which has to be inverted into $r(t)$ is

$$t = \sqrt{\frac{m}{2}}\int_{r(0)}^{r} \frac{dy}{\sqrt{E - V(y)}}. \tag{74}$$

Recall that

$$V(r) = -\int f(r)dr + \frac{L^2}{2mr^2},$$

and in case of the forces $f(r)$ given by inverse powers of r finding the integral in (73) is feasible, with some knowledge of special functions. It is in particular feasible, without any special functions, for the inverse square force governing planetary motion – this was done by Newton. We shall stop here short of doing this; there are many accounts on how to do this, see e.g. Wikipedia's http://en.wikipedia.org/wiki/Kepler's_laws_of_planetary_motion.

Let us further confine ourselves with the analysis of *circular* and nearby orbits, by essentially repeating the arguments of Section 4.3. Circular orbits are the solutions in the form $r(t) = const.$ which arise as those values of r which are critical points of the effective potential, i.e where

$$V'(r) = 0.$$

Equivalently, if for some $r = R$, the effective force $\tilde{f}(R)$ is zero, then the initial condition $r(0) = R, \dot{r}(0) = 0$ will persist, i.e. the particle will be moving around a circle, with radius R and constant angular velocity

$$\omega_R = \frac{L^2}{mR^2}.$$

Observe that in order to find itself on a circular orbit, the particle must have a specific values of the constant of motion L, as R satisfies the equation

$$-f(R) = \frac{L^2}{mr^3}.$$

E.g. in the attracting gravitational force created by the stationary Sun of mass $M \gg m$, we have $f(r) = -\frac{GmM}{r^2}$, the radius of a single circular orbit is

$$R = \frac{L^2}{Gm^2M}.$$

In section 4.3 we used Taylor expansions (36) to describe approximately the behaviour of one degree of freedom systems near equilibria. In particular, we called a critical point of the potential, or equilibrium, *stable* if the second derivative of the potential at that point was positive and *unstable* if the second derivative of the potential there was negative.

The formulae (37), (38) for the behaviour $x(t) - x_*$ near an equilibrium x_* transfer verbatim as approximate solutions $r(t) - R$ near circular orbits, the parameter $\omega = \sqrt{\frac{|U''(x_*)|}{m}}$ in these formulae being replaced by

$$\omega_O = \sqrt{\frac{|V''(R)|}{m}}.$$

Hence, circular orbits can be stable or unstable, depending on the sign of $V''(R)$:

$$r(t) - R \approx \begin{cases} a\sin(\omega_O t + \phi), & V''(R) > 0, \\ a_1 e^{\omega_O t} + a_2 e^{-\omega_O t}, & V''(R) < 0. \end{cases}$$

This is merely a restatement of (37), (38).

A stable orbit ($V''(R) > 0$) oscillates around the circular orbit, as shown qualitatively in the figure.

An orbit, obtained via the Taylor approximation $V(r) = V(R) + \frac{1}{2}V''(R)(r-R)^2$ near a stable circular orbit will be closed, when the ratio of the period of the circular orbit

$$T_R = \frac{2\pi}{\omega_R} = 2\pi\frac{mR^2}{L^2}$$

and the period of oscillations

$$T_O = 2\pi\sqrt{\frac{m}{V''(R)}}$$

are rational multiples of each other – in the figure $T_R = 3T_O$. Otherwise, the orbit near a stable circular orbit will wind around it without ever closing upon itself.

Unstable circular orbits occur when $V''(R) < 0$, in which case the two exponential solutions correspond to orbits winding onto or spiralling out of the circular orbit. In the above example with the gravitational field, an easy calculation, or merely looking at the shape of the effective potential

$$V(r) = -\frac{GmM}{r} + \frac{L^2}{2mr^2}$$

(the second term dominates and goes to $+\infty$ for small r and the first term dominates and approaches zero from below for large r) shows that $R = \frac{L^2}{Gm^2M}$ is the minimum of the potential, and therefore the circular orbit is stable.