Notes for Classical Mechanics I, CMI, Autumn 2022

Govind S. Krishnaswami, November 17, 2022

Please let me know at govind@cmi.ac.in of any comments or corrections

Course website http://www.cmi.ac.in/~govind/teaching/cm1-o22

Contents

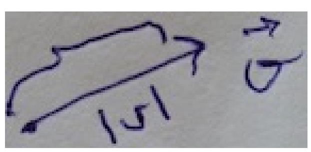
1 Primer on vectors, polar coordinates and kinematics

1.1 Vectors, dot and cross product

• Vectors provide a convenient way of writing the equations of physics in compact form. Newton's 2nd law $F_x = ma_x, F_y = ma_y, F_z = ma_z$ becomes the single equation of motion $\boldsymbol{F} = m\boldsymbol{a}$ where $\boldsymbol{F} = (F_x, F_y, F_z)$ and $\boldsymbol{a} = (a_x, a_y, a_z)$ are ordered triples.

• Vectors also make it easier to understand structural features (behavior under rotations or other transformations which could be symmetries) of physical quantities and equations. They allow us to exploit tools from linear algebra and geometry.

• A nonzero vector $\boldsymbol{v}$ in 3d Euclidean space is a directed line segment emanating from a chosen origin. It has a magnitude or length or norm denoted $|\boldsymbol{v}| = v$ (which is a positive real number determined using a measuring scale) and a direction (which is specified with respect to the origin and relative to other objects). Additionally there is an exceptional vector called the zero vector $\boldsymbol{0}$, which has zero length. It is convenient to think of the zero vector as pointing in all directions!

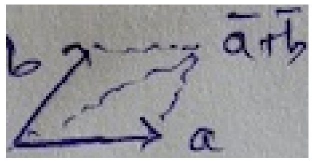


• Examples of vectors are the velocity and acceleration of a particle, the force acting on it, the electric field at a point in space etc.

• The multiplication of a vector $\boldsymbol{v}$ by the real number α (also called a real scalar) denoted $\alpha\boldsymbol{v}$ is a vector in the same or opposite direction as $\boldsymbol{v}$ (according as $\alpha \geq 0$ or $\alpha \leq 0$) that has the length $|\alpha||\boldsymbol{v}|$. For example $-\frac{1}{2}\boldsymbol{v}$ is a vector of half the length that points in the direction opposite to $\boldsymbol{v}$. Moreover, $0\boldsymbol{v} = \boldsymbol{0}$ is the zero vector.

• A **unit vector** is one with unit length. Given a nonzero vector $\boldsymbol{a}$, the associated unit vector $\hat{a}$ is obtained by 'normalization', i.e., dividing it by its length: $\hat{\boldsymbol{a}} = \frac{\boldsymbol{a}}{|\boldsymbol{a}|}$. Conversely, $\boldsymbol{a} = a\hat{\boldsymbol{a}}$. For a unit vector $|\hat{a}| = 1$.

• The sum of the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ is the vector that points (from the common origin) along the diagonal of the parallelogram with adjacent sides $\boldsymbol{a}$ and $\boldsymbol{b}$.



• Evidently, the order does not matter: $\boldsymbol{a}+\boldsymbol{b}=\boldsymbol{b}+\boldsymbol{a}$. We say that addition of vectors is commutative or an abelian operation.

• The zero vector has the special property that $\mathbf{0}+\boldsymbol{v}=\boldsymbol{v}$ for any vector $\boldsymbol{v}$. We say that $\mathbf{0}$ is the additive identity. Moreover $1\boldsymbol{v}=\boldsymbol{v}$ for any vector, so 1 is the multiplicative identity.

• The zero vector $\mathbf{0}$ is not the same as the real number 0. The former lies at the chosen origin of 3d space while the latter is a point on the real line. They live in different spaces.

• The set of all vectors in 3d space (with a fixed origin) equipped with the operations of addition of vectors and multiplication of vectors by real numbers is called a 3d real vector space and denoted $\mathbb{R}^3$.

• The space of vectors on the 2d Euclidean plane with a chosen origin forms the 2d real vector space denoted $\mathbb{R}^2$.

Dot or scalar product. Geometry has to do with angles, lengths, notions of parallel and perpendicular etc. Geometry enters through the dot product of vectors. For two vectors in 3d space, we define their dot product as $\boldsymbol{a}\cdot\boldsymbol{b}=ab\cos\theta$ where θ is the angle between the vectors. It does not matter whether we measure θ from $\boldsymbol{a}$ to $\boldsymbol{b}$ or vice versa.

• Notice that $\boldsymbol{a}\cdot\boldsymbol{b}=\boldsymbol{b}\cdot\boldsymbol{a}$, so the dot product is commutative.

• Turning things around, the angle between vectors can be expressed in terms of the dot product $\theta=\arccos(\frac{\boldsymbol{a}\cdot\boldsymbol{b}}{ab})$.

• As a consequence of the definition, $\boldsymbol{a}\cdot\boldsymbol{a}=a^2$. Thus the length of a vector can also be expressed in terms of the dot product.

• The dot product is also called the scalar product since the result is a real number (scalar) and not a vector.

• Show the law of cosines for $\boldsymbol{c} = \boldsymbol{a} + \boldsymbol{b}$:

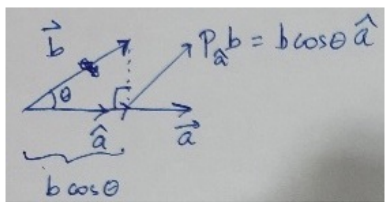
$$c^2 = a^2 + b^2 + 2ab\cos\theta. \tag{1}$$

• Two vectors are orthogonal or perpendicular if $\boldsymbol{a} \cdot \boldsymbol{b} = 0$.

• **Component of $\boldsymbol{b}$ in the direction of $\boldsymbol{a}$** Note that $b\cos\theta = \boldsymbol{b}\cdot\boldsymbol{a}/a = \boldsymbol{b}\cdot\hat{\boldsymbol{a}}$ is the component of $\boldsymbol{b}$ in the direction of $\boldsymbol{a}$. The component is just a real number, it can be positive or negative or even zero. It does not depend on the length of $\boldsymbol{a}$.

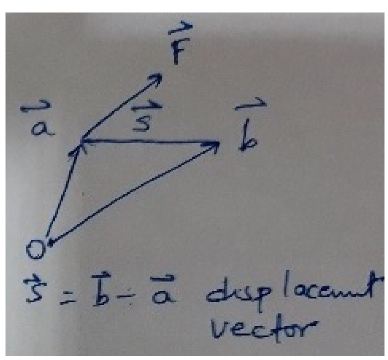
• The vector $P_{\hat{a}}\boldsymbol{b} = b\cos\theta\hat{a}$ is called the orthogonal projection of $\boldsymbol{b}$ on $\boldsymbol{a}$. It is a vector that points in the direction of $\hat{a}$ or $-\hat{a}$ and has a magnitude $|b\cos\theta|$ equal to the absolute value of the component of $\boldsymbol{b}$ along $\hat{a}$.

• Similarly, $\boldsymbol{a}\cdot\hat{\boldsymbol{b}}$ is the component of $\boldsymbol{a}$ in the direction of $\boldsymbol{b}$.



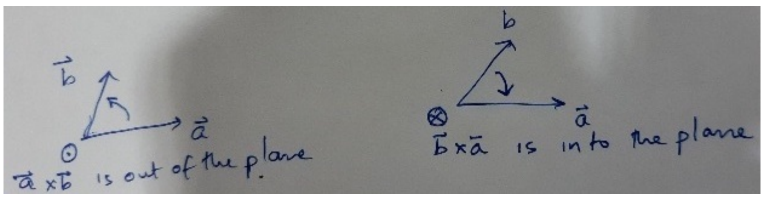
• The norm $||\boldsymbol{v}||$ of a vector $\boldsymbol{v}$ is defined as $||\boldsymbol{v}|| = \sqrt{\boldsymbol{v}\cdot\boldsymbol{v}}$. It is also called the Euclidean norm. The norm is the length of the vector, it is ≥ 0. The zero vector is the only one with zero norm.

• **Work as a scalar product.** Suppose a force $\boldsymbol{F}$ acts on a particle and displaces it from position vector $\boldsymbol{a}$ to an infinitesimally nearby position vector $\boldsymbol{b}$. The resulting infinitesimal displacement vector is $d\boldsymbol{s} = \boldsymbol{b} - \boldsymbol{a}$. Then the infinitesimal work done by the force is the scalar product $dW = \boldsymbol{F}\cdot d\boldsymbol{s}$.



• **Vector or cross product.** The vector or cross product $\boldsymbol{a}\times\boldsymbol{b}$ is a vector with magnitude $ab \sin\theta \geq 0$ where $0 \leq \theta \leq \pi$ is the angle between the vectors. Its direction is determined by the right hand thumb rule. If the fingers of the right hand curl from $\boldsymbol{a}$ to $\boldsymbol{b}$ then the cross product points in the direction of the thumb. [Another rule that gives the same direction is called the screw rule.]

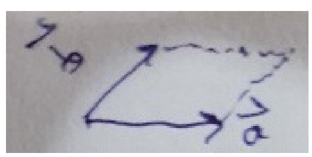
• In particular, $\boldsymbol{a} \times \boldsymbol{b}$ points in a direction perpendicular to both $\boldsymbol{a}$ and $\boldsymbol{b}$. Moreover, $\boldsymbol{a} \times \boldsymbol{b} = -\boldsymbol{b} \times \boldsymbol{a}$ is antisymmetric and so $\boldsymbol{a} \times \boldsymbol{a} = \boldsymbol{0}$ (the zero vector). Thus, the cross product is not commutative in general.



• The torque due to a force $\boldsymbol{F}$ on a particle located at the position vector $\boldsymbol{r}$ relative to a given origin is the cross product $\boldsymbol{\tau} = \boldsymbol{r} \times \boldsymbol{F}$.

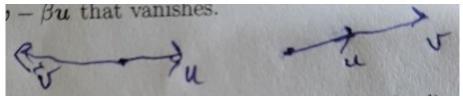
• How might a force point so that it imparts no torque on the particle?

• The magnitude of the area of a parallelogram spanned by the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ is equal to $|\boldsymbol{a} \times \boldsymbol{b}| = |\boldsymbol{b} \times \boldsymbol{a}|$. Sometimes it is useful to view the area as a vector: the area vector of the parallelogram spanned by $\boldsymbol{a}$ and $\boldsymbol{b}$ is $\boldsymbol{a} \times \boldsymbol{b}$.



• **Linear combination.** Given vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, $\alpha\boldsymbol{u} + \beta\boldsymbol{v}$ where α, β are real numbers is called a linear combination.

• **Linear dependence.** Two vectors $\boldsymbol{u}, \boldsymbol{v}$ are linearly dependent if one can be expressed as a multiple of the other, i.e., if they point in the same or opposite directions. In other words, $\boldsymbol{u} = \alpha\boldsymbol{v}$ or $\boldsymbol{v} = \beta\boldsymbol{u}$ for some real numbers α and β. So when they are dependent, there is a linear combination $\boldsymbol{u} - \alpha\boldsymbol{v}$ or $\boldsymbol{v} - \beta\boldsymbol{u}$ that vanishes. We need to allow for both possibilities. For instance if $\boldsymbol{u} = 0\boldsymbol{v}$ and $\boldsymbol{v} \neq \boldsymbol{0}$ then β is formally infinite and we do not have a relation of the sort $\boldsymbol{v} = \beta\boldsymbol{u}$. In summary, $\boldsymbol{u}$ and $\boldsymbol{v}$ are linearly dependent if there are real numbers a, b (not both zero) such that $a\boldsymbol{u} + b\boldsymbol{v} = 0$.



• $\boldsymbol{u}$ and $\boldsymbol{v}$ are said to be linearly independent if they point in different directions (i.e., are neither parallel not antiparallel.).

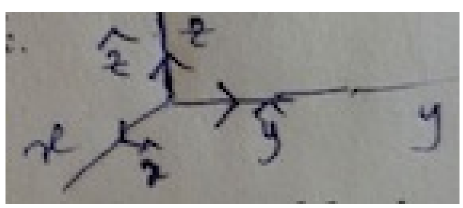
• Show that $\boldsymbol{u}$ and $\boldsymbol{v}$ are linearly independent if and only if $\boldsymbol{u} \times \boldsymbol{v} \neq \mathbf{0}$. In other words, two vectors are linearly independent if the parallelogram they span has nonzero area. This idea can be generalized to 3d. 3 vectors in $\mathbb{R}^3$ are linearly independent if the volume of the parallelepiped they define is nonzero.

• The zero vector $\mathbf{0}$ and any other vector $\boldsymbol{u}$ are always linearly dependent, since $\mathbf{0} = 0\boldsymbol{u}$.

• **Linear independence.** Vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \cdots$ are said to be linearly independent if the only linear combination $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 + \cdots$ that vanishes is the one where $c_1 = c_2 = c_3 = \cdots = 0$. In other words, the only linear combination that vanishes, is the trivial one.

• In $\mathbb{R}^3$ we can have at most three linearly independent vectors.

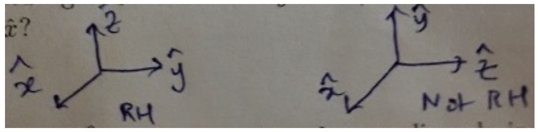
• **Cartesian axes** are any choice of mutually perpendicular axes x, y, z in $\mathbb{R}^3$.



• **Cartesian orthonormal frame.** We denote the unit vectors along the Cartesian axes by $\hat{x}, \hat{y}$ and $\hat{z}$ or $\hat{i}, \hat{j}$ and $\hat{k}$. They are orthonormal in the sense that they each have unit norm and are mutually perpendicular:

$$\hat{x} \cdot \hat{x} = \hat{y} \cdot \hat{y} = \hat{z} \cdot \hat{z} = 1 \quad \text{and} \quad \hat{x} \cdot \hat{y} = \hat{y} \cdot \hat{z} = \hat{z} \cdot \hat{x} = 0. \tag{2}$$

The $(\hat{x}, \hat{y}, \hat{z})$ frame is called right-handed if $\hat{x} \times \hat{y} = \hat{z}$ (rather than $-\hat{z}$, in which case it is left-handed). What are $\hat{y} \times \hat{z}$ and $\hat{z} \times \hat{x}$? We will work with right-handed frames.

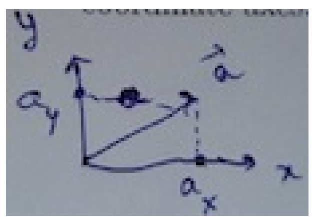


• $\hat{x}, \hat{y}$ and $\hat{z}$ form a basis for $\mathbb{R}^3$ in the sense that they are linearly independent and any vector can be written (uniquely) as a linear combination of them:

$$\boldsymbol{a} = a_x\hat{x} + a_y\hat{y} + a_z\hat{z}. \tag{3}$$

The three real numbers a_x, a_y and a_z are the components of $\boldsymbol{a}$ along the three coordinate frame vectors, verify that

$$a_x = \boldsymbol{a} \cdot \hat{x}, \quad a_y = \boldsymbol{a} \cdot \hat{y} \quad \text{and} \quad a_z = \boldsymbol{a} \cdot \hat{z}. \tag{4}$$



• Notice that $a_x\hat{x} = P_{\hat{x}}\boldsymbol{a}$ is the projection of $\boldsymbol{a}$ along $\hat{x}$. We say that the vector $\boldsymbol{a}$ has been resolved into its components and written as a sum of its orthogonal projections along the orthonormal basis vectors.

• In fact, (a_x, a_y, a_z) are the Cartesian coordinates of the location of the tip of the vector $\boldsymbol{a}$.

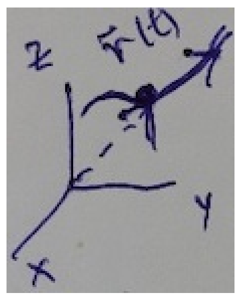
• Verify that the dot product of $\boldsymbol{a} = a_x\hat{x}+a_y\hat{y}+a_z\hat{z}$ and $\boldsymbol{b} = b_x\hat{x}+b_y\hat{y}+b_z\hat{z}$ is $\boldsymbol{a} \cdot \boldsymbol{b} = a_xb_x + a_yb_y + a_zb_z$.

• Express the Cartesian components of the cross product $\boldsymbol{a} \times \boldsymbol{b}$ in terms of those of $\boldsymbol{a}$ and $\boldsymbol{b}$.

1.2 Position coordinates and velocity and acceleration vectors

• The instantaneous location of a particle moving in 3d Euclidean space may be specified by its Cartesian coordinates $(x(t), y(t), z(t))$. As the particle moves, it traces out a curve parametrized by time, called its trajectory. It is a directed curve, the direction being that of increasing time.

• The instantaneous location of a particle is the same no matter which coordinate system we use to describe it. The latter is simply a convenient way of specifying its 'address'. The coordinates x, y, z depend on the choice of origin and orientation of coordinate axes. If we change the origin of our coordinates or orientation of the axes, we will get a different set of coordinates to describe the location of the particle at a given time. Two people following different coordinate systems will nevertheless meet each other at the common instantaneous location of the particle. [For instance, a courier delivers a letter to the same geographic location irrespective of whether the address on the envelope says CMI, Old number 2, 2nd Avenue or CMI, New number 5, 2nd Avenue].



• The vector that points from the origin of Cartesian coordinates to the instantaneous position of the particle, has components $(x(t), y(t), z(t))$. It is called the position vector and is denoted $\boldsymbol{r}(t)$. While such a designation is convenient for some purposes, it is important to bear in mind that the location of a particle is not really a vector: it is not physically associated to a direction and the location of the particle does not come with any intrinsic notion of an origin or a length.

• The infinitesimal displacement of a particle over a time $[t, t + \delta t]$ *does* define a vector, albeit a vector with infinitesimal length:

$$\delta \boldsymbol{r}(t) = \boldsymbol{r}(t + \delta t) - \boldsymbol{r}(t). \tag{5}$$

The concept of infinitesimal displacement defines an origin, namely the initial location of the particle (at time t). The infinitesimal displacement vector then points from this origin to the final location of the particle (at time $t + \delta t$).

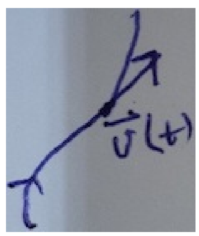
• The concept of infinitesimal displacement does not define a coordinate frame, it only defines an origin and a vector $\delta \boldsymbol{r}$. We may resolve $\delta \boldsymbol{r}$ along the axes of any frame. Here, we will parallel transport $\delta\boldsymbol{r}$ to the origin of our Cartesian coordinate frame, its components with respect to this frame are

$$\delta\boldsymbol{r}(t) = \boldsymbol{r}(t+\delta t)-\boldsymbol{r}(t) = (x(t+\delta t)-x(t), y(t+\delta t)-y(t), z(t+\delta t)-z(t)) \quad (6)$$

• The arbitrarily chosen origin of the Cartesian coordinate system (x, y, z) has no physical relevance to the infinitesimal displacement vector. We have parallel transported it to this origin for ease of some later calculations.

• The velocity of the particle is defined as the limiting value of the difference quotient

$$\boldsymbol{v}(t) = \lim_{\delta t \to 0} \frac{\delta\boldsymbol{r}(t)}{\delta t}. \quad (7)$$



• The velocity of the particle is a vector. It is the time derivative of the position along the trajectory. The velocity

$$\boldsymbol{v}(t) = \frac{d\boldsymbol{r}(t)}{dt} \quad (8)$$

defines a vector that points from the instantaneous location in the direction of motion. It is a tangent vector to the trajectory. Its magnitude is called the instantaneous speed of the particle.

• The origin from which the velocity vector points moves with the particle.

• For many purposes, it is convenient to resolve $\boldsymbol{v}$ along the Cartesian coordinate axes by moving the origin of the Cartesian frame to the current location of the particle. With this understanding, and denoting time derivatives with an over-dot,

$$\boldsymbol{v}(t) = (\dot{x}(t), \dot{y}(t), \dot{z}(t)) = \dot{x}\hat{x} + \dot{y}\hat{y} + \dot{z}\hat{z} \quad (9)$$

• The time-derivative of the velocity is the acceleration, which may be viewed as the vector

$$\boldsymbol{a}(t) = \ddot{x}(t)\hat{x} + \ddot{y}(t)\hat{y} + \ddot{z}(t)\hat{z}. \quad (10)$$

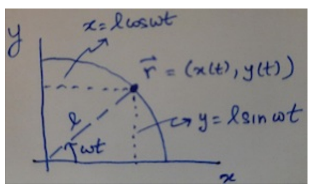
It is the limit of the difference quotient $(\boldsymbol{v}(t+\Delta t) - \boldsymbol{v}(t))/\Delta t$ as $\Delta t \to 0$. It may be regarded as a vector emanating from the instantaneous location of the particle.

1.3 Uniform circular motion

Suppose a particle moves counterclockwise on the circle $x^2 + y^2 = \ell^2$ of radius ℓ in the x-y plane at a constant angular speed $\omega > 0$ radians per second. Assuming it starts from the point $(\ell, 0)$ at $t = 0$, its instantaneous location may be given by the Cartesian coordinates

$$x(t) = \ell \cos \omega t \quad \text{and} \quad y(t) = \ell \sin \omega t. \tag{11}$$

Sometimes, it is convenient to regard $\boldsymbol{r}(t) = \ell \cos \omega t \hat{x} + \ell \sin \omega t \hat{y}$ as a vector that points radially outwards from the center of the circle. Thus $\boldsymbol{r}(t)$ is the position vector of the particle relative to the origin chosen to lie at the center of the circle.



• Let us see why this formula is justified. Notice that $x^2 + y^2 = \ell^2$ at all times and that the motion is counterclockwise. The velocity vector is given by

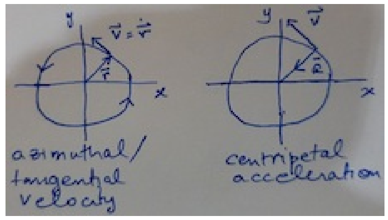
$$\boldsymbol{v}(t) = -\ell\omega \sin \omega t \; \hat{x} + \ell\omega \cos \omega t \; \hat{y}. \tag{12}$$

The speed of the particle $v = |\boldsymbol{v}(t)| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}} = \ell\omega$ is constant ensuring uniform circular motion. The particle goes round the circle once in a time $T = 2\pi\ell/v = 2\pi/\omega$. Thus, the particle covers 2π radians in a time $2\pi/\omega$ resulting in an angular speed of ω radians per second (angular speed is sometimes called angular frequency).

• Notice that $\boldsymbol{v} \cdot \boldsymbol{r} = 0$. Thus, the velocity vector is tangent to the circle. The acceleration is given by

$$\boldsymbol{a}(t) = \dot{\boldsymbol{v}}(t) = -\ell\omega^2 \cos \omega t \hat{x} - \ell\omega^2 \sin \omega t \hat{y} = -\omega^2 \boldsymbol{r}(t). \tag{13}$$

We see that the acceleration has the constant magnitude $|\boldsymbol{a}(t)| = \ell\omega^2$ and points radially inwards towards the center of the circle. The latter feature justifies the name centripetal acceleration. Centripetal means 'seeking the center' in Latin.



• The time derivative of acceleration $\dot{\boldsymbol{a}}$ is sometimes called jerk or jolt. Show that $\dot{\boldsymbol{a}} \cdot \boldsymbol{a} = 0$. Which way does the jerk point in uniform circular motion?

• We observe that if there is a radially inward force $\boldsymbol{F}$ (like a person tugging at a string with a stone tied at the other end and rotated) that is responsible for this circular motion, then $\boldsymbol{F} \cdot d\boldsymbol{s} = 0$. Here $d\boldsymbol{s}$ is the infinitesimal displacement of the particle, which is tangent to the circle. This dot product is called the infinitesimal work done by the force in displacing the particle $d\boldsymbol{s}$. Thus we see that a radially inward force does no work in moving a particle along a circular trajectory. This is not the case if the motion is due to a force that is tangential - like an agent pushing the particle along the rim of the circle.

1.4 Nonuniform circular motion

• We may model nonuniform circular motion of a particle around a circle of radius ℓ via the position vector

$$\boldsymbol{r}(t) = \ell(\cos\theta(t), \sin\theta(t)) = \ell(\cos\theta(t)\hat{x} + \sin\theta(t)\hat{y}). \quad (14)$$

If we denote $\omega(t) = \dot{\theta}$, then the angular speed of such a particle is $|\omega(t)| = |\dot{\theta}|$, which we suppose is not constant.

• The velocity of such a particle is

$$\boldsymbol{v}(t) = \ell\dot{\theta}(-\sin\theta(t), \cos\theta(t)). \quad (15)$$

Notice that $\boldsymbol{v} \cdot \boldsymbol{r} = 0$. So $\boldsymbol{v}$ always points tangent to the circle as it must for a particle confined to the circle. However, the linear speed $\ell|\dot{\theta}|$ may vary with time.

• Using the product rule of differentiation, the acceleration is given by

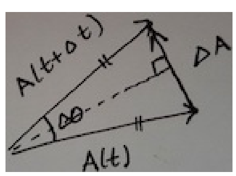
$$\boldsymbol{a} = \dot{\boldsymbol{v}} = \ell\ddot{\theta}(-\sin\theta, \cos\theta) - \ell\dot{\theta}^2(\cos\theta, \sin\theta). \tag{16}$$

The first term points tangentially (counterclockwise or clockwise depending on the sign of $\ddot{\theta}$), and is called the angular acceleration while the second term points radially inwards and is called the centripetal acceleration. Thus $\boldsymbol{a} \cdot \boldsymbol{v} \neq 0$ in general for nonuniform circular motion.

1.5 Rotating vectors

• We observed that the velocity of a uniformly rotating particle is orthogonal to its radius vector, i.e., the time derivative of the position vector $\boldsymbol{r}$ is perpendicular to $\boldsymbol{r}$: $\boldsymbol{r} \cdot (d\boldsymbol{r}/dt) = \boldsymbol{r} \cdot \boldsymbol{v} = 0$. In fact, this is true even if the rotation is not uniform. Let us comment on the significance of this.

• Suppose $\boldsymbol{A}$ is any vector such that its time derivative is orthogonal to $\boldsymbol{A}$ at all times. Then the infinitesimal change in $\boldsymbol{A}$ in a short time is perpendicular to $\boldsymbol{A}$ and not along $\boldsymbol{A}$. This means the magnitude of $\boldsymbol{A}$ cannot change (it cannot shrink or elongate), and the vector can only rotate. Let us obtain a formula for $|d\boldsymbol{A}/dt|$. Suppose $\Delta\boldsymbol{A} = \boldsymbol{A}(t+\Delta t) - \boldsymbol{A}(t)$ is the infinitesimal change in $\boldsymbol{A}$. Since the length of $\boldsymbol{A}$ does not change, the three vectors $\boldsymbol{A}(t)$, $\boldsymbol{A}(t+\Delta t)$ and $\Delta\boldsymbol{A}$ form an isosceles triangle with $\Delta\boldsymbol{A}$ as base. Let us denote the angle at the apex of this isosceles triangle by $\Delta\theta$, which is the angle of rotation.



• Then

$$|\Delta\boldsymbol{A}| = |2A\sin(\Delta\theta/2)| \approx A|\Delta\theta| \quad \text{for} \quad \text{small } \Delta\theta. \tag{17}$$

We have used the linear Taylor approximation for the sine function (more on this soon). Taking the limit $\Delta t \to 0$,

$$\left|\frac{d\boldsymbol{A}}{dt}\right| = A\left|\frac{d\theta}{dt}\right|. \tag{18}$$

Here, $|d\theta/dt|$ is the angular speed of $\boldsymbol{A}$.

• We may apply this to circular motion where $\boldsymbol{A} = \boldsymbol{r}$ is the radius vector of the particle and $d\boldsymbol{A}/dt = \boldsymbol{v}$ is its velocity. Then the linear speed of the particle is $v = r|\omega|$ where $\omega = d\theta/dt$ is the angular speed (positive for counterclockwise motion). Note that v, ω need not be constant. Uniform circular motion is a special case where ω is a constant and

$$\boldsymbol{r} = r(\cos\omega t\,\hat{x} + \sin\omega t\,\hat{y}) \quad \text{and} \quad \boldsymbol{v} = r\omega(-\sin\omega t\,\hat{x} + \cos\omega t\,\hat{y}) \qquad (19)$$

Notice that $\boldsymbol{v}\cdot\boldsymbol{r} = 0$ since $\boldsymbol{v}$ points tangentially/azimuthally while $\boldsymbol{r}$ is radial. This ensures that the length of $\boldsymbol{r}$ does not change with time. What is more, we showed that the acceleration $\boldsymbol{a} = \dot{\boldsymbol{v}} = -\omega^2\boldsymbol{r}$ so that $\dot{\boldsymbol{v}}$ is perpendicular to $\boldsymbol{v}$ for uniform circular motion. Thus, the velocity vector cannot change in magnitude and must also simply rotate! Verify that the same is true of $\boldsymbol{a}$ as well, for uniform circular motion.

1.6 Integration of kinematical equations

As we will soon learn, if the forces on a particle are known, then one may use Newton's second law to find its acceleration. This is called the dynamical part of the problem of motion, since it depends on the forces and interactions. The kinematical part of the problem of motion is to determine the velocity of the particle and its trajectory from its acceleration.

• Suppose we are given the acceleration of a particle as a function of time. Then the velocity must satisfy $\frac{d\boldsymbol{v}}{dt} = \boldsymbol{a}(t)$. This is a first order differential equation since it only involves the first derivative. The independent variable is t and the vector $\boldsymbol{v}(t)$ is called the unknown or dependent variable. Integrating this equation with respect to time from t_0 to t, we get

$$\boldsymbol{v}(t) = \boldsymbol{v}(t_0) + \int_{t_0}^{t} \boldsymbol{a}(t')dt'. \qquad (20)$$

In addition to knowledge of the acceleration, here we needed an 'initial condition' $\boldsymbol{v}(t_0)$ (actually three ICs, the three Cartesian components of $\boldsymbol{v}(t_0)$) to determine the velocity. The problem of determining velocity has been reduced to quadratures i.e., to evaluating integrals (one each for the three Cartesian components of velocity).

• The step from velocity to position involves one more integration and another initial condition:

$$\dot{\boldsymbol{r}}(t) = \boldsymbol{v}(t) \quad \Rightarrow \quad \boldsymbol{r}(t) = \boldsymbol{r}(t_0) + \int_{t_0}^{t} \boldsymbol{v}(t')dt'. \tag{21}$$

Evaluating these integrals for a specific acceleration may or may not be feasible analytically.

• We have solved the 2nd order ordinary differential equations $\ddot{\boldsymbol{r}}(t) = \boldsymbol{a}(t)$ in two steps. Being 2nd order, the process required two initial conditions or pieces of initial data $\boldsymbol{v}(t_0)$ and $\boldsymbol{r}(t_0)$ (each of which is a vector with three components).

• A simple example is that of uniform acceleration, i.e., where $\boldsymbol{a}(t)$ is a constant vector $\boldsymbol{a}$. In this case,

$$\boldsymbol{v}(t) = \boldsymbol{v}(t_0) + (t - t_0)\boldsymbol{a}, \tag{22}$$

and integrating once more,

$$\boldsymbol{r}(t) = \boldsymbol{r}(t_0) + (t - t_0)\boldsymbol{v}(t_0) + \frac{1}{2}(t^2 - t_0^2)\boldsymbol{a} - (t - t_0)t_0\boldsymbol{a}. \tag{23}$$

The formula simplifies if $t_0 = 0$:

$$\boldsymbol{r}(t) = \boldsymbol{r}(0) + t\boldsymbol{v}(0) + \frac{1}{2}t^2\boldsymbol{a} \tag{24}$$

This formula applies to the case of constant acceleration.

1.7 Plane polar coordinates

• For many problems, especially those where there is rotational symmetry around a central object, polar coordinates are more convenient than Cartesian coordinates.

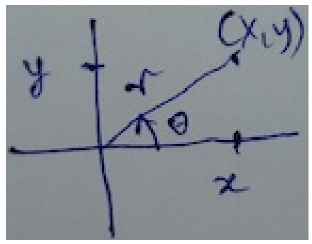
• For simplicity, we consider polar coordinates (r, θ) on the plane. Suppose we are given an origin O and horizontal and vertical x and y axes. Given a point $P(x, y)$, r is the distance of P from the origin, and θ is the counterclockwise angle the radius vector $\boldsymbol{r} = (x, y)$ makes with the horizontal axis. Note that x and y are called the abscissa and ordinate of the point P. In other words, $\cos\theta = x/r$ or $\tan\theta = y/x$. Thus,

$$r = \sqrt{x^2 + y^2} \quad \text{and} \quad \theta = \arctan(y/x) = \arccos(x/r). \tag{25}$$

Conversely,

$$x = r\cos\theta \quad \text{and} \quad y = r\sin\theta. \tag{26}$$

• Notice that θ is defined modulo 2π. $\theta = 0$ and $\theta = 2\pi$ both correspond to the positive x-axis. One often chooses a convenient 'fundamental domain' for θ such as $[0, 2\pi)$ or $(-\pi, \pi]$.



It is important to observe that the polar coordinate system breaks down (or is singular) at the origin where $x = y = r = 0$. At this point, θ is not defined. In a sense, the point at the origin could be assigned any value of θ, depending on how we approach the origin. Said differently, the map between x, y and r, θ fails to be 1-1 at the origin.

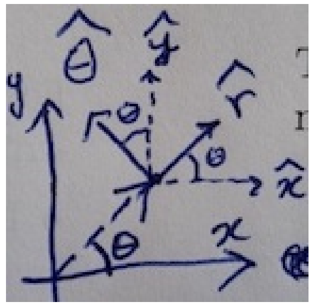
• Notice that the constant x and constant y curves (also known as the level curves[1] of x and y) are mutually orthogonal straight lines parallel to the y and x axes respectively. By contrast, the constant θ curves are rays emanating radially outwards from the origin while the constant r curves are concentric circles centered at O. This explains why polar coordinates are called curvilinear coordinates. Despite being curvilinear, the level curves of r and θ are mutually orthogonal.

• A real-valued function on the plane is any function of x and y that assigns a real number to each point (x, y). The simplest of these functions are the 'coordinate functions' x and y themselves.

• Analogously, r and θ are the coordinate functions in polar coordinates.

• **Unit vectors $\hat{r}$ and $\hat{\theta}$.** Recall that at a point (x, y) on the plane, $\hat{x}$ and $\hat{y}$ are unit vectors in the directions of increasing x holding y fixed and vice versa. Similarly, we define the unit vectors $\hat{r}$ and $\hat{\theta}$ at any point (r, θ). $\hat{r}$ points radially outwards while $\hat{\theta}$ points counterclockwise tangentially to the circle of radius r. The direction in which $\hat{\theta}$ points is called the azimuthal direction.

[1] A level curve of a quantity is a curve on which the quantity is a constant.



• $\hat{x}, \hat{y}$ furnish one basis for vectors at any point on the plane. Similarly, $\hat{r}, \hat{\theta}$ furnish another basis at points away from $r = 0$. We can therefore expand $\hat{r}$ and $\hat{\theta}$ in the $\hat{x}$, $\hat{y}$ basis at any point.

• A figure shows that we may decompose $\hat{r}$ and $\hat{\theta}$ as

$$\begin{aligned} \hat{r} &= \cos\theta\,\hat{x} + \sin\theta\,\hat{y} = \frac{x}{r}\hat{x} + \frac{y}{r}\hat{y} \\ \text{and} \quad \hat{\theta} &= -\sin\theta\,\hat{x} + \cos\theta\,\hat{y} = -\frac{y}{r}\hat{x} + \frac{x}{r}\hat{y}. \end{aligned} \tag{27}$$

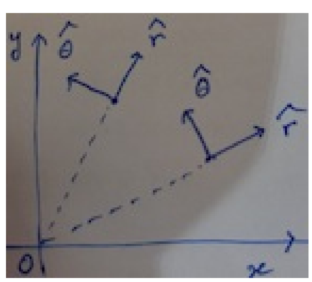
In fact, the figure shows that the $\hat{r} - \hat{\theta}$ frame is obtained from the $\hat{x} - \hat{y}$ frame through a rotation by angle θ counterclockwise.

• Check that $\hat{r}$ and $\hat{\theta}$ are orthonormal:

$$\hat{r} \cdot \hat{\theta} = 0 \quad \text{and} \quad \hat{r} \cdot \hat{r} = \hat{\theta} \cdot \hat{\theta} = 1. \tag{28}$$

• Exercise: Express $\hat{x}$ and $\hat{y}$ as linear combinations of $\hat{r}$ and $\hat{\theta}$.

• Unlike $\hat{x}$ and $\hat{y}$ which point in the same direction everywhere, the directions of $\hat{r}$ and $\hat{\theta}$ change with location.



• **Position coordinate and velocity vector.** The position vector $\boldsymbol{r}$ of a location P with coordinates (x, y) can now be expressed in polar coordinates. We parallel transport the position vector $\boldsymbol{r}$ from the origin to P and then write it as a linear combination of $\hat{x}$ and $\hat{y}$ or $\hat{r}$ and $\hat{\theta}$ at P:

$$\boldsymbol{r} = x\hat{x} + y\hat{y} = r\cos\theta\hat{x} + r\sin\theta\hat{y} = r\hat{r}. \tag{29}$$

We wish to find the velocity and acceleration vectors in plane polar coordinates. These are the polar coordinate analogues of $\boldsymbol{v} = \dot{x}\hat{x} + \dot{y}\hat{y}$ and $\boldsymbol{a} = \ddot{x}\hat{x} + \ddot{y}\hat{y}$.

• Now, suppose $\boldsymbol{r}(t) = r(t)\hat{r}(t)$ is the position vector of a particle at time t. As it moves along a trajectory, the radial coordinate r can change, but so can the direction of the unit vector $\hat{r}(t)$. Thus, its velocity is given by

$$\boldsymbol{v}(t) = \dot{\boldsymbol{r}}(t) = \dot{r}\,\hat{r} + r\frac{d\hat{r}}{dt}. \tag{30}$$

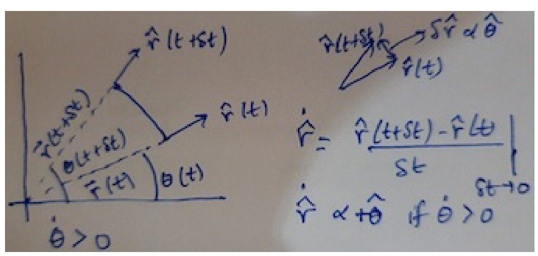
$\dot{r}\hat{r}$ is what we might naively guess as the radial velocity. The other term comes from the change in direction of the basis vector $\hat{r}$.

• Let us take a moment to find the rates of change of the basis unit vectors $\hat{r}$ and $\hat{\theta}$. Being unit vectors, their change can come only from a change in their direction. For instance,

$$\hat{r} = \cos\theta\hat{x} + \sin\theta\hat{y} \quad \Rightarrow \quad \frac{d\hat{r}}{dt} = -\sin\theta\,\dot{\theta}\,\hat{x} + \cos\theta\,\dot{\theta}\,\hat{y} = \dot{\theta}\hat{\theta}, \tag{31}$$

where we recalled that $\hat{\theta} = -\sin\theta\hat{x} + \cos\theta\hat{y}$.

• Thus, the change in $\hat{r}$ is always in the azimuthal $\hat{\theta}$ direction. Understand this through the figure



• Similarly,

$$\hat{\theta} = -\sin\theta\hat{x} + \cos\theta\hat{y} \quad \Rightarrow \quad \frac{d\hat{\theta}}{dt} = -\cos\theta\,\dot{\theta}\,\hat{x} - \sin\theta\,\dot{\theta}\,\hat{y} = -\dot{\theta}\,\hat{r}. \tag{32}$$

The rate of change of $\hat{\theta}$ always points radially.

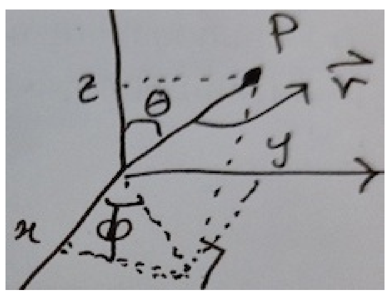
• Putting these together, we get the decomposition of the velocity of the particle in the polar coordinate basis.

1.8 Spherical polar coordinates

• The analogue of plane polar coordinates in 3d ($\mathbb{R}^3$) are called spherical polar coordinates (r, θ, ϕ). They are called the radial, polar and azimuthal coordinate respectively. They are particularly useful in dealing with systems where there is spherical symmetry about a central object such a the Sun in the solar system or the nucleus in an atom.

• Given a point P with Cartesian coordinates (x, y, z), the radial coordinate r is the distance of P from the origin $r = \sqrt{x^2 + y^2 + z^2}$. Evidently, $0 \le r < \infty$.

• If P has position vector $\boldsymbol{r}$ relative to the origin, then the *polar angle* θ is the angle $\boldsymbol{r}$ makes with respect to the upward vertical z axis. Thus $\theta = \arccos(z/r)$. Notice that $0 \le \theta \le \pi$ with $\theta = 0$ and $\theta = \pi$ corresponding to the positive and negative z axis. If we regard P as a point on the Earth, then θ specifies the latitude through P.



• Finally, suppose we orthogonally project the position vector onto the x-y plane. This projected vector has length $r \sin\theta = x^2 + y^2$.

• The azimuthal angle ϕ is defined as the angle that this projection makes with the x axis, measured counterclockwise. Thus, $\phi = \arccos(x/\sqrt{x^2 + y^2})$ and $\tan\phi = y/x$. Notice that $0 \le \phi < 2\pi$. If we view P as a point on the Earth, ϕ specifies the longitude passing through the point P.

• ϕ is the azimuthal angle for plane polar coordinates on the x-y plane (it was called θ in that discussion!)

• Unfortunately, the spherical polar coordinate system breaks down along the z axis, where ϕ is not uniquely defined (it can be assigned any value $0 \le \phi < 2\pi$). [When a coordinate system does not cover the whole of $\mathbb{R}^3$, we could introduce another set of coordinates that work in the excluded region.]

• Alternatively, we may write

$$z = r\cos\theta, \quad x = r\sin\theta\cos\phi \quad \text{and} \quad y = r\sin\theta\sin\phi. \tag{33}$$

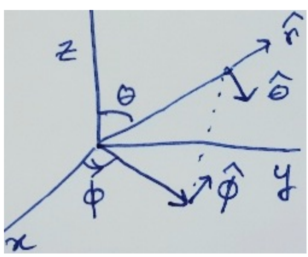
Check that $\tan\theta = \sqrt{x^2+y^2}/z$.

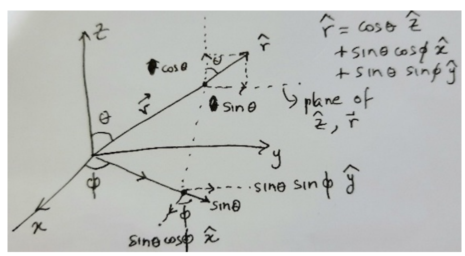
• The position vector of a particle located at (x,y,z) is then given by $\boldsymbol{r} = x\hat{x} + y\hat{y} + z\hat{z} = r\hat{r}$.

• The other formulae we derived for plane polar coordinates may be generalized to spherical polar coordinates.

• For instance, if $\hat{r}, \hat{\theta}, \hat{\phi}$ are the unit vectors in the directions of increasing r, θ, ϕ, then the figure helps us express

$$\begin{aligned} \hat{r} &= \cos\theta\hat{z} + \sin\theta(\cos\phi\hat{x} + \sin\phi\hat{y}), \\ \hat{\theta} &= -\sin\theta\hat{z} + \cos\theta(\cos\phi\hat{x} + \sin\phi\hat{y}) \quad \text{and} \\ \hat{\phi} &= -\sin\phi\hat{x} + \cos\phi\hat{y}. \end{aligned} \tag{34}$$





• Verify that $(\hat{r}, \hat{\theta}, \hat{\phi})$ is a right-handed orthonormal system.

1.9 Taylor approximation

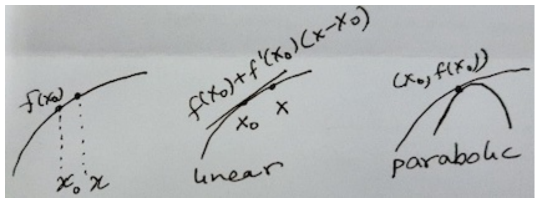
• Taylor series for one variable. Given a function of one real variable $f(x)$ that is continuous and hopefully differentiable a few times, we are interested in approximately evaluating it in the neighborhood of a point x_0.

• By continuity, $f(x) \approx f(x_0)$ is of course our zeroth order approximation to the value of the function for x near x_0.

• The next possibility is to approximate f by a linear function near x_0. It is natural to take the slope of this linear function to be the derivative of

f at x_0 (assuming f is differentiable at x_0), so that we approximate the graph of f by the tangent through the point $(x_0, f(x_0))$. This leads to the first order or linear Taylor approximation

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0). \tag{35}$$



• To indicate that $x - x_0$ is small, we will denote it by $\Delta x = x - x_0$, and denote $f(x) - f(x_0) = \Delta f$. Then we have $\Delta f \approx f'(x_0)\Delta x$. This is only an approximation

• It is also convenient to introduce the *differential* of f, which at x is defined as $df(x) = f'(x)dx$. dx is called the differential of x. The derivative denoted df/dx is the limit of $\Delta f/\Delta x$ as $\Delta x \to 0$. For example, $d\sin x = \cos x \, dx$. The differential of a function is also called a 1-form.

• More generally, if f is n times differentiable at x_0, we have the n^{th} order Taylor polynomial approximation for small $x - x_0$:

$$\begin{aligned} f(x) \;&\approx\; f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 + \frac{1}{3!}f'''(x_0)(x - x_0)^3 \\ &\quad + \cdots + \frac{1}{n!}f^{(n)}(x_0)(x - x_0)^n, \end{aligned} \tag{36}$$

where $f^{(n)}(x_0)$ is the n^{th} derivative of f at x_0.

• For many of the functions we encounter, the Taylor series, obtained by letting $n \to \infty$, converges to the function $f(x)$ for x in a neighborhood of x_0. Such functions are called real analytic.

• A real-valued function that is continuous in some domain is said to be of type C^0 in that domain. A function that is differentiable with continuous first derivative is said to be of class C^1 in that domain. Similarly we have the notion of C^k functions for $k = 1, 2, 3, \ldots$: k times continuously differentiable functions in some domain. A function that is C^k for all $k = 1, 2, 3, 4, \ldots$ is said to be smooth or C^∞. A function whose Taylor series converges to the function in some domain is said to be real analytic or of type C^ω.

• For example, show that the Taylor series for $1/(1-x)$ around $x=0$ is given by a geometric series:

$$\frac{1}{1-x} = 1 + x + x^2 + \cdots . \tag{37}$$

This series converges to $1/(1-x)$ for $|x|<1$. Also verify that $(1-x)(1+x+x^2+\cdots) = 1$ by multiplying things out and canceling. Note that $1/(1-x)$ does not admit a Taylor expansion around $x=1$.

• Show that the Taylor series for e^x, is given by

$$e^x = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots . \tag{38}$$

This series has an infinite radius of convergence. Find the Taylor series for $\sin x$ and $\cos x$.

• The binomial series is a very useful Taylor series around $x=0$:

$$(1+x)^\nu = 1 + \nu x + \frac{\nu(\nu-1)}{1\cdot 2}x^2 + \frac{\nu(\nu-1)(\nu-2)}{1\cdot 2\cdot 3}x^3 + \cdots \tag{39}$$

which converges for $|x|<1$ and any (real or complex) number ν. For a positive integer $\nu = n$, this series terminates and we recover the binomial expansion with coefficients given by combinatorial factors: $(1+x)^n = \sum_{r=0}^{n} \binom{n}{r} x^r$.

• In particular, show that

$$\frac{1}{\sqrt{1+x}} = 1 - \frac{x}{2} + \frac{x^2}{8} + \cdots . \tag{40}$$

• **Taylor series for more variables.** For a real function $f(x,y)$ of two variables, we have the Taylor expansion of f around a point (x_0, y_0):

$$\begin{aligned} f(x,y) &= f(x_0,y_0) + \frac{\partial f}{\partial x}|_{(x_0,y_0)}(x-x_0) + \frac{\partial f}{\partial y}|_{(x_0,y_0)}(y-y_0) \\ &+ \frac{1}{2}\left[\frac{\partial^2 f}{\partial x^2}(x-x_0)^2 + \frac{\partial^2 f}{\partial y^2}(y-y_0)^2 + \frac{\partial^2 f}{\partial x \partial y}(x-x_0)(y-y_0) + \frac{\partial^2 f}{\partial y \partial x}(y-y_0)(x-x_0)\right] + \cdots \end{aligned} \tag{41}$$

where all the partial derivatives are evaluated at (x_0, y_0). The mixed second partials $\frac{\partial^2 f}{\partial x \partial y}$ and $\frac{\partial^2 f}{\partial y \partial x}$ are equal (Clairaut's or Schwarz's Theorem, assuming the second partials are continuous).

• To calculate a partial derivative with respect to y we simply differentiate the function with respect to y treating x as fixed.

• One way to obtain this series is to treat y as fixed and first write down a Taylor series in x around x_0 with coefficients being functions of y. Then we expand these coefficients in a Taylor series in y.

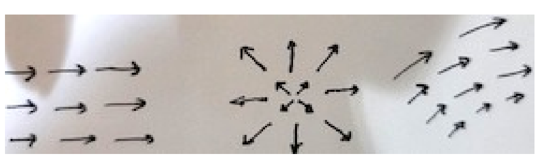
• Calculate the mixed second partials of $f(x,y) = \cos xy$ and show that they are both equal to $-\sin xy - xy\cos xy$

1.10 Some vector calculus

• **Scalar fields.** At a given instant of time, the pressure $p(\boldsymbol{r})$ in the atmosphere is a real number that depends on height, and more generally on location $\boldsymbol{r}$. The density $\rho(\boldsymbol{r})$ of sea water at a given instant of time depends on depth. Similarly, the salt concentration of sea water $c(\boldsymbol{r})$ depends on location. The potential energy $V(\boldsymbol{r})$ of a massive particle in Earth's gravity depends on height above the Earth's surface as well as the latitude and longitude. All these are examples of real valued functions in 3d space. We will also refer to real-valued functions as scalar fields. A scalar field assigns a real number to each location $\boldsymbol{r}$. Typically, the real number would vary smoothly (or at least continuously differentiably) as the location changes.

• Note that the notion of a field introduced here is different from the algebraic notion of a field (e.g., field of real or complex numbers). Here, field refers to something that depends on location.

• **Vector fields.** Similarly, we have the concept of a vector field: a smoothly varying vector $\boldsymbol{v}(\boldsymbol{r})$ at each location $\boldsymbol{r}$. The gravitational force felt by a point mass m at various locations and heights above the Earth's surface defines a vector field.



• In the figure, we have displayed three vector fields on the plane. Since such a vector field has two components in Cartesian coordinates, a vector field on the plane may be regarded as a map from $\mathbb{R}^2 \to \mathbb{R}^2$. If x and

y are the horizontal and vertical directions, then the first vector field is plausibly $\boldsymbol{v} \propto \hat{x}$. The second vector field points radially outwards with a magnitude increasing with radial distance and is circularly symmetric. It is plausible that the 2nd vector field $\boldsymbol{v} \propto x\hat{x} + y\hat{y} = \boldsymbol{r}$. The 3rd vector field could be the velocity vector field of a steadily flowing fluid.

• **Gradient of a scalar field.** Given a scalar field $\phi(\boldsymbol{r})$, its gradient is a kind of derivative that produces a vector field denoted $\boldsymbol{\nabla}\phi(\boldsymbol{r})$. In Cartesian coordinates $\boldsymbol{r} = (x, y, z)$,

$$\text{grad}\,\phi = \boldsymbol{\nabla}\phi(\boldsymbol{r}) = \frac{\partial\phi}{\partial x}\hat{x} + \frac{\partial\phi}{\partial y}\hat{y} + \frac{\partial\phi}{\partial z}\hat{z}. \tag{42}$$

• Example 1: If $\phi(x, y) = x$ then $\boldsymbol{\nabla}\phi = \hat{x}$ is a constant vector field pointing in the x direction at all points of $\mathbb{R}^2$.

• Example 2: If $\phi = \frac{1}{2}(x^2 + y^2 + z^2)$, then

$$\boldsymbol{\nabla}\phi = x\hat{x} + y\hat{y} + z\hat{z} = r\hat{r} \tag{43}$$

is a radially outward pointing vector field on $\mathbb{R}^3$, with magnitude equal to the distance from the origin.

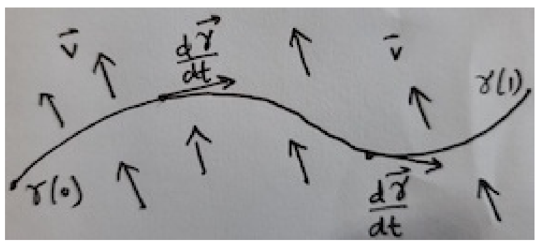
• At any location $\boldsymbol{r}$, $\boldsymbol{\nabla}\phi$ is a vector that points in the direction of most rapid increase of ϕ. To see why, it is helpful to introduce the level surfaces of ϕ, which are surfaces in $\mathbb{R}^3$ on which ϕ is a constant. For ϕ to change most rapidly, we must move from $\boldsymbol{r}$ along a vector that has no component along the level surface through $\boldsymbol{r}$. We will argue that at any point $\boldsymbol{r}$, $\boldsymbol{\nabla}\phi(\boldsymbol{r})$ is orthogonal to the level surface through $\boldsymbol{r}$. Suppose $\boldsymbol{v}$ is a vector at $\boldsymbol{r}$ of small magnitude, then the linear Taylor approximation gives $\phi(\boldsymbol{r}+\boldsymbol{v}) \approx \phi(\boldsymbol{r}) + \boldsymbol{v}\cdot\boldsymbol{\nabla}\phi$. Now, $\boldsymbol{v}$ is tangent to the level surface through $\boldsymbol{r}$ if $\phi(\boldsymbol{r}+\boldsymbol{v}) - \phi(\boldsymbol{r})$ vanishes to first order in $\boldsymbol{v}$. This happens precisely when $\boldsymbol{v}\cdot\boldsymbol{\nabla}\phi = 0$. Thus, $\boldsymbol{\nabla}\phi$ must be perpendicular to the level surface of ϕ and must point either in the direction of most rapid increase or decrease of ϕ. Taking $\boldsymbol{v} = \epsilon\boldsymbol{\nabla}\phi$ for $0 < \epsilon \ll 1$, we find that $\phi(\boldsymbol{r}+\epsilon\boldsymbol{\nabla}\phi) \approx \phi(\boldsymbol{r})+\epsilon|\boldsymbol{\nabla}\phi|^2 > \phi(\boldsymbol{r})$. Thus we conclude that $\boldsymbol{\nabla}\phi$ must point in the direction of most rapid increase of ϕ.

• If ϕ is regarded as a potential function, then its level surfaces are referred to as equipotential surfaces.

• E.g., For $\phi(x,y) = x$, the level curves are lines parallel to the y axis, and $\boldsymbol{\nabla}\phi = \hat{x}$ points perpendicular to these lines in the direction of most rapid increase in ϕ. For $\phi = \frac{1}{2}(x^2+y^2+z^2)$, the level surfaces are concentric spheres centered at the origin and $\boldsymbol{\nabla}\phi = \boldsymbol{r}$ is perpendicular to these surfaces.

• **Line integral.** Given a vector field $\boldsymbol{v}(\boldsymbol{r}) = (v_x, v_y, v_z)(\boldsymbol{r})$ in 3d space and a parametrized curve $\boldsymbol{\gamma}(t) = (x(t), y(t), z(t))$ for $0 \le t \le 1$, we may define the 'line integral' of $\boldsymbol{v}$ along γ as the real number

$$\int_\gamma \boldsymbol{v}\cdot d\boldsymbol{\gamma} = \int_0^1 \boldsymbol{v}\cdot \frac{d\boldsymbol{\gamma}}{dt} dt = \int_0^1 \left[v_x \frac{dx}{dt} + v_y \frac{dy}{dt} + v_z \frac{dz}{dt} \right] dt. \qquad (44)$$



• Here, $\dot{\boldsymbol{\gamma}} = \frac{d\gamma}{dt} = \dot{x}\hat{x} + \dot{y}\hat{y} + \dot{z}\hat{z}$ is a vector field along the curve $\boldsymbol{\gamma}$ (it is not defined elsewhere in $\mathbb{R}^3$). At each fixed t, it is the tangent vector to the curve at the point $\boldsymbol{\gamma}(t)$.

• For example, if $\boldsymbol{\gamma}$ is the helix $(\cos t, \sin t, t)$, then $d\boldsymbol{\gamma} = (-\sin t, \cos t, 1)dt$. We may consider $d\gamma$ as the differential of the map $\gamma : [0,1] \to \mathbb{R}^3$.

• The work done by a force field $\boldsymbol{F}(\boldsymbol{r})$ in moving a particle along a curve γ is an important example of a line integral: $W_{\boldsymbol{F}}(\gamma) = \int_\gamma \boldsymbol{F}\cdot d\boldsymbol{\gamma}$.

• In general, the line integral depends on the values of $\boldsymbol{v}$ all along the curve $\boldsymbol{\gamma}$. However, if $\boldsymbol{v}$ is the gradient of a scalar, $\boldsymbol{v} = \boldsymbol{\nabla}\phi$, then the line integral can be evaluated in terms of the values of ϕ at the endpoints:

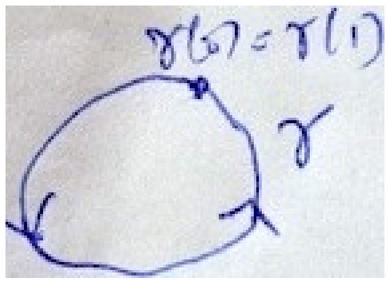
$$\begin{aligned} \int_\gamma \boldsymbol{\nabla}\phi \cdot d\boldsymbol{\gamma} &= \int_0^1 \left(\frac{\partial\phi}{\partial x}\frac{dx}{dt} + \frac{\partial\phi}{\partial y}\frac{dy}{dt} + \frac{\partial\phi}{\partial z}\frac{dz}{dt} \right) dt \\ &= \int_0^1 \frac{d\phi(\boldsymbol{r}(t))}{dt} dt = \phi(\boldsymbol{r}(1)) - \phi(\boldsymbol{r}(0)). \end{aligned} \qquad (45)$$

Here, we viewed $\phi(x(t), y(t), z(t))$ as a function of t and used the chain rule to differentiate it with respect to t.

• In particular, if γ is a closed curve, then $\boldsymbol{r}(0) = \boldsymbol{r}(1)$ and the line integral of a gradient vanishes

$$\oint_\gamma \boldsymbol{\nabla}\phi \cdot d\boldsymbol{\gamma} = 0. \tag{46}$$

Here $\oint$ denotes a line integral around a closed contour.



• A vector field that is the gradient of a scalar field is called a gradient vector field. In mechanics, if a force field $\boldsymbol{F}(\boldsymbol{r})$ is the gradient of a scalar field (or 'potential' $\phi(\boldsymbol{r})$), then it is called a conservative force field. The work done by a conservative force field $\boldsymbol{\nabla}\phi$ depends only on the initial and final locations of the particle, and not on the rest of the details of the path taken. A conservative force field does no work in moving a particle around a closed curve.

2 Newton's laws and forces

2.1 Time, light, simultaneity, space & time intervals, masses.

• To describe the dynamics (motion or more precisely the evolution in time) of mechanical systems, observers find it helpful to have a notion of time (measured with a clock) to index a sequence of events.

• In Newtonian mechanics, one assumes that if there is a flash of light somewhere, then all observers (irrespective of their locations) receive the flash instantaneously. In effect, light is assumed to travel infinitely fast.

• Using such flashes of light, all observers can synchronize their clocks and assign the same time for a given event.

• Another consequence is that two events (possibly at different locations) that occur at the same time for one observer occur simultaneously for any other observer. We say that simultaneity is absolute, not relative.

• These assumptions about light, time and simultaneity were in line with common human experience in Newton's time (as well as today!).

• In Newtonian mechanics, one also assumes that masses of particles, scales of length and time are the same for all observers. In other words, distances (like the length of a meter stick) and time intervals (like that between two ticks of a clock) are the same for all observers.

• **Causality.** Our experience with physical systems indicates that they respect the principle of causality: cause precedes effect. For example, a stone that is stationary is seen to move when it is pushed and not before that.

• The principle of causality postulates that it is not possible to send a signal from an event to its past. Given our Newtonian concept of time, all observers have a common notion of the past and future of an event. The future of an event that occurs at time t_0 consists of all events that occur at $t > t_0$ and the past consists of events that occurred at $t < t_0$.

• These notions of time, simultaneity and universality of masses, space and time intervals had to be discarded and replaced with more accurate concepts in the special theory of relativity, where speeds of bodies or observers could be comparable to that of light, which is a large but finite constant in vacuum ($c \approx 3 \times 10^8$ m/s).

• Newtonian or nonrelativistic mechanics is a limiting case of special relativistic mechanics where the speed of light is infinite (very large compared to other speeds). The principle of causality continues to apply in special relativity, though the notions of past and future need to be revised.

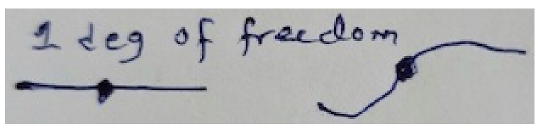
2.2 Degrees of freedom, instantaneous configurations, trajectories

• A point particle moving in three-dimensional space has three *degrees of freedom*: we need three coordinates (say (x, y, z) or (r, θ, ϕ)) to specify the location of the particle at the initial instant of time. We could locate the particle anywhere initially, so x, y and z can be chosen arbitrarily at the initial instant of time.

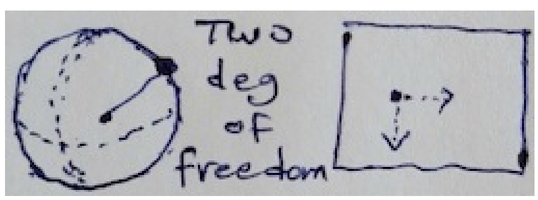
• For a system of particles, the number of degrees of freedom is the number of real parameters (coordinates) needed to specify the locations of all the particles in the system at the initial instant of time.

• The number of degrees of freedom does not depend on the nature of forces. An isolated particle, i.e., a particle that is not subject to any external influences (feels no force) is called a free particle. A free particle and a particle subject to Earth's gravity moving in three-dimensional space both have three degrees of freedom.

• On the other hand, a point particle that is constrained to move along a fixed wire has only one degree of freedom. We need one coordinate, say the distance (arc length) from one end of the wire, measured along the wire to specify the location of the particle at a given instant of time.



• A particle constrained to move on a spherical surface (such as a bob at the end of a rigid rod whose other end is attached to a pivot) has only two degrees of freedom.



• Two point masses moving in three dimensional space have six degrees of freedom. We need six coordinates (x_1, y_2, z_1) and (x_2, y_2, z_2) to specify the locations of the two particles. E.g., the Sun and the Earth regarded as point masses is a system with 6 degrees of freedom. Here, we do not restrict to a particular orbit of the Earth around the Sun but ask how many coordinates are needed to specify all possible locations of the Sun and the Earth at any fixed instant of time, without reference to the nature of the force between the two.

• A general rigid body like a stone has six degrees of freedom. For convenience, we may enumerate them as follows: 3 translational degrees of freedom to fix the location of a marked point in the body and 3 rotational degrees of freedom to orient the body holding the marked point fixed.

• A fluid consisting of $N \sim 10^{24}$ molecules in a bucket has a very large number of degrees of freedom, which can be taken to be the $3N$ Cartesian coordinates needed to specify the instantaneous locations of the N molecules, treated as point masses.

• An *instantaneous configuration* of a system of two point particles is any possible location of the two particles.

• **Zeroth law of classical mechanics.** The path followed by a particle in time is called its trajectory. It is a curve parametrized by time and directed towards increasing time. The *zeroth law of mechanics* can be regarded as saying that the trajectory $\boldsymbol{r}(t)$ of a particle is a (twice) differentiable function of time.

• This not an assumption but rather an assertion about natural phenomena, deduced by observing the motion of terrestrial and celestial bodies. This assertion applies to the motion of planets, pendulum bobs, cricket balls etc. But it fails for Brownian motion (movement of pollen grains in water, which are observed to follow very jagged paths). It also fails for electrons in an atom, which require a quantum mechanical treatment.

• Isaac Newton formulated three laws of classical mechanics in his Principia (1687).

2.3 Newton's 1^{st} law

• Newton's 1^{st} law, or the law of inertia, says that "Every body continues in its state of rest, or of uniform motion in a straight line, unless it is compelled to change that state by a force impressed upon it". In other words, the momentum $\boldsymbol{p} = m\boldsymbol{v} = m\dot{\boldsymbol{r}}$ of a particle that is free (isolated or far from physical interactions) does not change with time. Note that if $\dot{\boldsymbol{r}}$ does not change with time, i.e., $\ddot{\boldsymbol{r}} = 0$, then the trajectory $\boldsymbol{r}(t) = \boldsymbol{r}(0) + \dot{\boldsymbol{r}}(0)t$ is a straight line that is uniformly traversed.

• In general, it is found that macroscopic interactions decrease with distance, so it is possible to isolate a particle by taking it far from other bodies.

• We have been a bit imprecise in our statement of Newton's first law. Newton's first law generally holds only in certain reference frames. To specify the components of the position and velocity vectors, we need a frame of reference, i.e., an origin and coordinate axes.

• It is found that a particle that is not subject to any forces (i.e., an isolated body) could fail to follow a constant velocity trajectory in certain reference frames.

• A frame in which Newton's first law (as stated above) holds is said to be an inertial frame. To a reasonable approximation (if one ignores some effects of the rotation of the Earth), a frame that is fixed in a tennis court is an inertial frame for the motion of tennis balls, racquets etc. In particular, if the effects of gravity and friction are ignored (or somehow cancelled, see below), then tennis balls in this frame would always move uniformly in straight lines. Note that this same frame may not be considered inertial for describing certain other phenomena such as the motion of planets!

• It is possible to (essentially arbitrarily) reduce the effects of gravity and friction in some cases. Consider the horizontal motion of a ball on a ping-pong table placed in the tennis court. A ball that starts out at rest on the table is seen to remain at rest. A ping-pong ball that starts out rolling on the table continues to roll in the same direction though it is seen to slow down. We attribute this to friction with the surface. The magnitude of this frictional force depends on the weight of the ball, but by polishing the surface and ball we may reduce the effect of friction and find that the ball maintains its velocity for a long time. In this way, we arrive at a body that is essentially free in so far as its horizontal motion is concerned. Its uniform motion is evidence in favor of the inertial nature of a frame fixed to the Earth.

• However, it is found that Newton's first law for tennis balls fails to hold in a frame that is attached to a swinging pendulum or a rotating merry-go-round beside the tennis court. Such frames are called accelerated or noninertial.

• For instance, a frame that is attached to a bee as it flies irregularly in a faraway spaceship is not inertial, since a free particle at rest in the same spaceship would appear to move in a nonuniform manner.

• Note: here, we use the metaphor of the bee for the limited purpose of defining a frame that moves nonuniformly relatively to the spaceship. A flying bee is not a free particle - it does not move uniformly, it makes use of its internal energy and friction with the air to change direction, speed up or maintain its speed etc.

• Similarly, a frame that is attached to a top (spinning on the floor of the spaceship) and participates in its rotational motion is noninertial.

• To summarize, Newton's first law is the assertion that there is a frame of reference (called an inertial frame) in which all isolated bodies (far from physical interactions) move at constant velocity.

• As we will see shortly, from a principle enunciated by Galileo, an inertial frame is not unique.

• Henceforth, unless otherwise stated, all quantities will be specified with respect to an inertial frame of reference.

• There are indirect ways to check whether a frame is inertial even if one cannot isolate particles. This makes use of Newton's second law (which is a statement about inertial frames) and its consequences. Roughly, suppose we assume a frame is inertial, deduce consequences using Newton's second law and find that they are experimentally violated. Then, one possible reason for the discrepancy can be that the frame was not inertial to begin with. The Foucault pendulum gives a concrete realization of this idea and strongly suggests that the Earth is not quite an inertial frame, due to its rotation on its axis.

2.4 Newton's 2^{nd} law

• The departure from rest or uniform motion along a straight line (in an inertial frame) is caused by forces. For example, tugging at a string that is attached to a ball exerts a force on the ball and makes it accelerate.

• Forces typically arise from interactions between objects. The Earth exerts a force on a ball that is dropped, making it accelerate downwards.

• Newton's 2^{nd} law says that in an inertial frame, the rate of change of momentum $\dot{\boldsymbol{p}}$ is equal to the impressed force. In particular, it is in the direction in which the force acts.

• For a single particle of mass m, the acceleration $\boldsymbol{a} = \ddot{\boldsymbol{r}} = \dot{\boldsymbol{p}}/m$ along the trajectory $\boldsymbol{r}(t)$, due to the force $\boldsymbol{F}$ is determined by the 'equation of motion'

$$m\ddot{\boldsymbol{r}} = \boldsymbol{F} \quad \text{or} \quad \dot{\boldsymbol{p}} = \boldsymbol{F}. \tag{47}$$

The mass (more precisely, inertial mass) m of the particle is postulated to be independent of time.

• We may use Newton's 2nd law to give a way of assigning inertial masses to bodies. We begin by selecting a reference body A and choose units in which its inertial mass is assigned the value 1. We apply a force to A (e.g., let a compressed spring push it), and record its acceleration $\boldsymbol{a}_A$. Given body B whose inertial mass we wish to determine, we apply the same force to it, and measure its acceleration $\boldsymbol{a}_B$. Then we assign the mass $m_B = |\boldsymbol{a}_A|/|\boldsymbol{a}_B|$.

• The force is generally a vector field $\boldsymbol{F}(\boldsymbol{r})$, it could depend on the location of the particle. To begin with, the force field may not be known to us, so we do experiments with particles, observe their trajectories (measure their accelerations) and thereby deduce what the force field may be. Having done some such experiments, we develop a formula or picture of the force field. This is called the inverse problem: determination of the force from observed motion of particles. Having done this to our satisfaction, we may then make predictions of what a given particle may do when subjected to this force field by solving Newton's equation with prescribed initial conditions and the available information on $\boldsymbol{F}(\boldsymbol{r})$. This latter problem is called the direct problem: finding trajectories given a force field. We then compare these predicted trajectories with new observations to validate our formula/picture for the force field. If discrepancies are found, we may need to update our formula for the force field. Thus, one goes back and forth between the inverse and direct problems.

• In *Cartesian coordinates*, the trajectory is given by $\boldsymbol{r}(t) = (x^1, x^2, x^3) = (x, y, z)$ and Newton's second law becomes $m\ddot{x}^i = F^i$. It is conventional to use superscripts for coordinates, here $y = x^2$ is not the square of x. This component form of Newton's equation changes in curvilinear coordinates, such as spherical polar coordinates (i.e., it does not simply say $m\ddot{r} = \boldsymbol{F} \cdot \hat{r}$ etc.). For instance, there could be terms involving products of first derivatives of coordinates in addition to naive second derivative 'acceleration' terms as we found in plane polar coordinates:

$$m(\ddot{r} - r\dot{\theta}^2) = \boldsymbol{F} \cdot \hat{r} \quad \text{and} \quad m(r\ddot{\theta} + 2\dot{r}\dot{\theta}) = \boldsymbol{F} \cdot \hat{\theta}. \tag{48}$$

One may transform the equation from Cartesian coordinates to the desired system to find the form it takes.

• Being 2nd order in time, Newton's equation requires both the initial position $\boldsymbol{r}$ and velocity or momentum ($\dot{\boldsymbol{r}}$ or $\boldsymbol{p}$) as initial conditions. For a particle with 3 degrees of freedom, these would amount to 6 pieces of initial data (6 real numbers), say $x(0), y(0), z(0)$ and $p_x(0), p_y(0), p_z(0)$. The knowledge of the current position and momentum determines the trajectory via Newton's 2nd law. Bearing this in mind, we define the *state* of the particle as being specified by giving its instantaneous position *and* momentum. Thus, the knowledge of the current state of the particle along with Newton's second law determines its future evolution (trajectory). For instance, for a particle moving on a line subject to a force field $f(x)$, if we know $x(t)$ and $p(t)$, then at the next instant of time, $p(t + \delta t) \approx p(t) + (\delta t) f(x(t))$ and $x(t + \delta t) \approx x(t) + (\delta t) p(t)/m$.

• The path of the particle $\boldsymbol{r}(t)$ (satisfying Newton's equation and initial conditions) is called its trajectory. Trajectories are oriented by arrows specifying forward time evolution.

2.5 Galileo's relativity principle, space-time homogeneity and isotropy of space

• Notice that Newton's 2^{nd} law relates the force to the *second derivative* of position along a trajectory, as opposed to the first, third or other derivative. This is to incorporate Galileo's relativity principle which says roughly that there is no dynamical way of telling if a frame is at rest or moving uniformly relative to an inertial frame.

• In 1632, Galileo Galilei observed that it was not possible to detect the uniform motion (constant velocity motion without rocking) of a ship relative to the shore by performing mechanical experiments under the deck of the ship (i.e., without looking out or by using external forces etc.).

• These experiments could include observing the motion of projectiles (e.g. how long it takes for a ball thrown horizontally at a given speed to reach a wall), the manner in which water drips from a jug, how flies and fish move and so on.

• It is important to note that Galileo's principle is not concerned with forces external to the lab. In other words, if the acceleration due to gravity varies with location, then it may be possible for an observer in the uniformly moving ship to infer that the ship's frame is moving relative to the shore. Thus, Galileo's principle asserts that the relative motion of bodies in the lab is the same whether observed when the ship is docked or when it is uniformly moving.

• This idea is elevated to the principle of Galilean relativity, which states that the laws of mechanics must take the same form in two inertial frames that are in uniform motion relative to each other.

• Galileo's principle of relativity continues to hold in special relativity. To accommodate the constancy of the speed of light, Einstein modified the transformation rule that relates coordinates in two frames that are in uniform motion relative to each other.

• In Galilean relativity, it is assumed that

(a) the mass of a particle is the same in two frames that are in uniform relative motion,

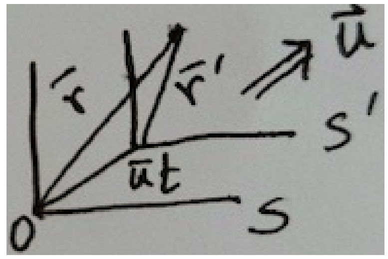
(b) both observers use the same scale for measuring distances and

(c) both observers agree on the time interval between any pair of events.

• In other words, uniformly moving measuring sticks have the same length as when they are observed at rest and a clock that is moving at a constant velocity neither slows down nor speeds up relative to a clock at rest.

• With these assumptions, the appearance of acceleration $\ddot{\boldsymbol{r}}$ (rather than velocity $\dot{\boldsymbol{r}}$) in Newton's 2^{nd} law can be motivated. It ensures that when referred to a frame S' moving at constant velocity $\boldsymbol{u}$ relative to an inertial frame S, Newton's 2^{nd} law takes the same form for a system of (interacting) particles.

• For example, suppose the frames coincide at $t = 0$ so that $\boldsymbol{r}' = \boldsymbol{r} - \boldsymbol{u}t$. Then $\ddot{\boldsymbol{r}} = \ddot{\boldsymbol{r}}'$ as $m\frac{d^2}{dt^2}(\boldsymbol{u}t) = 0$. Then the equation for a free particle is the same in both frames $m\ddot{\boldsymbol{r}} = 0$ and $m\ddot{\boldsymbol{r}}' = 0$. A particle is free in one frame



iff it is free in the other.

• Similarly, Newton's second law for a pair of particles subject to an interparticle force that depends on the relative distance between the two particles,

$$m\ddot{\boldsymbol{r}}_1 = \boldsymbol{F}_{2 \text{ on } 1}(|\boldsymbol{r}_1 - \boldsymbol{r}_2|) \quad \& \quad m\ddot{\boldsymbol{r}}_2 = \boldsymbol{F}_{1 \text{ on } 2}(|\boldsymbol{r}_2 - \boldsymbol{r}_1|) \qquad (49)$$

takes the same form in frame S' with $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ replaced with $\boldsymbol{r}_1'$ and $\boldsymbol{r}_2'$ since

$$\boldsymbol{r}_1 - \boldsymbol{r}_2 = \boldsymbol{r}_1' + \boldsymbol{u}t - \boldsymbol{r}_2' - \boldsymbol{u}t = \boldsymbol{r}_1' - \boldsymbol{r}_2'. \qquad (50)$$

The velocity $\boldsymbol{u}$ cancels out from the difference in position vectors. The relative velocity between the frames $\boldsymbol{u}$ makes no appearance.

• $\boldsymbol{u}$ also cancels out from the difference in velocity vectors $\dot{\boldsymbol{r}}_1 - \dot{\boldsymbol{r}}_2$ so the same conclusion also applies to interparticle forces that depend on velocities (such as friction).

• Thus, practically speaking, (1) projectiles move in exactly the same way when observed in two frames in uniform relative motion (ignoring a possible variation of the external gravitational acceleration) and (2) a brick sliding on a plank subject to friction displays the same dynamics irrespective of whether the experiment is performed below the deck of a docked ship or a uniformly moving ship.

• Newton's 2nd law for a particle subject to an *external* force $m\ddot{\boldsymbol{r}} = \boldsymbol{F}(\boldsymbol{r})$ however is not the same in the two frames. It becomes $m\ddot{\boldsymbol{r}}' = \boldsymbol{F}(\boldsymbol{r}' + \boldsymbol{u}t)$. The appearance of $\boldsymbol{u}$ in this frame would mean that one could find out which frame corresponds to the moving ship and which to the docked ship. For instance, variation in the acceleration due to gravity at different locations could be used to determine that the experiment was performed in a moving ship rather than at a fixed location on the shore. This does not violate Galileo's principle of relativity since the latter is not concerned with external forces but with interparticle forces.

• If Newton's 2nd law for a particle involved velocity instead of acceleration, say $\nu\dot{\boldsymbol{r}} = \boldsymbol{G}$, then in a frame moving at velocity $\boldsymbol{u}$, the equation for a free particle would take the form $\nu(\dot{\boldsymbol{r}}' + \boldsymbol{u}) = 0$. The appearance of $\boldsymbol{u}$ on the left of this equation implies that it does not have the same form as the equation $\nu\dot{\boldsymbol{r}} = 0$ in S and would allow us to determine the velocity $\boldsymbol{u}$ of the frame S' relative to S and thereby tell the frames apart, in violation of Galileo's principle.

• Relating the force to the second derivative of position (as opposed to, say, the third derivative of $\boldsymbol{r}(t)$) is the simplest way of incorporating Galileo's principle.

• Fortunately, experiments and observations confirm that Newton's 2^{nd} law accurately describes both terrestrial and celestial mechanical phenomena (motion of tennis balls, planets etc.), so there is no need to include higher time derivatives in Newton's second law, although they would not violate Galileo's principle.

• There are other reasons to avoid higher time derivatives on the LHS of Newton's equation. Indeed, suppose Newton's second law for the position $x(t)$ of a particle moving along a line had a 3rd derivative term: $m\ddot{x} + \nu\dddot{x} = f$, for some constant $\nu \neq 0$, where f is the force. Now, consider a free particle, $f = 0$. The equation $m\ddot{x} + \nu\dddot{x} = 0$ may be integrated once to get $m\dot{x} + \nu\ddot{x} = \alpha$ and integrated a second time to get $mx + \nu\dot{x} = \alpha t + \beta$ for constants of integration α, β. It can be shown that the solution of this first order equation is

$$x(t) = \gamma e^{-mt/\nu} + m^{-2}\left[m(\beta + \alpha t) - \alpha\nu\right]. \tag{51}$$

where γ is a third constant of integration. For $\gamma, \nu \neq 0$, this is clearly not of the constant speed form $x(0) + v(0)t$. Thus, in violation of Newton's 1st law, this free particle trajectory does not have constant speed. We conclude that a third derivative term on the left of Newton's 2nd law equation would not be consistent with Newton's first law.

• Can the LHS of Newton's equation $m\ddot{\boldsymbol{r}} = \boldsymbol{F}$ include a term such as $\lambda\boldsymbol{r}$ for some constant λ? No, for more than one reason. (a) This would violate Newton's first law, free particles would not always follow straight line trajectories with constant speed. Indeed for $\lambda \neq 0$, $m\ddot{x} + \lambda x = 0$ does not admit the constant speed solution $x(t) = vt + x_0$. (b) It would violate the homogeneity of space which requires that the laws of mechanics be the same at all locations. Given the same external conditions, the results of mechanical experiments do not depend on where they are performed. Mathematically, Newton's 2nd law equation must be translation-invariant. Suppose we make a translation $x' = x + a$, then $m\ddot{x} + \lambda x = f$ would become $m\ddot{x}' + \lambda(x' - a) = f$, so that Newton's equation would not take the same form in a frame that is shifted by distance a relative to the original frame. Note that interparticle forces are translation invariant, since they depend on the relative locations of particles; the problem lies in the appearance of a on the LHS of the equation of motion in the shifted frame.

• We also postulate that the laws of mechanics do not pick out any particular direction at any given location. We say that space is isotropic. The orientation of a frame has no dynamical significance. Holding external conditions the same, rotating the experimental apparatus does not change the results of experiments.

• Along with homogeneity and isotropy of space, we also postulate the homogeneity of time. Given identical external conditions, the results of mechanical experiments must not depend on when they are done. In other words, the equations of mechanics must be invariant under translations of time $t \to t' = t + t_0$. For instance, the masses of a pair of particles that interact through interparticle forces cannot change with time. Since $\frac{d}{dt} = \frac{d}{dt'}$, Newton's equation of motion $m\ddot{\boldsymbol{r}} = \boldsymbol{f}$ takes the same form (for a force that is not explicitly time-dependent) whether we use t or t'.

• Space rotation invariance, space and time translation invariance along with the invariance under a change from an inertial frame to one moving at constant velocity are together termed the Galilean invariances of the laws of mechanics.

2.6 Linear superposition of forces

• Suppose a particle is acted upon by two forces $\boldsymbol{F}_1$ and $\boldsymbol{F}_2$. Then according to the superposition principle, the total force $\boldsymbol{F}$ on the particle or resultant of the two forces is the vector sum $\boldsymbol{F} = \boldsymbol{F}_1 + \boldsymbol{F}_2$. This is to be expected from Newton's 2nd law $m\boldsymbol{a} = \boldsymbol{F}$: acceleration is a vector and so force must also be a vector and vectors can be added using the parallelogram law to obtain the total force. Newton does not mention the superposition principle for forces as a separate law, but states it as Corollary 1.

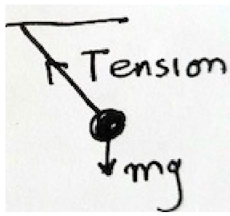
• It is noteworthy that Newton had more than three laws (about 5 or 6) in his manuscript *De motucorporum in mediis regulariter cedentibus* that he wrote a couple of years before the Principia. Some postulates/laws, such as a version of the principle of relativity were later demoted to corollaries of what we now call his three laws of motion.

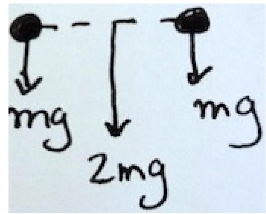
• The superposition principle is very useful. It allows us to separately determine individual forces on a body, which may have distinct origins (gravitational, frictional, electric etc.) before adding them up to find the total force.

• Note that we do not have a superposition principle for solutions of Newton's equation in general. For example, suppose $m\ddot{\boldsymbol{r}}_1 = \boldsymbol{f}_1(\boldsymbol{r}_1)$ and $m\ddot{\boldsymbol{r}}_2 = \boldsymbol{f}_2(\boldsymbol{r}_2)$ are trajectories in the presence of individual forces. Then putting $\boldsymbol{r} = \boldsymbol{r}_1 + \boldsymbol{r}_2$, we get $m\ddot{\boldsymbol{r}} = \boldsymbol{f}_1(\boldsymbol{r}_1) + \boldsymbol{f}_2(\boldsymbol{r}_2)$. However, the latter is generally *not* equal to the vector field $\boldsymbol{f}_1 + \boldsymbol{f}_2$ evaluated at $\boldsymbol{r}_1 + \boldsymbol{r}_2$. So we cannot in general 'add' trajectories in the presence of separate forces to get a trajectory when both forces are present. The 1d example of a superposition $f = -kx + c$ of a linear restoring force $f_1 = -kx$ and a constant force $f_2 = c$ provides a counterexample. Suppose $m\ddot{x}_1 = -kx_1$ and $m\ddot{x}_2 = c$ are trajectories in the presence of the separate forces. Verify that the sum of these trajectories $x = x_1 + x_2$ satisfies $m\ddot{x} = -kx_1 + c$ which differs from the desired equation $m\ddot{x} = -kx + c = -k(x_1 + x_2) + c$. In general, it is meaningless to add solutions to Newton's equation: they typically do not form a linear space.

• Moreover, the superposition principle does not say that we can superpose solutions to Newton's equation for a given force field to get new solutions in the same force field. For instance, suppose $m\ddot{\boldsymbol{r}}_1 = \boldsymbol{f}(\boldsymbol{r}_1)$ and $m\ddot{\boldsymbol{r}}_2 = \boldsymbol{f}(\boldsymbol{r}_2)$, i.e., $\boldsymbol{r}_1(t)$ and $\boldsymbol{r}_2(t)$ are solutions of Newton's equation for the same force field $\boldsymbol{f}$. Then in general $\boldsymbol{r} = \boldsymbol{r}_1 + \boldsymbol{r}_2$ is not a solution of the equation

$m\ddot{\boldsymbol{r}} = \boldsymbol{f}(\boldsymbol{r})$. This is because the force could depend nonlinearly on the location.





• Similarly, suppose a composite body is made up of several constituent parts (e.g. a rigid body made of several point masses). Then the total force on the composite body is the vector sum of the forces on its constituents.

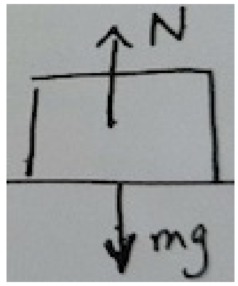
2.7 Newton's 3^{rd} law

Newton's 3^{rd} law says that 'to every action there is always opposed an equal reaction'. In other words, if body A exerts a force $\boldsymbol{F}$ on body B, then B exerts a force $-\boldsymbol{F}$ on A. These two forces are called the *impressed* and *expressed* forces. While the third law is *not* needed to understand the motion of a particle subject to given external forces, it *is* needed to understand the motion of bodies subject to interparticle forces. The third law concerns forces in an inertial frame.

• E.g. 1: The Sun attracts the Earth with a force equal in magnitude and opposite in direction to the force exerted by the Earth on the Sun.

• E.g. 2: A cubical block of concrete of mass m that lies on the floor exerts a downward force on the floor of magnitude equal to mg, where g is the magnitude of the acceleration due to gravity. (Note: this is not the force of the Earth on the block!) On the other hand, the floor exerts an upward 'normal reaction' force $\boldsymbol{N}$ of the same magnitude mg on the block. Notice that the impressed and expressed forces act on different bodies. In this example, both forces can be called normal surface forces, it is just conventional to call the force of the floor on the block by the name normal reaction force. The force of the block on the floor is equally well a normal reaction force. In a fluid, such equal and opposite normal surface forces between small neighboring volumes of fluid go by the name of pressure.

• For future reference. Newton's third law also helps us distinguish be-



tween a real force and a fictitious force in an accelerated or noninertial reference frame that we will encounter in §8.1. According to Newton's third law, an acceleration due to a real force felt by a body is distinguished by the presence of an equal and opposite reaction force on some other body.

2.8 Dynamics, kinematics and statics: what do they refer to?

• Dynamics refers to the evolution of a system in time. By this we mean the behavior of the system with the passage of time due to the forces in operation. To understand the dynamics, we need to know the forces and interactions present. An aim of dynamics is to find the trajectory by solving the equations of motion (EOM), given the initial conditions. By Newton's second law, doing this requires the knowledge of forces. The result of this exercise for some initial conditions could be that the parts of the system do not move while for other initial conditions the parts of the system may move. The equations of motion are also called the dynamical equations.

• Statics refers to situations where a system does not change with time. For example, a static solution of the EOM for a particle is one where $\boldsymbol{r}(t)$ is independent of time. To discover static solutions, we need to know the forces and solve the equations of motion.

• In a somewhat different direction, kinematics refers to those aspects of the system and its motion that do not depend on the specific forces that act, but on other features like the number of degrees of freedom, the range of values that physical variables can take, how these values change in different frames of reference, etc. The problem of finding the trajectory $x(t)$ from a known acceleration $\ddot{x}(t)$ function is kinematical, while finding the trajectory from the forces involves dynamics.

2.9 Dimensional analysis and units

• Physical quantities like the Cartesian components of the position of a particle, a time interval or a force can each be assigned a dimension. It turns out that the dimensions of all the mechanical quantities we encounter can be expressed in terms of three basic dimensions: mass M, length L and time T.

• More precisely, any quantity F can be assigned a dimension $[F] = M^\alpha L^\beta T^\gamma$ for some real numbers α, β, γ (which are typically rational numbers). If $\alpha = \beta = \gamma = 0$, the quantity is said to be dimensionless. Consistency requires that all terms in an equation have the same dimensions. This gives a quick way of eliminating some errors in equations.

• Examples of dimensions of physical quantities

$$\begin{aligned} [\text{mass}] &= M, \quad [\text{length}] = L, \quad [\text{time}] = T, \quad [\text{velocity}] = LT^{-1}, \\ [\text{acceleration}] &= LT^{-2}, \quad [\text{momentum}] = MLT^{-1}, \\ [\text{force}] &= MLT^{-2}, \quad [\text{energy}] = ML^2T^{-2}. \end{aligned} \tag{52}$$

• What is the dimension of an angle?

• **Units.** The most commonly used systems of units in mechanics are the SI and CGS systems.

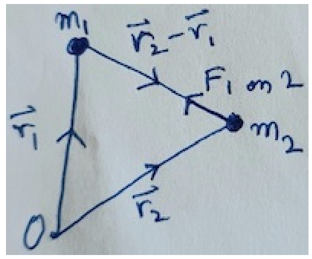
• In the International System of Units (SI), the basic units of length, mass and time are the meter, kilogram and the second. The unit of force is called the Newton $1N = 1$ kg m / s^2.

• In the CGS system, the corresponding base units are the centimeter, gram and second. The unit of force is called the dyne, (one Newton is 10^5 dyne).

2.10 Examples of forces

• Gravity. The gravitational force played a central role in the development of mechanics. Newton proposed his universal law of gravitation and developed his laws of mechanics in large part to understand the motion of the planets.

• Newton's law of gravity says that the force between two point masses m_1



and m_2 (called gravitational masses, more on this shortly) is attractive and proportional to the product of the masses and is inversely proportional to the square of the distance of separation. The constant of proportionality is Newton's gravitational constant G. If $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are the position vectors of the two particles, then the force exerted by m_1 on m_2 points from m_2 towards m_1 and is given by:

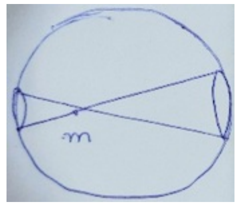
$$\boldsymbol{F}_{1\text{ on }2} = -G\frac{m_1 m_2}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|^3}(\boldsymbol{r}_2 - \boldsymbol{r}_1) \tag{53}$$

By Newton's third law, m_2 exerts an equal and opposite force on m_1 given by $\boldsymbol{F}_{2\text{ on }1} = -\boldsymbol{F}_{1\text{ on }2}$.

• **Principle of superposition of forces.** Newton's law of gravitation applies to point particles. However, it may be used to calculate the force due to extended bodies on point masses or on other extended objects. In order to do this, one repeatedly uses the principle of linear superposition, which states that the force due to two objects on a given particle is the vector sum of the individual forces. Furthermore, the force on a composite body is the vector sum of forces on all its constituents.

• By using the principle of superposition, one may show that the force on a point mass m that lies outside a spherically symmetric mass distribution is the same as the force due to a point particle (with mass equal to the total mass M of the distribution) located at the center of the distribution.

• For example, the force due to a thin spherical shell of radius R and mass M centered at the origin on a point mass m located at $\boldsymbol{r}$ (with $r > R$) is given by $-GmM\frac{\hat{\boldsymbol{r}}}{|\boldsymbol{r}|^2}$. It turns out that the force vanishes inside the shell due to a cancellation of forces due to diametrically opposite parts of the spherical shell. These diametrically opposite surface elements have mass proportional to their area, which grows as the square of the distance from the apex of a cone. This quadratic growth is cancelled by the inverse square fall off in Newton's gravitational force so that the two elements exert equal and opposite forces on a mass inside the shell.



• This is an opportunity so say a bit more on the concept of mass. The mass m_i that appears in Newton's second law $\boldsymbol{F} = m_i \boldsymbol{a}$ is called the inertial mass. Using Newton's 2nd law, we have described a way of assigning inertial masses to bodies, relative to that of a reference body.

• On the other hand, the masses that appear in Newton's law of gravity $\boldsymbol{F}_{2 \text{ on } 1} = G m_{1g} m_{2g} \hat{r}/r^2$ are the gravitational masses, with $\boldsymbol{r} = \boldsymbol{r}_2 - \boldsymbol{r}_1$. (We will say more on this in the context of Kepler's laws of planetary motion.)

• Let us now focus on the gravitational force of the Earth on small bodies near its surface. At the surface of the Earth, the force on a gravitational mass m_g is $\boldsymbol{F} = -G M_e m_g \hat{\boldsymbol{r}}/R_e^2$ where M_e is the gravitational mass of the Earth and R_e the radius of the Earth. The acceleration due to gravity of the body is then $\boldsymbol{a} = \boldsymbol{F}/m_i = -(G M_e (m_g/m_i)/R_e^2)\hat{r}$. This downward pointing acceleration is denoted $\boldsymbol{g}$. On the face of it, the acceleration due to gravity depends on the body through its inertial and gravitational mass.

• **Equivalence principle.** The surprising experimental observation is that the acceleration due to gravity is the same for all bodies. The story of different objects falling in unison from the leaning tower of Pisa is a way to remember this. The magnitude of this acceleration g is approximately $9.8\ m/s^2$.

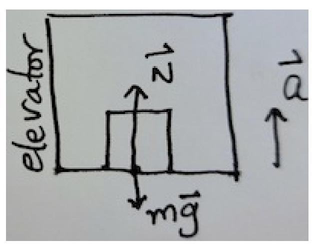
• Thus, the ratio m_g/m_i must be the same for all bodies. By absorbing this constant ratio into G, we arrive at the equality $m_i = m_g$. This is called the principle of equivalence of inertial and gravitational masses. Since we have already assigned inertial masses to bodies, the Equivalence principle gives a way of assigning gravitational masses as well. Henceforth we will not make a distinction between inertial and gravitational masses.

• The **weight** of a body is defined as the Earth's gravitational force acting on it. At the the surface of the Earth, the weight of a body of mass m

is $\boldsymbol{W} = -(GM_e m/R_e^2)\hat{r} = m\boldsymbol{g}$. Since weight is a force, it is measured in Newtons in SI units. The weight, as defined above, is independent of the motion of the body.

• **Normal reaction.** There is a concept related to the weight of a body that is sometimes confused with it. Consider a body of mass m at rest on the Earth. Balance of forces in the vertical direction implies that the floor must exert a 'normal reaction' force $\boldsymbol{N} = -m\boldsymbol{g}$ upwards on the body. This is the force that the floor exerts to support the body.

• Now suppose the body is in an **elevator that accelerates** upwards at the rate $\boldsymbol{a}$. The force due to gravity on the mass is still its weight $\boldsymbol{W} = m\boldsymbol{g}$. Then Newton's second law implies that $m\boldsymbol{a} = \boldsymbol{N} + \boldsymbol{W}$ or $\boldsymbol{N} = m(\boldsymbol{a} - \boldsymbol{g})$. Since $\boldsymbol{N}$ and $\boldsymbol{a}$ point upwards, while $\boldsymbol{g}$ points downwards, the magnitude of the normal reaction is $N = m(a+g)$. This is the force that the floor of the elevator must exert upwards to support the body. The magnitude of $\boldsymbol{N}$ exceeds the magnitude of the weight of the body. A 'weighing' scale is usually calibrated to read the value $|\boldsymbol{N}|/g$.



• **Electrostatic force.** Charged particles exert electrostatic forces on each other. They are found to attract or repel. The force between point charges is summarized in Coulomb's law, which is very similar to Newton's law of gravitation between point masses, it is proportional to the product of electric charges and falls off inversely with the square of the distance of separation. However, while masses always attract, like charges (of the same sign) repel and unlike charges (with opposite signs) attract.

• Many forces we are familiar with have their microscopic origin in electrostatic forces between molecules. The frictional force between a body and a surface, the viscous force between layers of a fluid or between a fluid and a body moving through it are macroscopic manifestations of electric forces. The force that a stretched string exerts also has its microscopic origin in electric forces between molecules.

• For this reason, the electric force is called a fundamental force while friction and viscosity are called emergent or phenomenological forces.

• Friction and viscosity are also examples of contact forces, exerted when bodies are in contact.

• Though these forces arise from electrostatic interactions, it is often not practical to deduce their strength from microscopic considerations. They are usually described via effective macroscopic formulae based on experimental measurements like Hooke's law for a spring.

• For instance,

1. The drag force on a sphere moving slowly through a fluid is approximately proportional to its speed and points in a direction opposite to its velocity.
2. Neighboring segments of a stretched string are found to exert a tensional force on each other tending to elongate each segment. The tension can usually be modeled as a constant force along the length of the stretched string.

• Aside from gravity and electromagnetic forces, there are two more fundamental forces: the weak and strong nuclear forces. The latter are very short-ranged and act typically over nuclear and subnuclear scales. They are responsible for radioactive decay and for binding neutrons and protons in nuclei.

3 Momentum, Energy, Work, Angular momentum, Dynamical variables

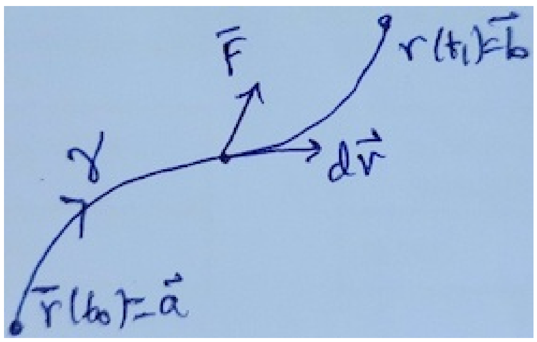
• The (linear) momentum of a particle of mass m moving at velocity $\boldsymbol{v}$ is defined as $\boldsymbol{p} = m\boldsymbol{v}$. If there is no force, then each of the components p_x, p_y, p_z of momentum is conserved, since $\dot{\boldsymbol{p}} = \boldsymbol{F} = 0$ (this is Newton's first law). If the force only acts downwards, then the horizontal components of momentum p_x, p_y are conserved.

3.1 Work done by a force and conservative forces

• **Work done by a force.** Suppose a particle is moved from position $\boldsymbol{a}$ to $\boldsymbol{b}$ along a trajectory γ given by $\boldsymbol{r}(t)$ for $t_0 \leq t \leq t_1$, then the work done by the force $\boldsymbol{F}(\boldsymbol{r})$ is defined as the line integral

$$W(\gamma) = \int_\gamma \boldsymbol{F} \cdot d\boldsymbol{r} = \int_{t_0}^{t_1} \boldsymbol{F} \cdot \frac{d\boldsymbol{r}}{dt}\, dt. \tag{54}$$

In general, this work depends on the trajectory and not just on the end-points $\boldsymbol{r}(t_0) = \boldsymbol{a}$ and $\boldsymbol{r}(t_1) = \boldsymbol{b}$.



• However, there is a special class of forces where this work depends only on the endpoints. In fact, suppose the force field is given by the negative gradient of a potential function $\boldsymbol{F}(\boldsymbol{r}) = -\boldsymbol{\nabla} V(\boldsymbol{r})$. Then the work done by such a 'conservative' force is

$$W(\gamma) = -\int_{t_0}^{t_1} \boldsymbol{\nabla} V \cdot \frac{d\boldsymbol{r}}{dt} dt = -\int_{t_0}^{t_1} \frac{dV(\boldsymbol{r}(t))}{dt}\, dt = V(\boldsymbol{a}) - V(\boldsymbol{b}), \tag{55}$$

which is seen to be the drop in the potential. Here we used the chain rule to write

$$\boldsymbol{\nabla} V \cdot \dot{\boldsymbol{r}} = \left(\frac{\partial V}{\partial x}, \frac{\partial V}{\partial y}, \frac{\partial V}{\partial z}\right) \cdot \left(\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}\right) = \frac{\partial V}{\partial x}\frac{dx}{dt} + \frac{\partial V}{\partial y}\frac{dy}{dt} + \frac{\partial V}{\partial z}\frac{dz}{dt} = \frac{dV}{dt}. \tag{56}$$

3.2 Conserved energy for a conservative force

• Many interesting forces such as the gravitational force, the electrostatic force and the simple harmonic restoring force (but not friction, see (60)) are conservative $\boldsymbol{F} = -\boldsymbol{\nabla} V$. For such conservative forces, Newton's second law becomes $m\ddot{\boldsymbol{r}} = -\boldsymbol{\nabla} V(\boldsymbol{r})$. This implies that the sum of kinetic and potential energies, $E = \frac{1}{2}m\dot{\boldsymbol{r}}^2 + V(\boldsymbol{r})$ is conserved along trajectories:

$$\dot{E} = m\dot{\boldsymbol{r}} \cdot \ddot{\boldsymbol{r}} + \boldsymbol{\nabla} V(\boldsymbol{r}) \cdot \dot{\boldsymbol{r}} = \dot{\boldsymbol{r}}(m\ddot{\boldsymbol{r}} + \boldsymbol{\nabla} V) = 0 \quad \text{since} \quad m\ddot{\boldsymbol{r}} = -\boldsymbol{\nabla} V. \tag{57}$$

We have used Newton's equation of motion, which means that we have shown that E is constant along a trajectory assuming the force is conservative. A curve qualifies as a trajectory if it satisfies the equation of motion $m\ddot{\boldsymbol{r}} = \boldsymbol{F}$.

• We say that the energy is a conserved quantity or a constant of motion if the forces are conservative (expressible as the gradient of a potential).

• This total energy may also be expressed in terms of momentum $\boldsymbol{p} = m\dot{\boldsymbol{r}}$ rather than velocity: $E = \boldsymbol{p}^2/2m + V(\boldsymbol{r})$. In the latter form, the energy is also called the Hamiltonian.

• One can obtain this conserved energy by integrating Newton's equation of motion once using an integrating factor.

• Consider motion on a line (one degree of freedom). For a conservative force $f = -V'(x)$, Newton's equation $m\ddot{x} = f$ becomes $m\ddot{x} + V'(x) = 0$.

• Multiplying by the 'integrating factor' $\dot{x}$, we get

$$m\ddot{x}\dot{x} + V'(x)\dot{x} = 0 \quad \text{or} \quad \frac{d}{dt}\left(\frac{1}{2}m\dot{x}^2 + V(x(t))\right) = 0. \tag{58}$$

Thus, the total energy $E = \frac{1}{2}m\dot{x}^2 + V(x)$ is conserved. This energy is the sum of a kinetic energy $\frac{1}{2}m\dot{x}^2$ (which accrues from the particle's motion) and the previously introduced potential energy $V(x)$.

• **Work-kinetic energy relation** Interestingly, even if the force is not conservative and there is no potential V, the work done by the force while moving a particle along a trajectory can be expressed as the increase in kinetic energy $\frac{1}{2}m\boldsymbol{v}^2$. Indeed, suppose $\boldsymbol{r}(t)$ for $t_a \leq t \leq t_b$ is a trajectory from $\boldsymbol{a}$ to $\boldsymbol{b}$, then using Newton's second law,

$$\begin{aligned} W(\gamma) &= \int_{t_a}^{t_b} \boldsymbol{F} \cdot \frac{d\boldsymbol{r}}{dt}\, dt = \int \frac{d\boldsymbol{p}}{dt} \cdot \boldsymbol{v}\, dt = \int \frac{d(m\boldsymbol{v})}{dt} \cdot \boldsymbol{v}\, dt \\ &= \int_{v_a}^{v_b} m\boldsymbol{v} \cdot d\boldsymbol{v} = \int \frac{1}{2} m d(\boldsymbol{v}^2) = \frac{1}{2}m\boldsymbol{v}_b^2 - \frac{1}{2}m\boldsymbol{v}_a^2. \end{aligned} \tag{59}$$

• This result is called the *work-energy theorem or principle.*

• Let us contrast this with formula (55) for the work done by a conservative force (which depends only on the difference in potential energies at the endpoints of a trajectory). Despite appearances, the RHS of (59), (where we make no assumption about the force being conservative) generally depends on the trajectory and not just the endpoints $\boldsymbol{a}$ and $\boldsymbol{b}$.

• In fact, the kinetic energies at $\boldsymbol{a}$ and $\boldsymbol{b}$ depend on the velocities (or momenta) at $\boldsymbol{a}$ and $\boldsymbol{b}$. As we noted in our discussion of Newton's second law, the specification of initial position and momentum is enough to determine an entire trajectory.

3.3 Example of a nonconservative force: damping force

• Of course, not every system is conservative. In fact, most real-world systems are *not* conservative due to interaction with the environment: dissipation, external driving etc.

• For instance, suppose a particle moves under the influence of both a conservative force as well as a *frictional* or *damping* force proportional to its velocity

$$m\ddot{x} = -V'(x) - \gamma\dot{x} \quad \text{with damping coefficient} \quad \gamma > 0. \tag{60}$$

• The frictional force $-\gamma\dot{x}$, being dependent on $\dot{x}$, cannot be written as the derivative (with respect to x) of some function of x. Thus, this frictional force is not conservative.

• In this case, we may show that the above-defined energy is nonincreasing. Multiplying (60) by $\dot{x}$, we get

$$\frac{d}{dt}\left(\frac{1}{2}m\dot{x}^2 + V(x)\right) = -\gamma\dot{x}^2 \leq 0. \tag{61}$$

3.4 Angular momentum

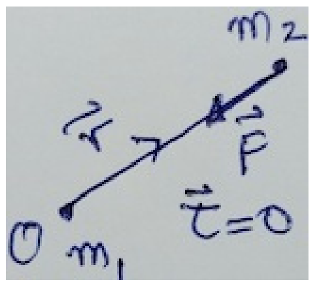
• The angular momentum (about a chosen origin) of a particle moving in 3d space is $\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p}$, where $\boldsymbol{r}$ is the position vector of the particle from the chosen origin. In components

$$L_x = yp_z - zp_y, \quad L_y = zp_x - xp_z, \quad L_z = xp_y - yp_x. \tag{62}$$

Angular momentum is also called the moment of momentum. Formulae for successive components are obtained by cyclically permuting $x \to y \to z \to x$.

• Newton's force law then implies that the rate of change of angular momentum is the torque (or moment of force) about the same origin:

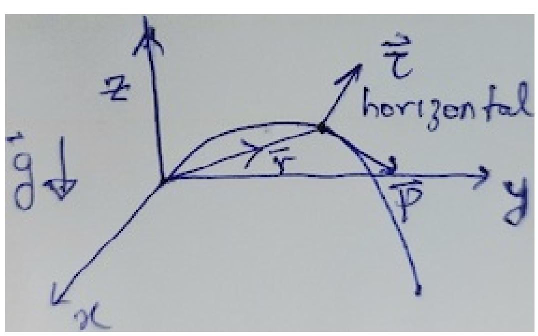
$$\dot{\boldsymbol{L}} = \dot{\boldsymbol{r}} \times \boldsymbol{p} + \boldsymbol{r} \times \dot{\boldsymbol{p}} = \frac{1}{m}\boldsymbol{p} \times \boldsymbol{p} + \boldsymbol{r} \times \boldsymbol{F} = \boldsymbol{r} \times \boldsymbol{F} \equiv \boldsymbol{\tau}. \qquad (63)$$



• A particularly important example is a *central force*, i.e., one which points radially along the line from the origin to the particle. Newton's gravitational force between point masses, as well as Coulomb's electrostatic force between point charges, are central forces.

• The torque due to a central force about the force center vanishes since $\boldsymbol{r}$ and $\boldsymbol{F}$ are collinear. Thus, we conclude that the angular momentum of a particle moving in a central force field is independent of time, it is a conserved quantity.

• For a projectile moving under the vertical gravitational force, the torque must be in the horizontal plane.



• So the vertical component of angular momentum $L_z = xp_y - yp_x$ must be conserved. Since p_x and p_y are also conserved, we conclude that the trajectory $(x, y, z)(t)$ must be such that its projection on the horizontal plane is a straight line $L_z = xp_y - yp_x$.

• In fact, one can show that the trajectory of a projectile is a parabola over the x-y plane. Knowledge of conserved quantities allowed us to clarify the nature of the trajectory.

3.5 Dynamical variables, phase & configuration spaces and conserved quantities

• **Dynamical variables.** The components of position $\boldsymbol{r}$, momentum $\boldsymbol{p}$, angular momentum $\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p}$ and energy $E = \frac{\boldsymbol{p}^2}{2m} + V(\boldsymbol{r})$ are interesting physical quantities associated with the dynamics of a particle. They are examples of *dynamical variables* or observables (a term that is used more in the quantum mechanical context). They can change with time.

• In general, any real function $f(\boldsymbol{r}, \boldsymbol{p})$ of the components of the position and momentum of a particle is a dynamical variable. The potential $V(\boldsymbol{r})$ is a dynamical variable. The components of position x, y, z and those of momentum p_x, p_y, p_z are the basic dynamical variables. In general, dynamical variables change along the trajectory.

• Note that the mass of a particle, a spring constant or the charge of a particle are not dynamical variables. They are called parameters and are used to specify the nature of the particle or system. They are not functions of positions and momenta.

• Recall that the set of possible instantaneous positions of the particles in a system is called its **configuration space** (denoted Q). The number of degrees of freedom is the dimension of the configuration space.

• For one particle moving along a line, its position is specified by the position coordinate x, which can take any real value. Thus the configuration space of such a particle is $\mathbb{R}^1$. A point particle moving in 3d space has $Q = \mathbb{R}^3$ with coordinates x, y, z. Two point particles moving in 3d has $Q = \mathbb{R}^6$ with coordinates given by the positions coordinates of both the particles $x_1, y_1, z_1, x_2, y_2, z_2$. What is the configuration space of a rigid rotor - a rigid stick of zero thickness and length l?

• Newton's equation of motion $m\ddot{\boldsymbol{r}} = \boldsymbol{f}$ is 2nd order in time. Knowledge of the initial configuration is not sufficient to determine the trajectory of a mechanical system. For instance, for a particle, we need to know both the initial position $\boldsymbol{r}(0)$ and momentum $\boldsymbol{p}(0)$ to determine the trajectory by solving Newton's equation of motion. The pair $(\boldsymbol{r}(t), \boldsymbol{p}(t))$ is called the **state** of the particle at time t.

• The set of possible instantaneous states of the particle is called its **state space or phase space** M. For a particle moving on the real line, the phase space is $\mathbb{R}^2$ parametrized by the pair of coordinates (x, p). For a particle moving in 3D space, its configuration space is $\mathbb{R}^3$ and its phase space is $\mathbb{R}^6$ (locations and momenta).

• With the concept of the phase space at hand, a **dynamical variable** may be defined as a (sufficiently smooth) real-valued function on phase space. For one particle moving on a line, a dynamical variable is a function $f(x, p)$. f gets its time-dependence from that of x and p. Sometimes, one includes explicitly time-dependent functions $f(x, p, t)$ as dynamical variables.

• The path of the particle $\boldsymbol{r}(t)$ (satisfying Newton's equation and initial conditions) is called its (configuration space) **trajectory**. Trajectories are oriented by arrows specifying forward time evolution.

• Also of interest is the **trajectory in phase space**, the curve $(\vec{x}(t), \vec{p}(t))$ in phase space. A **phase portrait** is a sketch of trajectories on phase space. Draw the phase portrait of a free particle that can move on a straight line.

• **Conserved quantities.** Conserved quantities are dynamical variables that are constant along *every* trajectory. This means the value of a conserved quantity does not change as the system evolves. The value of a conserved quantity may differ from trajectory to trajectory. For example, momentum is a conserved quantity for free particle motion. But the value of momentum in general differs from trajectory to trajectory, depending on how fast the particle is moving. In general, the value of a conserved quantity is determined by initial conditions.

• Conserved quantities are useful. They help us solve/understand Newton's equation for the trajectory. E.g., for a particle moving on a line subject to a conservative force, Newton's 2nd order equation $m\ddot{x} = -V'(x)$ can be reduced to a first order equation stating the conservation of energy.

We may then integrate once more and get an (implicit) expression for $x(t)$:

$$E = \frac{1}{2}m\dot{x}^2 + V(x) \quad \Rightarrow \quad \frac{dx}{dt} = \pm\sqrt{\frac{2}{m}\left(E - V(x)\right)}$$
$$\Rightarrow \qquad t - t_0 = \pm\int_{x_0}^{x} \frac{dx'}{\sqrt{\frac{2}{m}\left(E - V(x')\right)}}. \tag{64}$$

• Conservation of energy has allowed us to reduce the order of Newton's original differential equation by one.

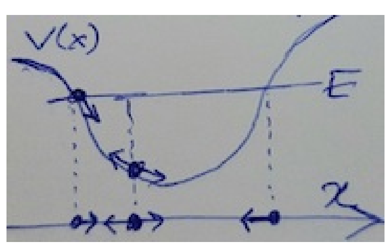
• It is noteworthy that both signs correspond to forward time evolution ($t \geq t_0$). The positive sign corresponds to a situation where $x > x_0$ (rightward motion) while the negative sign corresponds to $x < x_0$ (leftward motion). More on this below.

• In effect, we have solved Newton's second order equation of motion in two steps. Energy is the constant of integration in the first step and x_0 is the second constant of integration. We can think of x_0 and E as specifying (partial, see below) initial conditions at time t_0. Though t_0 is a constant of integration we do not regard it as as an initial condition, but rather designate it as the initial time.

• Our answer expresses t as a function of x. We must invert it to find trajectories $x(t)$ with energy E and initial location x_0 at t_0.

• Interestingly, there is often more than one trajectory with fixed energy and initial location, corresponding to the $\pm$ signs.

• This is to be expected, since specification of energy allows two possible initial velocities in general $v_0 = v(t_0) = \pm\sqrt{(2/m)(E - V(x_0))}$.



• There are exceptions. If the particle is at a turning point of the potential $E = V(x_0)$, initially, then $v_0 = 0$ and the particle has only one way to go, 'down hill'.

• So specification of energy and initial location is, in general, not a complete specification of the instantaneous state of the particle.

• Explain in qualitative terms the motion of a particle in a potential of the sort shown in the figure for given energy E and initial position x_0. Which way does the force point at various locations. Which parts of the x-axis are forbidden from being explored? Where does the particle move fastest? Identify equilibrium and turning points for the motion. Argue that the force may be called restoring. In doing the integration, specify when to use the $+$ and $-$ signs. Argue that the motion is oscillatory and periodic in time.

4 Collisions or scattering and conservation laws

• We will now illustrate the use of conserved quantities (mass, momentum and energy) in the context of collisions.

• By a collision of point particles, we shall mean an interaction among particles that behave as free particles in the asymptotic past and future so that each of them has a constant velocity as $t \to \pm\infty$.

• Such a situation arises if the forces between particles are sufficiently short-ranged and particles are separated by distances large compared to the range of forces as $t \to \pm\infty$.

• A collision does not necessarily mean the particles come into contact.

• For example (a) two particles may collide, each suffering a deflection in direction of motion, (b) a particle may disintegrate/decay into two or more particles, (c) two or more particles may coalesce (merge), etc.

• Evidently, the number of incoming and outgoing particles in a collision need not be equal.

• Though particle number need not be conserved, collisions may be fruitfully treated using the conservation laws of mass, momentum and energy even without a detailed knowledge of the forces of interaction.

• (Inertial) mass, regarded intuitively as the amount of matter, is conserved in nonrelativistic mechanical processes: it is neither created nor destroyed.

• The conservation of momentum and energy should not come as a surprise. If external forces may be neglected, the total momentum of a system is conserved. For a system of particles, this is a consequence of Newton's 2^{nd} and 3^{rd} laws, with the latter allowing cancellation of interparticle forces in computing the rate of change of total momentum.

• The total energy of an isolated system is also conserved. However, kinetic energy could arise from or be transformed into other types of energy (potential energy in a spring, chemical bond energy etc.) and we must account for this.

• Suppose we have a collision among p ('past') incoming particles resulting in f ('future') outgoing particles. Let the masses and velocities (as $t \to \mp\infty$) of the particles be denoted $(m_i, \boldsymbol{v}_i)$ for $i = 1, \ldots, p, p+1, \ldots, p+f$.

• Then the law of conservation of mass states that

$$m_1 + \cdots + m_p = m_{p+1} + \cdots + m_{p+f}. \qquad (65)$$

• The conservation of linear momentum in a collision is the statement that

$$m_1 \boldsymbol{v}_1 + \cdots + m_p \boldsymbol{v}_p = m_{p+1} \boldsymbol{v}_{p+1} + \cdots + m_{p+f} \boldsymbol{v}_{p+f}. \qquad (66)$$

• The conservation of energy [initial energy = final energy (e.g., kinetic + potential)] can be rewritten as

$$\sum_{i=1}^{p} \frac{1}{2} m_i \boldsymbol{v}_i^2 = \sum_{i=p+1}^{p+f} \frac{1}{2} m_i \boldsymbol{v}_i^2 + Q. \qquad (67)$$

• The difference Q between initial and final kinetic energies is positive if, say, kinetic energy is stored in a compressed spring or released as heat.

• It is negative if internal potential energy is converted into kinetic energy, for instance in the decay of a particle that was initially at rest.

• If $Q = 0$, the collision is called elastic: kinetic energy is conserved but may be redistributed among the particles.

• An example of a collision is the elastic '$2 \to 2$' scattering of two particles which retain their identities (including their masses).

• In this case, conservation of mass is automatic and we have effectively 4 conservation laws (for energy and the three components of momentum).

• Given the initial velocities $\boldsymbol{v}_1, \boldsymbol{v}_2$, these 4 equations are insufficient to determine the final velocities $\boldsymbol{v}_3, \boldsymbol{v}_4$, which comprise 6 unknowns. The conservation equations are underdetermined. Although conservation laws place restrictions on the final velocities, one needs information on the nature of forces to determine the latter (by solving Newton's equation of motion).

• However, in the special case of collisions in 1d, we have 2 conservation laws and 2 unknown final velocity components and the system of equations is even-determined.

• If we denote the initial and final velocities by v_1, v_2 and v_1', v_2', then the conservation laws for elastic $2 \to 2$ scattering become

$$\begin{aligned} m_1 v_1 + m_2 v_2 &= m_1 v_1' + m_2 v_2' \equiv p \quad \text{and} \\ \frac{1}{2} m_1 v_1^2 + \frac{1}{2} m_2 v_2^2 &= \frac{1}{2} m_1 v_1'^2 + \frac{1}{2} m_2 v_2'^2 \equiv T. \end{aligned} \tag{68}$$

• Eliminating $v_1' = (p - m_2 v_2')/m_1$ and writing $M = m_1 + m_2$, we get

$$\begin{aligned} v_2' &= M^{-1} \left[p \pm \sqrt{p^2 - M(m_1/m_2)(p^2/m_1 - 2T)} \right] \\ &= (2 m_1 v_1 + (m_2 - m_1) v_2)/M \quad \text{or} \quad v_2. \end{aligned} \tag{69}$$

• The second solution is the trivial one, where the particles retain their velocities $v_1' = v_1, v_2' = v_2$, this happens if the particles do not interact at all. In the first, the scattering is nontrivial.

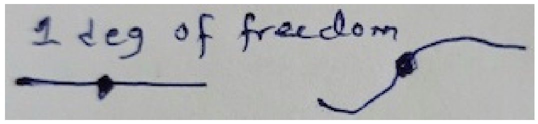
5 Motion in one dimension

• Consider a particle of mass $m > 0$ moving on the real line $\mathbb{R}$ with instantaneous position $x(t)$. We do this for simplicity, though some of the features we discuss are valid more generally.

• It is said to possess one degree of freedom since precisely one coordinate (x) is needed to specify its location.

• A particle moving on a circle or other curve also has one degree of freedom.

• It is a free particle if it is isolated from physical influences (no 'forces'



act on it), in which case Newton's 1$^{\text{st}}$ law states that it must either be at rest or moving at a constant velocity $\dot{x} = \frac{dx}{dt}$ to the right or left.

• On the other hand, if a force F acts on it, Newton's 2$^{\text{nd}}$ law says that the particle accelerates according to the *equation of motion* mass $\times$ acceleration $=$ force or $m\ddot{x} = F$.

• To find the trajectory $x(t)$ of the particle, we need to solve this second order ordinary differential equation subject to a pair of initial conditions, which could be the initial location and velocity $(x(0), \dot{x}(0))$.

• If the force depends only on location $F = F(x)$, then in one dimension, we may define a potential function such that $F(x) = -V'(x)$. The latter is a negative antiderivative or primitive of F:

$$V(x) = V(0) - \int_0^x F(x')\, dx'. \tag{70}$$

• For a conservative force, Newton's equation becomes $m\ddot{x} + V'(x) = 0$. Multiplying by the 'integrating factor' $\dot{x}$ we get

$$m\ddot{x}\dot{x} + V'(x)\dot{x} = 0 \quad \text{or} \quad \frac{d}{dt}\left(\frac{1}{2}m\dot{x}^2 + V(x(t))\right) = 0. \tag{71}$$

Thus, the total energy $E = \frac{1}{2}m\dot{x}^2 + V(x)$ is conserved. This energy is the sum of a kinetic energy $\frac{1}{2}m\dot{x}^2$ (which accrues from the particle's motion) and the previously introduced potential energy $V(x)$.

5.1 Turning points, bound and unbound motion

• Having a conserved energy is helpful in understanding the dynamics.

• Indeed, for a given energy E, the nature of the motion can be deduced from a graph of the potential $V(x)$.

• On account of the positivity of kinetic energy $\frac{1}{2}m\dot{x}^2$, the motion is confined to the region where $E \geq V(x)$.

• This region, when nonempty, may be a union of several intervals/points (see Fig 2). However, due to 'potential barriers' a particle cannot jump between two disconnected intervals, so we may discuss each in isolation

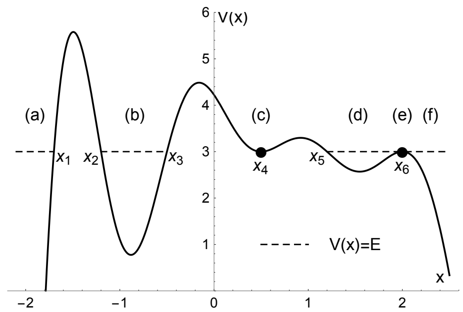


Figure 1: Qualitative characterization of motion of a particle in a 1d potential.

• In Fig. 2, the closed interval $[x_2, x_3]$ (the square brackets mean the endpoints are included) is a connected set, any two points in it can be joined by a curve lying in it. On the other hand, the disjoint union $[x_2, x_3] \cup \{x_4\}$ is disconnected.

• The points x where the energy-E horizontal line intersects the graph of $V(x)$ are the places where $\dot{x}$ vanishes momentarily and the energy is purely potential. They are called 'turning points' since the particle turns around at such a point if reached in finite time [this happens if $V' \neq 0$ at a turning point].

• For the potential in Fig. 2, the classically allowed region corresponding to the indicated energy is a union of six connected sets, with initial conditions determining in which one the motion takes place. From left to right, try to argue that the qualitative motion is of the following sorts:

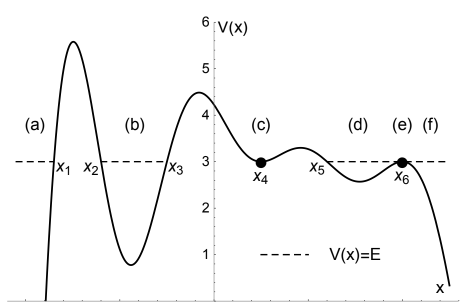


Figure 2: Qualitative characterization of motion of a particle in a 1d potential.

(a) $(-\infty, x_1]$: Particle can come in from any point to the left of x_1, collide against the barrier at x_1, turn around and escape/scatter to $-\infty$. In this process, the particle reaches x_1 in a finite time. This is indicated via the square bracket, which means the interval is closed at the x_1 end.

(b) $[x_2, x_3]$: Particle oscillates with finite time period between the turning points at x_2 and x_3.

(c) x_4: Particle remains at rest at the stable equilibrium point x_4.

(d) $[x_5, x_6)$: For instance, particle starting at x_5 accelerates and moves rightward but then slows down and takes infinitely long to reach x_6. Consequently, x_6 is not part of the interval, and this is indicated via the round bracket. By contrast, the turning point at x_5 is reached in finite time, at which the particle comes instantaneously to rest and reverses direction. $[x_5, x_6)$ is a 'closed-open' interval.

(e) x_6: Particle remains at rest at the unstable equilibrium point x_6.

(f) (x_6, ∞): Particle can come leftwards from large x, but slows down and takes infinitely long to reach x_6. Starting from $x > x_6$ with rightward velocity, particle speeds up and escapes to infinity.

(g) (x_1, x_2), (x_3, x_4) and (x_4, x_5) are forbidden intervals. The particle cannot be found in any of these intervals since its kinetic energy would have to be negative.

• While in (b)-(d) the particle is 'bound' or 'confined', in (a) and (f) the motion can be unbounded. The foregoing statements about the finite or infinite time taken to reach turning points can be established by solving Newton's equation.

5.2 Time-reversal

• Newton's equation for motion in a potential $m\ddot{x} = -V'(x)$ is time-reversal invariant in the sense that if $x(t)$ is a solution, then so is $x(-t)$. Here, for simplicity, we assume that the solution $x(t)$ exists for all time t. If it exists for $t_{\rm min} \le t \le t_{\rm max}$, then the assertion is that $x(t_{\rm max} - t)$ is also a solution for $0 \le t \le t_{\rm max} - t_{\rm min}$.

• To see this, suppose we consider the first case where the solution $x(t)$ exists for all time $-\infty < t < \infty$. Let $\tilde{t} = -t$. Then $\frac{d}{d\tilde{t}} = -\frac{d}{dt}$ and $\frac{d^2}{d\tilde{t}^2} = \frac{d^2}{dt^2}$. So Newton's equation $m\frac{d^2}{dt^2}x(t) = -V'(x(t))$ implies that

$$m\frac{d^2}{d\tilde{t}^2}x(-\tilde{t}) = -V'(x(-\tilde{t})). \tag{72}$$

We see that $x(-\tilde{t})$ satisfies the same equation as $x(\tilde{t})$ satisfied [$\tilde{t}$ is just a dummy variable, it could be renamed t]. This equation implies that $x(-t)$ is a trajectory if $x(t)$ was one.

• In other words, a movie of a solution played backwards is also an admissible motion for any conservative force.

• We often indicate the effect of time-reversal succinctly by writing that under $t \to -t$, $x(t) \to x(-t)$, $\dot{x}(t) \to -\dot{x}(-t)$, $\ddot{x}(t) \to \ddot{x}(-t), F(x(t)) \to F(x(-t)), V(x(t)) \to V(x(-t))$ and $\dot{x}^2(t) \to \dot{x}^2(-t)$ etc.

• A force that is linear in velocities (and therefore not conservative) leads to a Newton equation that is *not* time-reversal invariant. For e.g., $m\ddot{x} = -\gamma\dot{x}$, under $t \to -t$ becomes $m\ddot{x}(-t) = \gamma\dot{x}(-t)$. In this case, $x(-t)$ does not satisfy the same equation as $x(t)$.

6 Oscillations

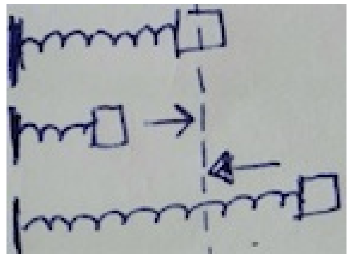
6.1 Simple harmonic motion

• The linear or simple harmonic oscillator is one of the simplest of mechanical systems with one degree of freedom.

• It describes, for instance, small oscillations of a particle of mass m due to a linear restoring force (say, due to a spring) $F = -kx$ proportional to the particle's displacement x from equilibrium. Here, the positive constant k (what are its dimensions?) is called the force constant.

• The negative sign indicates that the force tends to restore the particle to its equilibrium position rather than push it further away. When the spring is compressed, it tends to expand; when it is stretched, it tends to contract.

• As we shall see, a linear restoring force results in oscillatory motion of the



particle around the point of equilibrium. The specific type of oscillatory motion that results is called harmonic motion since it involves the sine and cosine functions of time.

• The force law $F = -kx$ is called Hooke's law after Robert Hooke who used it to model such motion. More complicated 'anharmonic' forces are possible, such as a cubic restoring force $F = -gx^3$ for a constant $g > 0$.

• Newton's second law $F = ma$ for a particle subject to a linear restoring force leads to the differential equation

$$m\ddot{x} = -kx. \tag{73}$$

• This equation could describe oscillations of the extension x of a spring of force constant k. One end of the spring is held fixed while the particle of mass m is attached to the other end.

• Suppose ℓ_0 is the natural or equilibrium length of the spring while $\ell(t)$ is its length as it expands and contracts. Then $x = \ell(t) - \ell_0$. Notice that the neither ℓ_0 nor $\ell(t)$ enter the equation of motion, which involves only the departure x from equilibrium length.

• It is conventional to define the parameter $\omega = \sqrt{k/m}$ (with dimensions of inverse time or a frequency) so that the EOM becomes $\ddot{x} = -\omega^2 x$. This equation is linear in x. This means any linear combination $ax_1(t) + bx_2(t)$ of solutions $x_{1,2}(t)$ is again a solution. But how do we solve it?

• If the force were absent ($k = 0$ and $\omega = 0$), then we have $\ddot{x} = 0$ and we may integrate this once $\dot{x} = a$ and then again $x = at + b$ to find the general solution.

• Notably, the general solution depends on 2 constants of integration (a and b): this is generally true of second order ODEs. The integration constants are to be fixed using initial conditions (such as $x(0)$ and $\dot{x}(0)$).

• The ODE $\ddot{x} = -\omega^2 x$ cannot simply be integrated once since we do not a priori know the integral of x.

• We can, however, make a guess for the sort of function that may be a solution. Thus, we try $x = e^{rt}$ for some constant r. Putting this in, we get $r^2 e^{rt} = -\omega^2 e^{rt}$ leading to $r^2 = -\omega^2$. There are two possibilities $r = \pm i\omega$. This leads to two linearly independent solutions $x_1 = e^{i\omega t}$ and $x_2 = e^{-i\omega t}$. Taking arbitrary linear combinations of these, we get a 2 parameter family of solutions $x(t) = \tilde{a} e^{i\omega t} + \tilde{b} e^{-i\omega t}$ ($\tilde{a}$ is pronounced 'a-tilde').

• However, the solution must be real since x is a displacement from equilibrium. This imposes conditions on $\tilde{a}$ and $\tilde{b}$: $\tilde{a}^* = \tilde{b}$ and $\tilde{b}^* = \tilde{a}$, i.e., they are complex conjugates. So if we express the complex number $\tilde{a}$ in terms of its real and imaginary parts $\tilde{a} = \tilde{A} + i\tilde{B}$, then we must have $\tilde{b} = \tilde{A} - i\tilde{B}$. Consequently,

$$x = \tilde{A}(e^{i\omega t} + e^{-i\omega t}) + i\tilde{B}(e^{i\omega t} - e^{-i\omega t}) = 2\tilde{A}\cos\omega t - 2\tilde{B}\sin\omega t \qquad (74)$$

Let us now denote $c_1 = 2\tilde{A}$ and $c_2 = -2\tilde{B}$. Then we may express the solution as

$$x(t) = c_1 \cos\omega t + c_2 \sin\omega t \qquad (75)$$

where c_1 and c_2 are two real constants of integration.

• This may also be expressed as

$$x(t) = a\cos(\omega t + \alpha) \quad \text{with} \quad a \geq 0 \quad \text{and} \quad 0 \leq \alpha < 2\pi. \qquad (76)$$

Using $\cos(\omega t + \alpha) = \cos\omega t \cos\alpha - \sin\omega t \sin\alpha$, we find that

$$c_1 = a\cos\alpha \quad \text{and} \quad c_2 = -a\sin\alpha \qquad (77)$$

while

$$a^2 = c_1^2 + c_2^2 \quad \text{and} \quad \cos\alpha = c_1/a. \qquad (78)$$

• We say that the particle displays simple harmonic oscillations with amplitude $a = \sqrt{c_1^2 + c_2^2}$ and initial phase $\alpha = \arccos(c_1/a)$, which are determined by initial conditions. Note that $\alpha = -\arctan(c_2/c_1)$, though this formula does not uniquely determine $\alpha \in [0, 2\pi)$ since the tangent function has period π unlike the cosine and sine functions that have periods of 2π.

• Plot the solution for various values of a and α.

• The material constant ω is called the angular frequency and determines the time period $T = 2\pi/\omega$ of the small oscillations, which is independent of amplitude. We say that simple harmonic oscillations are isochronous.

• This isochronous nature is special to a linear restoring force. If the force were nonlinear (e.g. cubic rather than linear), the motion would still be periodic, but the time period would depend on amplitude.

• It follows that a stiff spring (larger force constant k) has a smaller time period T. On the other hand, a particle with larger inertia m would have a longer time period.

• Newton's second oder equation $m\ddot{x} = -kx$ may also be written as a pair of first order equations by introducing the momentum $p = m\dot{x}$. Indeed, they are given by

$$\dot{x} = p/m \quad \text{and} \quad \dot{p} = -kx. \tag{79}$$

The dynamical variables x and p are called the dependent variables and t the independent variable. This is a system of two homogeneous linear ordinary differential equations. Explain the qualifying terms.

• This pair of linear equations can be written in matrix form

$$\frac{d}{dt}\begin{pmatrix} x \\ p \end{pmatrix} = \begin{pmatrix} 0 & 1/m \\ -k & 0 \end{pmatrix}\begin{pmatrix} x \\ p \end{pmatrix}. \tag{80}$$

• This system is of the form $\dot{\psi} = A\psi$ where

$$\psi = \begin{pmatrix} x \\ p \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} 0 & 1/m \\ -k & 0 \end{pmatrix}. \tag{81}$$

This form is nice since it makes it clear why the exponential guess was a good one.

• Comparing with $\dot{y} = ay$ whose solution is $y(t) = e^{at}y(0)$, it is natural to propose that the solution must be given by $\psi = e^{At}\psi(0)$ where e^{At} is the matrix exponential, defined via the exponential series $\sum_{n=0}^{\infty}(At)^n/n!$. One can check by substituting the exponential series and differentiating term by term that $\frac{de^{At}}{dt} = Ae^{At}$.

• Upon calculating the matrix exponential we get

$$\exp At = \begin{pmatrix} \cos\omega t & (1/m\omega)\sin\omega t \\ -m\omega\sin\omega t & \cos\omega t \end{pmatrix} \tag{82}$$

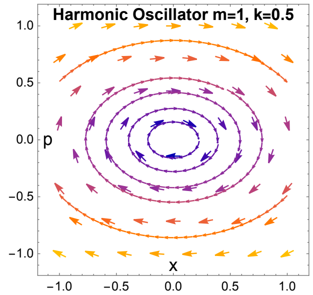


Figure 3: Phase portrait of the harmonic oscillator showing concentric ellipse phase trajectories.

Thus the solution of the 'initial value problem' (IVP) for $x(t)$ and $p(t)$ given $x(0)$ and $p(0)$ is

$$\begin{aligned} x(t) &= x(0)\cos\omega t + \frac{p(0)}{m\omega}\sin\omega t, \\ p(t) &= -m\omega x(0)\sin\omega t + p(0)\cos\omega t. \end{aligned} \tag{83}$$

Verify that this solves the equations $\dot{x} = p/m$ and $\dot{p} = -kx = -m\omega^2 x$ and also satisfies the initial conditions.

• The linear restoring force $F = -kx$ arises from the potential $V = \frac{1}{2}kx^2$ in the sense that $F = -V'(x)$. The conserved energy of the linear oscillator is then

$$E = \frac{1}{2}m\dot{x}^2 + V(x) = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}m\omega^2 x^2. \tag{84}$$

Introducing the momentum $p = m\dot{x}$, we may also write the energy as $E = \frac{1}{2m}p^2 + \frac{1}{2}m\omega^2 x^2$. In this form, the energy is called the Hamiltonian of the oscillator.

• The phase space of the oscillator is $\mathbb{R}^2$ since the initial values of x and p can each be any real number.

• Sketch a phase portrait for the simple harmonic oscillator. Show that phase trajectories are clockwise directed ellipses centered at the origin. What are the semi-axes of these ellipses? The origin $x = 0, p = 0$ is the only static solution of the equations of motion and is a one-point trajectory.

6.2 Simple pendulum

6.2.1 Qualitative description and equation of motion

• Consider a bob of mass m suspended from a massless rigid rod of length ℓ clamped at a pivot, as shown in Fig. 4. The bob is free to move subject to Earth's constant downward gravitational force.

• Suppose the rod is initially deflected from the downward position. The direction of the rod and the downward pointing acceleration due to gravity together, define a plane. We will suppose that the initial velocity of the bob lies in this plane. If this is the case, the motion of the pendulum will be confined to this plane.

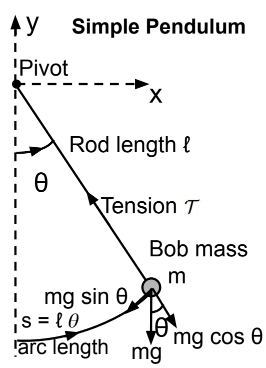


Figure 4: Simple pendulum suspended from a fixed support with massless rod and oscillating in a vertical plane.

• This system is an idealized simple pendulum, it is used in clocks. A heavy pendulum bob (wrecking ball) may be used to demolish buildings! A simple pendulum is also a conceptually interesting system which we will use to illustrate many concepts of mechanics.

• The qualifier 'simple' means the mass is concentrated in the bob, which is assumed point-like. A compound pendulum is one where the mass is not point-like but distributed, say over the rod.

• A spherical pendulum is one where the rod is not confined to a vertical plane. In this case the bob can move on a sphere of radius ℓ, which explains the name.

• From our experience, we know that the pendulum is in stable equilibrium when hanging vertically downward, the downward gravitational force being balanced by the upward 'tension' force in the rod. A small push makes the bob oscillate through small angles, always remaining close to equilibrium.

• The pendulum is also in equilibrium when it is balanced vertically upwards. But this is unstable equilibrium, a small push in either direction will take the bob far from the point of equilibrium.

• The above are the two time-independent motions of a simple pendulum.

• Let us now consider time-dependent motion. We are aware of two types of motion of a pendulum.

• (1) **Libration** is oscillation between a pair of turning points on either side of the vertical. In a sense, the bob is *bound* or *trapped* around its point of stable equilibrium.

• (2) **Rotation** ensues if the energy or initial speed of the bob is above a critical value. The bob then rotates around the pivot, in general at a nonuniform rate. It moves slower at the top and faster at the bottom. In rotational motion, the bob is not trapped near its point of stable equilibrium and the motion does not have any turning points.

• The pendulum has one degree of freedom, the (counterclockwise) angle of deflection θ from its stable equilibrium position. θ can be chosen to take values in the interval $0 \le \theta < 2\pi$. $\theta = 2\pi$ corresponds to the same angular configuration as $\theta = 0$ (bob hanging vertically downwards).

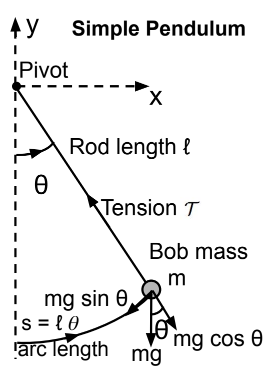


Figure 5: Simple pendulum suspended from a fixed support with massless rod and free to oscillate in a vertical plane. $\hat{r}$ is radially outward, $\hat{\theta}$ points azimuthally counterclockwise. $\hat{z}$ is out of the plane.

• The downward gravitational force $-mg\hat{y}$ is resolved into a radial component $mg\cos\theta$ and a tangential component $mg\sin\theta$ as shown in Fig. 5.

• Newton's second law in the radial direction is $\mathcal{T} - mg\cos\theta = m\ell\dot{\theta}^2$. The quantity on the right is the mass $\times$ the centripetal acceleration while $\mathcal{T}$

is the radially inward tension in the rod. $\mathcal{T}$ must vary with angle and is greatest when $\theta = 0$ since the angular speed $\dot{\theta}$ is largest at the bottom and $mg\cos\theta$ is also maximal when $\theta = 0$.

• The tangential component of the gravitational force tends to reduce the angle of deflection and causes the bob to accelerate towards its equilibrium position.

• Suppose $s = \ell\theta$ is the arc length corresponding to a counterclockwise deflection angle θ. Then the tangential velocity of the bob is $\dot{s} = \ell\dot{\theta}$ and its acceleration is $\ell\ddot{\theta}$ since ℓ is constant.

• Newton's second law then says that $m\ell\ddot{\theta}(t) = -mg\sin\theta(t)$. Thus, we arrive at the equation of motion

$$\ddot{\theta}(t) = -(g/\ell)\ \sin\theta(t) = -\omega^2\sin\theta(t) \quad \text{where} \quad \omega = \sqrt{g/\ell}. \qquad (85)$$

• ω has dimensions of a frequency. In this equation, t is called the independent variable and θ is called the dependent variable (since it depends on t).

• Since there is no radial motion, the radial Newton equation simply fixes $\mathcal{T}$ once the angular motion is determined. Thus the radial and tangential EOM form a sort of 'triangular' system.

• The mass of the bob cancelled out, so the time-dependence of θ, and the motion of the pendulum is independent of m; it can depend only on the constant (angular) frequency ω.

• In particular, the time period of oscillation must be independent of the mass (no matter how large the oscillation is). This was discovered experimentally by Galileo around 1602.

• The equation of motion $\ddot{\theta} = -\omega^2\sin\theta$ for the dependent variable θ is second order in time but very nonlinear due to the $\sin\theta$ on the RHS ($\sin\theta$ can be thought of as an infinite series in powers of θ).

• Even without solving it explicitly, we may determine many qualitative features of the motion.

6.2.2 Energy, angular momentum

• To find a conserved energy, we use the method of an integrating factor. Multiplying $m\ell\ddot{\theta} = -mg\sin\theta$ by the integrating factor $\dot{s} = \ell\dot{\theta}$ we get

$$\ell\dot{\theta}m\ell\ddot{\theta} = -\ell\dot{\theta}mg\sin\theta \quad \text{or} \quad \frac{d(m\ell^2\dot{\theta}^2/2)}{dt} = \frac{d(mg\ell\cos\theta)}{dt}. \tag{86}$$

Physically, $\dot{s}$ is the tangential component of velocity and multiplying it with the tangential component of force gives the work done by the tangential force per unit time.

• Integrating, we get a constant of integration that we denote $E - E_0$. Thus, we have a conserved quantity

$$E = \frac{1}{2}m\ell^2\dot{\theta}^2 - mg\ell\cos\theta + E_0. \tag{87}$$

• We notice that in the state of stable equilibrium, i.e., when $\theta = 0$ and $\dot{\theta} = 0$, $E = -mg\ell + E_0$.

• It is convenient to choose the constant of integration so that the energy vanishes in the state of stable equilibrium. Thus, we choose $E_0 = mg\ell$.

• With this choice, our conserved total energy is

$$E = \frac{1}{2}m\ell^2\dot{\theta}^2 + mg\ell(1 - \cos\theta). \tag{88}$$

• The first term is the kinetic energy $T = \frac{1}{2}m\ell^2\dot{\theta}^2$.

• The potential energy is $V(\theta) = mg\ell(1 - \cos\theta)$. We have chosen to write the constant of integration in such a way that $V = 0$ when the bob hangs downwards ($\theta = 0$).

• The extrema of V correspond to equilibria where the pendulum is stationary. Now $V'(\theta) = mg\ell\sin\theta = 0$ when $\theta = 0$ or π (modulo 2π), corresponding to the bob pointing vertically downwards or upwards.

• As the figure shows, the former is a local minimum of V (stable equilibrium) while the latter is a maximum (unstable equilibrium).

• **Angular momentum.** Suppose $\boldsymbol{r} = \ell\hat{r}$ is the radius vector of the bob with pivot as the origin. Its velocity is $\boldsymbol{v} = \dot{s}\hat{\theta}$ where $s = \ell\theta$. Thus, its

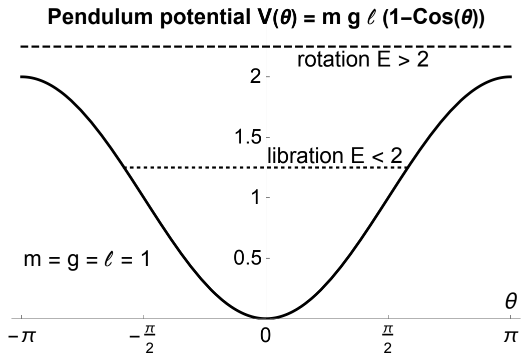


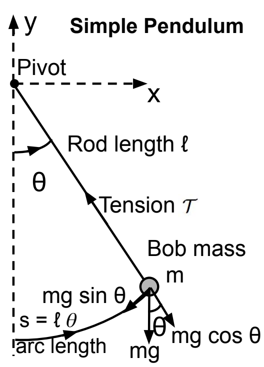
Figure 6: Potential $V(\theta)$ (for $m = g = \ell = 1$) showing stable equilibrium at $\theta = 0$ and unstable one at $\theta = \pm\pi$ as well as examples librational and rotational energies.

linear momentum is

$$\boldsymbol{p} = m\dot{s}\hat{\theta} = m\ell\dot{\theta}\hat{\theta}. \tag{89}$$

Its angular momentum about the pivot is therefore

$$\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p} = m\ell^2\dot{\theta}\hat{r} \times \hat{\theta} = m\ell^2\dot{\theta}\,\hat{z} \equiv p_\theta \hat{z}. \tag{90}$$



(a)

• The angular momentum component $p_\theta = m\ell^2\dot{\theta}$ is positive for counterclockwise motion and negative for clockwise motion. For librational motion, p_θ must keep changing sign along a trajectory. In the rotational phase p_θ must have a definite sign at all times: positive for counterclockwise rotation and negative for clockwise rotation.

• The **configuration space** of the pendulum is a circle, denoted S^1. Points on the circle are parametrized by the deflection angle θ which is defined modulo 2π. Specifying θ tells us the position of the bob.

• The **phase space** of the pendulum is the set of all states of the pendulum: all possible ordered pairs $(\theta, \dot{\theta})$ or equivalently (θ, p_θ). Since p_θ can take any real value, the phase space is the Cartesian product $S^1 \times \mathbb{R}$. This space is an infinite cylinder.

6.2.3 Oscillation through small angles: harmonic motion and clocks

• At low energies $E \gtrsim 0$, the bob always remains close to its point of stable equilibrium (small oscillations, $|\theta| \ll \pi/2$) and we may approximate $\sin\theta \approx \theta$.

• The EOM $\ddot{\theta} = -\omega^2 \sin\theta$ (85) may be approximated by the linear equation for simple harmonic motion $\ddot{\theta} = -\omega^2\theta$.

• The general solution is

$$\theta(t) = A\cos\omega t + B\sin\omega t \tag{91}$$

with A and B dimensionless constants of integration. They are related to the initial angle and initial angular velocity by $\theta(0) = A$ and $\dot{\theta}(0) = B\omega$.

• Putting $A = \theta_{\max}\sin\phi$ and $B = \theta_{\max}\cos\phi$, the solution is

$$\theta(t) = \theta_{\max}\sin(\omega t + \phi) \quad \text{with} \quad \theta(0) = \theta_{\max}\sin\phi \quad \text{and} \quad \dot{\theta}(0) = \omega\theta_{\max}\cos\phi. \tag{92}$$

• It is clear that the deflection angle is a sinusoidally varying function of time.

• The maximum angle of deflection $\theta_{\max}$ is called the 'amplitude' of small oscillations.

• Of course, for the small angle approximation to hold, $\theta_{\max}$ must be small, say, compared to $\pi/2$. (How small depends on the accuracy desired.)

• The time period of these small oscillations $T = 2\pi/\omega = 2\pi\sqrt{\ell/g}$ is not just independent of the bob's mass, but also independent of the energy or amplitude $\theta_{\max}$.

• We say that a pendulum executing small oscillations is isochronous, as discovered experimentally by Galileo. This is a feature that allowed pendulums to be used as the most accurate clocks ('chronometers') from the mid 1600s (Christiaan Huygens, 1656) to the early 20th century. The best pendulum clocks had an accuracy of about a second per day.

• Pendulums were also used as gravimeters, to measure the variation of the acceleration due to gravity over the surface of the Earth. Indeed, it was found that pendulums of the same length lose time (the pendulum clock 'runs slow') near the equator and gain time ('runs fast') at high latitudes.

• It was inferred that the acceleration due to gravity g is smaller near the equator and grows with latitude. This is explained by the fact that the Earth bulges out near the equator and is flattened at the poles. It may be modeled as an ellipsoid of revolution called an oblate spheroid or oblate ellipsoid. The difference between the equatorial (6378km) and polar (6357km) semiaxes is about 20 km.

• It was also found empirically that even if the oscillations are not small, the motion is still periodic, though the time period grows with amplitude $\theta_{\max}$.

• However, the nonlinear equation of motion $\ddot{\theta} = -\omega^2 \sin\theta$ cannot be solved in general using elementary functions like polynomials or trigonometric or exponential functions of time. The solution defines a new class of functions called elliptic functions.

7 From Kepler's laws of planetary motion to Newton's law of gravitation

7.1 Kepler's laws of planetary motion

• Based on the Danish astronomer Tycho Brahe's naked eye observations of planetary positions (accurate to better than 0.01°), the German mathematical astronomer Johannes Kepler formulated (1606-1619) three laws of planetary motion around the Sun:

1. Planetary orbits are ellipses with the Sun at a focus. In particular, each orbit lies on a plane, the ecliptic plane of the planet.
2. The radius vector connecting the Sun to a planet sweeps out equal areas in equal times ('constant areal speed').
3. The square of the period of revolution is proportional to the cube of the semimajor axis, with a proportionality constant that is approximately

the same for all planets

$$R^3 = KT^2 \quad \text{where} \quad K \approx 7.5\times10^{-6}(\mathrm{AU})^3/(\mathrm{day})^2 = 3.4 \times 10^{18}\ \mathrm{m}^3/\mathrm{s}^2 \tag{93}$$

is 'Kepler's constant'. An astronomical unit AU is roughly the mean Sun-Earth distance, approximately 150 million kilometers. By considering the case of the Earth, we may estimate the numerical value $K \approx 1/365^2\ (\mathrm{AU})^3/(\mathrm{day})^2$.

• In what follows, we will address the so-called inverse problem of deducing Newton's universal inverse-square force law of gravitation from Kepler's laws and Newton's second law $\boldsymbol{F} = m\boldsymbol{a}$.

• The general problem of deducing a 3d force field $\boldsymbol{F}(\boldsymbol{r})$ for the force felt by a planet in the neighborhood of the Sun is quite hard.

• We will nevertheless find a solution to this problem

1. by considering special cases (circular orbits),
2. by making some physically justified simplifying assumptions (that the force is central) and
3. by using some general features of Newtonian mechanics (angular momentum conservation in a central potential).

7.2 Ecliptic plane, polar coordinates and conservation of angular momentum

• To begin with, we set up a coordinate frame.

• In view of Kepler's first law, the planet's orbit lies on a plane (the ecliptic plane), which we take to be the x-y plane. [The word ecliptic is related to eclipses rather than to ellipses!]

• We will find it useful to define the z axis so that xyz becomes a right-handed system. The z-direction will be helpful to discuss angular momentum.

• We use spherical polar coordinates (r, θ, ϕ) for the planet's location $\boldsymbol{r} = (x, y, z)$ with the Sun at the origin (see Fig. 8):

$$z = r\cos\theta, \quad x = r\sin\theta\cos\phi \quad \text{and} \quad y = r\sin\theta\sin\phi. \tag{94}$$

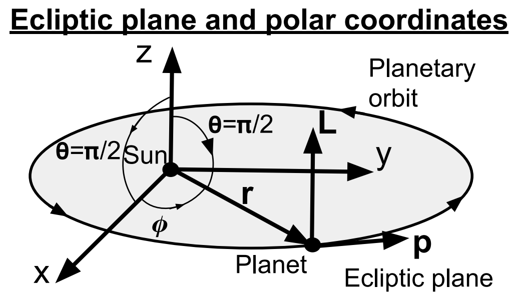


Figure 8: Spherical polar coordinates with the Sun at the origin and the planet moving on the x-y ecliptic plane.

• Both bodies are treated as point particles as they are small compared to the observed sizes of orbits. The Sun's radius is 7×10^8m while the mean Sun-Earth distance is 1.5×10^{11}m.

• Since the planet moves on the x-y plane, $z = 0$ and $\theta \equiv \pi/2$ and both the position and momentum vectors $\boldsymbol{r}, \boldsymbol{p}$ of the planet lie in the x-y plane.

• It follows that the angular momentum about the origin $\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p}$ must point in the $\hat{z}$ or $-\hat{z}$ direction.

• Since planets are observed to go around the Sun without switching direction, the direction of angular momentum cannot change, it must point along $\hat{z}$ or $-\hat{z}$ at all times.

• We will denote the magnitude of angular momentum by l, $\sqrt{\boldsymbol{L} \cdot \boldsymbol{L}} = l$ and its z-component by $\boldsymbol{L} \cdot \hat{z} = l_z$.

• In spherical polar coordinates, the angular momentum can be expressed as $\boldsymbol{L} = l_z \hat{z} = (x p_y - y p_x)\hat{z} = m r^2 \sin^2\theta \dot{\phi}\, \hat{z}$. Since $\theta = \pi/2$, $l_z = m r^2 \dot{\phi}$.

• Now, Kepler's second law may be used to deduce that the magnitude of angular momentum is constant in time.

• Indeed, as shown in Fig. 9, the infinitesimal area (dAr) swept out by the line joining the Sun to the planet in a small time dt while the planet's angular position changes $d\phi$ is $d\text{Ar} \approx \frac{1}{2} r \, r d\phi$.

• Dividing by dt and taking the limit, the constancy of the areal speed

$$\frac{d\text{Ar}}{dt} = \frac{1}{2} r^2 \dot{\phi} = \frac{l_z}{2m} \tag{95}$$

implies angular momentum is conserved.

• We ignore here the small change in area that results from a change in r,

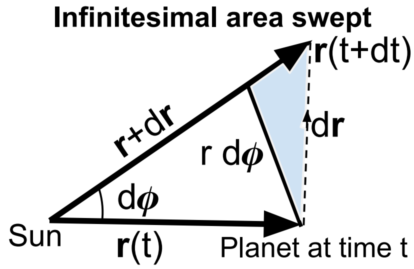


Figure 9: Infinitesimal area swept out by a planet. Roughly, we ignore the shaded region whose area is second order in the infinitesimals $d\phi$ and dr. The area of the isosceles triangle, if we approximate its base by the circular arc length is $\frac{1}{2}r(t)d\phi \times r(t)\cos(d\phi/2)$. For small $d\phi$, $\cos(d\phi/2) \approx 1$.

for this area, given approximately by the shaded triangle in Fig. 9 is 2^{nd} order in infinitesimals, $\propto d\phi\ dr$. When this is divided by dt and we let $dr, d\phi, dt \to 0$, this term vanishes, unlike the leading term given in (95).

• A force field $\boldsymbol{F}(\boldsymbol{r})$ where $\boldsymbol{r}$ is the radius vector from the Sun to the planet is called central if it points radially everywhere and has a magnitude that depends only on the radial distance r.

• It is an independent mathematical fact of Newtonian dynamics (following from Newton's second law) that angular momentum is conserved in a central force field. This suggests that the gravitational force is central $\boldsymbol{F} = -f(r)\hat{r}$ (with the negative sign for attraction).

• We now wish to use Kepler's 3rd law to fix $f(r)$ to the extent possible.

7.3 Newton's inverse square law of gravitation

• The inverse-square nature of the force is guessed from Kepler's third law $T^2 = r^3/K$.

• To see this, we first note that the eccentricity of several planetary orbits is fairly small ($\epsilon = .02$ for the Earth) and they approximately describe uniform circular motion around the Sun.

• Kepler's 3rd law certainly applies to these planets and let us see what it implies.

• Consider a planet such as the Earth moving uniformly around a circle of radius r at constant angular speed ω. It takes a time $T = 2\pi/\omega$ to go round once. Its linear speed is $v = 2\pi r/T = r\omega$.

• To find its acceleration we write its position vector in the x-y plane as $\boldsymbol{r} = (r\cos\omega t, r\sin\omega t)$, from which we find its velocity $\boldsymbol{v} = \dot{\boldsymbol{r}} = \omega r(-\sin\omega t, \cos\omega t)$. Its acceleration is $\boldsymbol{a} = \ddot{\boldsymbol{r}} = -\omega^2 r(\cos\omega t, \sin\omega t)$. We see that $\boldsymbol{a} = -\omega^2\boldsymbol{r} = -(v^2/r)\hat{r}$: the acceleration is directed radially inwards, and therefore called centripetal.

• Newton's 2$^{\text{nd}}$ law requires the (radially inward) centripetal acceleration times (inertial) mass of the earth to equal the inward gravitational force. This gives

$$-m_e\frac{v^2}{r}\hat{r} = -f(r)\hat{r} \quad \text{with} \quad v^2 = \frac{(2\pi r)^2}{T^2} = \frac{K(2\pi r)^2}{r^3} \quad \Rightarrow \quad f(r) = \frac{4\pi^2 K m_e}{r^2}. \tag{96}$$

• Besides its inverse-square nature, the above gravitational force on the Earth due to the Sun is proportional to the Earth's mass m_e (since K is independent of the planet) so that the Earth's acceleration is *independent* of its mass.

• Newton postulated that this must be true also of the force felt by the Sun due to the Earth (his 3$^{\text{rd}}$ law) and concluded that $K \propto m_s$. Thus, we guess the universal (both terrestrial and celestial) law of gravitation

$$\boldsymbol{F} = -\frac{G m_s m_e}{r^2}\hat{r} \quad \text{where} \quad G = \frac{4\pi^2 K}{m_s} = 6.67\times 10^{-11}\ \text{Nm}^2/\text{kg}^2. \tag{97}$$

• **Remark:** In this 'derivation' of Newton's law of gravitation, m_e was the inertial mass of the Earth, since it came from the mass $\times$ acceleration term in Newton's 2nd law. By Newton's 3rd law argument, m_s would then have to be the inertial mass of the sun. So one may wonder where there is room for gravitational masses that may be distinct from inertial masses in Newton's law of gravity. This apparent lack of room is because we assumed that Kepler's constant K is independent of the planet. This is analogous to Galileo's assertion that all bodies fall the same way when dropped from the same height above the Earth's surface. In fact, K is not exactly the same for all planets. To incorporate this dependence on the planet we could write $Km_e = km_e^g$ where k is independent of the planet and m_e^g is a property of the planet which we choose to call its gravitational mass. Then, by Newton's third law we would write $k = k'm_s^g$ where m_s^g is the gravitational mass of the Sun and k' is a constant independent of both bodies. Finally, denoting $4\pi^2 k'$ by the symbol G we get Newton's law for the gravitational force of the Sun on the Earth with magnitude $G m_s^g m_e^g / r^2$. [As it turns out, the dependence of K on planet is not due to a difference between gravitational and inertial masses and can be explained by other effects that we have ignored: effects of other planets, finite sizes of sun and planets etc.]

• We will often use the abbreviation $\alpha = G m_e m_s$. Note that $m_e \approx 6 \times 10^{24}$ kg and $m_s \approx 2 \times 10^{30}$ kg. Though we will not do it here for lack of time, Kepler's first law on elliptical orbits may now be derived by solving Newton's equation of motion using his universal law of gravitation.

• An important feature of the gravitational force is that it is derivable from the gravitational potential $V(r) = -\alpha/r$:

$$\boldsymbol{F} = -\frac{\alpha}{r^2}\hat{r} = -\boldsymbol{\nabla}_{\boldsymbol{r}}\left(-\frac{\alpha}{r}\right) = -\boldsymbol{\nabla}_{\boldsymbol{r}} V(r). \tag{98}$$

• Here $\boldsymbol{\nabla}_{\boldsymbol{r}}$ is the vector gradient. If $\boldsymbol{r} = (x, y, z)$ are the Cartesian components of the position vector, then $\boldsymbol{\nabla}_{\boldsymbol{r}} = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})$.

• The gravitational force is called central since it points radially and its magnitude depends only on the radial distance. The corresponding potential $V(r)$ depends only on the distance from the origin and is called a central or spherically symmetric potential.

• In obtaining the $1/r$ potential from Kepler's laws, we have in effect solved an *inverse problem*, i.e., to deduce a potential from features of trajectories. More specifically, we deduced a potential from the period of oscillations that it supports.

• To solve such a problem in general is very difficult, and Isaac Newton and his contemporaries like Edmond Halley and Robert Hooke are lucky to have succeeded in this case of central importance.

• **The lunar perigee problem.** It took time for Newton's inverse-square law of gravitational force to be tested and widely accepted. Newton (1643-1727) showed that it did give the observed elliptical orbits of some planets and comets.

• However, a difficulty arose in applying it to lunar motion. The moon's approximately elliptical orbit around the Earth was observed to precess [it isn't quite closed or periodic]. The lunar perigee (closest approach to the Earth) rotates by 40° of arc per annum.

• This precession was believed to be due to the Sun's effect on the Earth-Moon system. This leads to the famous three-body problem of celestial mechanics.

• Newton tried to estimate the motion of the lunar perigee due to the Sun's effect, but could account for only about half of it.

• There were lingering doubts about whether the $1/r^2$ force law was correct and Alexis Clairaut even proposed a small $1/r^3$ correction to it. A lively competition ensued among Euler, Clairaut and d'Alembert to develop a theory of solar perturbations to the Moon's orbit. [See S Bodenmann, *The 18th-century battle over lunar motion*, Physics Today, **63(1)**, 27 (2010) doi: 10.1063/1.3293410].

• Eventually, Clairaut (1759) was able to use the purely inverse-square force law to account for 85% of the motion of the lunar perigee using a third order perturbative treatment of the Sun's effect.

• This and subsequent work on other solar system trajectories led to widespread acceptance of Newton's laws of motion and gravity.

8 Motion in uniformly accelerating frames

8.1 Uniformly accelerating systems

• Having dealt with frames moving at a constant velocity with respect to an inertial frame, we now progress to the nature of dynamics as observed from a frame S' that is uniformly accelerating at the rate $\boldsymbol{A}$ relative to an inertial frame S. This is the first step in the study of dynamics in a non-inertial frame. One could subsequently consider a frame that accelerates nonuniformly or rotates.

• The acceleration of a particle in S' is related to that in S by $\boldsymbol{a}_i' = \boldsymbol{a}_i - \boldsymbol{A}$. Multiplying by m_i,

$$m_i \boldsymbol{a}' = m_i \boldsymbol{a} - m_i \boldsymbol{A} = \boldsymbol{F} - m_i \boldsymbol{A}. \tag{99}$$

We used the fact that S is inertial to write $m_i \boldsymbol{a} = \boldsymbol{F}$.

• We will call $\boldsymbol{F}' = m_i \boldsymbol{a}'$ the (apparent) force on the particle in the accelerated frame. We assume that the inertial mass of the particle is the same in both frames.

• Unlike in a frame that moves uniformly, the force seen in a noninertial frame is generally *not* the same as in an inertial frame.

• The apparent force in S' may be expressed as $\boldsymbol{F}' = \boldsymbol{F} + \boldsymbol{F}_{\text{fict}}$ where the fictitious force is given by $\boldsymbol{F}_{\text{fict}} = -m_i \boldsymbol{A}$.

• Notice that this fictitious force points oppositely to the direction of acceleration $\boldsymbol{A}$.

• Moreover, the fictitious force on a particle $\boldsymbol{F}_{\text{fict}}$ is uniform, i.e., independent of location and proportional to the particle's inertial mass.

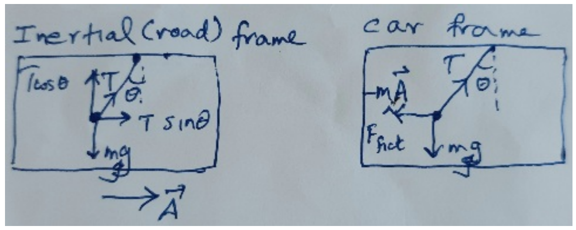
• This is similar to the gravitational force $m_g \boldsymbol{g}$ on a particle of gravitational mass m_g on the Earth's surface, which is roughly uniform and proportional to the gravitational mass. However, $\boldsymbol{F}_{\text{fict}}$ arises from the acceleration of the frame rather than from interactions between particles, which explains the name fictitious.

• **Example: Fictitious force in an accelerating car.** A small bob of inertial mass m_i is suspended by a stretched string from the roof of a car that accelerates at the rate $\boldsymbol{A}$. The situation is viewed from the inertial frame of the road as well as from the accelerated frame of the car. Viewed from either frame, the string is seen to settle into a configuration where it hangs at a nonzero angle relative to the vertical, with the bob displaced towards the back of the car. The angle that the string makes with the vertical as well as the tension in the string are found to be the same in both frames and given by

$$\theta = \arctan(m_i A / m_g g) \quad \text{and} \quad T = \sqrt{m_g^2 g^2 + m_i^2 A^2}. \tag{100}$$

Obtain these results by working (a) in the inertial frame of the road and (b) in the accelerated frame of the car.

• For instance, in the inertial road frame, the bob accelerates in the direction of motion of the car and the vertical component of the string tension is balanced by the weight of the bob. On the other hand, in the accelerated frame of the car, the bob is at rest and the tension in the string is balanced



by the resultant of the weight of the bob and the fictitious force.

8.2 Principle of equivalence

• We have seen that Newton's 2nd law in a frame that accelerates uniformly at the rate $\boldsymbol{A}$ is the same as that in an inertial frame, provided we introduce a fictitious force $\boldsymbol{F}_{\text{fict}} = -m_i \boldsymbol{A}$ on each particle in the accelerated frame.

• Aside from the appearance of m_i and m_g, the fictitious force due to a frame's acceleration $\boldsymbol{A}$ is similar to the gravitational force due to a uniform acceleration due to gravity $\boldsymbol{g} = -\boldsymbol{A}$.

• Einstein posed the question of whether the two are distinguishable in principle. For instance, in an inertial frame in which a local gravitational field is present (with constant acceleration due to gravity $\boldsymbol{g}$), a free particle of mass m_g is subject to a force $\boldsymbol{F} = m_g \boldsymbol{g}$.

• Now, suppose this particle is isolated far from all physical interactions and is observed in a frame that is uniformly accelerated at the rate $\boldsymbol{A} = -\boldsymbol{g}$. In this frame, it is subject to a fictitious force $\boldsymbol{F}_{\text{fict}} = -m_i \boldsymbol{A} = m_i \boldsymbol{g}$. Einstein asked whether it is possible for an observer in either of these frames to physically distinguish between these two situations.

• To make the question more concrete, he proposed the following thought ('gedanken') experiment.

• Suppose a man is on an elevator at rest in a uniform gravitational field of magnitude g that points downwards. If he releases a ball from rest, it falls downwards with an acceleration $\boldsymbol{a} = (m_g/m_i)\boldsymbol{g}$.

• On the other hand, suppose the same elevator is far from other physical interactions and accelerates upwards at the rate $a = g$. If the man releases the ball in the same manner in this accelerating elevator, it accelerates downward at the rate g.

• Now, if $m_i = m_g$, then as far as the man is concerned, the two situations are identical and he has no way to distinguish between an accelerating elevator and a gravitational field.

• This may be regarded as a 20th century version of Galileo's assertion from the 17th century that is not possible to dynamically distinguish between an inertial frame and one that is moving at constant velocity relative to it.

• **Freely falling elevator.** Another way to look at the equivalence between a uniform gravitational field and an accelerating frame is to consider an elevator that is freely falling in a uniform gravitational field.

• The elevator, the man and the ball accelerate downwards at the rate g. If the man releases the ball, it would remain suspended as though the elevator is motionless far from other physical interactions.

• If $m_i = m_g$, then the force of magnitude $m_g g$ due to the local gravitational field is exactly cancelled by the upward fictitious force of magnitude $m_i g$ due to the downward acceleration of the elevator. Einstein concluded that if inertial and gravitational masses are the same, then an observer in the elevator, cannot determine whether the elevator is falling in a uniform gravitational field or is isolated in outer space.

• There is strong experimental evidence that inertial and gravitational masses can be taken to be equal $m_i = m_g$.

• The **Principle of equivalence** asserts that the inertial and gravitational masses of any body are numerically equal $m_i = m_g$ and that it is not possible to distinguish locally between a uniform gravitational field with acceleration $\boldsymbol{g}$ (observed from an elevator that is at rest or moving uniformly relative to an inertial frame) and a coordinate system (elevator) that accelerates at the rate $\boldsymbol{A} = -\boldsymbol{g}$. The restriction to 'local' is to avoid inhomogeneities in real gravitational fields over large distances and the possibility of the man looking out of the elevator etc.

• The principle of equivalence lies at the foundation of Einstein's theory of gravitation.

9 Special theory of relativity

9.1 Difficulties with Newtonian mechanics

• Newton, in his formulation of mechanics, in effect, postulated that all observers, irrespective of their state of relative motion could synchronize their clocks to a common time and could agree about the simultaneity of a pair of events.

• In the late 1800s, Ernst Mach criticized this postulate as it was not backed by an examination of the physical processes involved in measuring times using clocks carried by various observers

• Nevertheless, Newton's assumption about time remained at the foundation of mechanics for nearly 200 years, since its consequences were largely in agreement with empirical observations.

• With the benefit of hindsight, we may say that the Newtonian assumption is a good approximation as long as speeds (of particles or of observers) are much less than that of light. In effect, Newton had assumed that light traveled infinitely fast. It is perhaps not a surprise that difficulties in the application of Newtonian mechanics first arose in matters concerning light.

• First, by the 1800s there was strong empirical evidence that light traveled at a large but finite speed $\approx 3 \times 10^8$ m/s in vacuum.

• Next, Maxwell's equations that had been hugely successful in understanding electricity and magnetism predicted the existence of electromagnetic waves, which included light. The equations did not seem to refer to any medium and the speed of the wave was a constant (denoted c) which did not seem to depend on how the waves were produced or observed (c did not depend on the motion of the source or observer).

• Other types of waves known at that time, such as waves in a stretched string, sound, elastic waves in a solid and water waves were traveling disturbances in a medium like air or water. Moreover, the observed speed of sound depended on the motion of the observer through the medium.

• Based on the analogy with sound (and the overwhelming success of mechanics in describing waves and other natural phenomena) it was assumed that light waves too must travel in an as yet unobserved medium, which was called ether. Ether had to have some peculiar properties: (a) to allow light to travel very fast, it had to be minimally deformable (i.e., very rigid, since sound travels faster in a solid than in a gas) but (b) it had to be very rare to have evaded detection through its effect on the motion of celestial or terrestrial bodies.

• Michelson set out to detect the effect of the motion of an observer (relative to the proposed ether medium) on the observed speed of light.

• Using an interferometer, he (in 1881) and later he along with Morley (in 1887), obtained a null result: the speed of light was the same for a variety of observers moving differently through the ether (the orientation of the arms of the interferometer and motion of the Earth altered the direction of motion through the proposed ether). This seemed to be at odds with expectations based on Newtonian mechanics.

• Attempts were made to explain this null result, but often seemed contrived and typically introduced other complications. For instance, FitzGerald and Lorentz proposed that motion through the ether produced just the right contraction of one arm of the interferometer to cancel the effect of the change in speed of light due to the relative motion.

9.2 Postulates of special relativity

• In 1905, Einstein brought clarity to this confusing situation when he introduced his special theory of relativity.

• Einstein said he was not influenced by the null result of Michelson and Morley, as he had already come to a similar conclusion based on H Fizeau's 1851 experiments that attempted to detect the effect of a moving medium (water and air) on the speed of light. Fizeau found an unexpectedly small effect in water and no effect in air.

• Mach's critique of Newtonian mechanics and Maxwell's electrodynamics significantly influenced Einstein's thinking.

• Einstein's insight was that the principles of Newtonian mechanics had to be revised so as to make mechanics compatible with Maxwell's electrodynamics and avoid the contradictions in interpreting phenomena involving light.

• Special relativity is based on a pair of postulates: the principle of relativity and the constancy of the speed of light. The qualifier 'special' is meant to convey that the theory deals with physical phenomena observed from inertial frames of reference. Relativity refers to the fact that certain concepts such as simultaneity of events or lengths of rods, which were considered absolute in Newtonian mechanics, are defined relative to an observer.

• **The principle of relativity** goes back to Galileo, who had proposed it in the context of mechanics. Galileo had pointed out that there is no mechanical way of distinguishing between an inertial frame and another frame that is in uniform motion relative to it. In particular, the concept of a frame that is 'absolutely at rest' is not meaningful in mechanics.

• Einstein argued that this principle must apply to all the laws of physics: *if a law holds in one inertial frame, it must also hold in any other frame moving at a constant velocity relative to it.*

• Accepting the principle of relativity meant discarding the ether hypothesis since the latter would mean that there is a distinguished inertial frame in which the ether is at rest and in which Maxwell's equations are valid.

• The failure to detect motion relative to the ether suggested that it was more economical to discard the ether concept while retaining the principle of relativity.

• Moreover, Maxwell's equations predicted a universal speed for light, without reference to any medium, so there was no role for ether in electromagnetism (though Maxwell erroneously believed in it). This brings us to the 2nd postulate of special relativity.

• *The speed of light in vacuum is the same constant c in all inertial frames.* Since Maxwell's theory seemed to predict a constant speed of light independent of the motion of the source or observer and without reference to any medium, Einstein postulated that the speed of light must be the same for all inertial observers. In particular, the speed of light is the same irrespective of the direction in which the light propagates.

• In other words, the analogy with sound waves, which require a medium for their propagation and whose speed depends on the motion of the observer and source, has its limitations.

• These postulates seemed to be the simplest way to accommodate the available theoretical knowledge (constancy of speed of light in Maxwell theory, absence of distinguished frame in Galilean relativity) and experimental/observational data (observer independence of speed of light, ether going undetected).

• In addition to these postulates, Einstein was guided by what one might call the correspondence principle: the relativistic generalization of the laws of mechanics and rules for transforming physical quantities between inertial frames must reduce to those of Newton and Galileo when speeds are small compared to that of light.

• Thus, although Galileo's principle of relativity is postulated to apply to all the laws of physics, the Galilean transformation formula ($t' = t, \boldsymbol{r}' - \boldsymbol{r} - \boldsymbol{v}t$) telling us how coordinates transform between frames moving at relative velocity $\boldsymbol{v}$ needs to be revised.

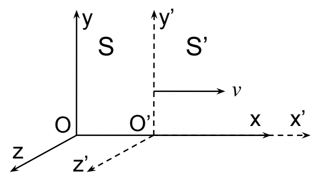
• The name *correspondence principle* was introduced by Neils Bohr in postulating that the laws of quantum mechanics must reduce to those of classical mechanics in the semiclassical limit of large quantum numbers, where appropriate physical quantities with dimensions of action (angular momentum) have numerical values that are large compared to Planck's constant $h = 6.6 \times 10^{-34}$ Js.

9.3 Incompatibility of Galilean transformations with constancy of the speed of light

• To describe events, an observer finds it convenient to set up a coordinate system or frame of reference S. For simplicity, we will consider Cartesian coordinate systems for space consisting of an origin and a right-handed frame of x, y and z axes. In Newtonian mechanics, any event is then specified by the time t at which it occurs, along with the (x, y, z) coordinates of the place where it occurs.

• Galilean transformations. Now consider an inertial frame of reference $S : (t, x, y, z)$. If Galileo's principle of relativity applies to all physical phenomena, then the laws of physics must take the same form in S as they do in a frame S' moving uniformly relative to S. Let us denote the coordinates in frame S' by (t', x', y', z'). How are the coordinates of an event referred to the two systems related?

• In Newtonian mechanics, one can define a common time t for all observers, so $t' = t$. For simplicity, we will suppose that S' and S coincide at $t = 0$ and that S' moves at speed v to the right along the x direction.



• Then the coordinates of the event at (t, x, y, z) when viewed from S' are given by

$$t' = t, \quad x' = x - vt, \quad y' = y, \quad z' = z. \tag{101}$$

This transformation of coordinates is called a Galilean transformation (more precisely a Galilean boost by velocity $v\hat{x}$).

• We observed in §2 that Newton's second law for two gravitating masses is invariant under such a Galilean transformation.

• Now, let us examine how the speed of a light signal behaves under a Galilean boost. We suppose that at $t = 0$, a pulse of light is emitted from the origin of S and travels along the x-axis at the speed c.

• Then, at time t, the light pulse is located at $x = ct$. When viewed from frame S', the location of the light pulse at time t is given by $x' = ct - vt = (c - v)t$. Thus, in frame S', the light pulse moves at speed $\frac{dx'}{dt'} = c - v$.

• However, this is not consistent with the observed universality of the speed of light. Thus, Galilean transformations are not compatible with Einstein's postulate that the speed of light is the same for all inertial observers.

9.4 Synchronization of clocks & simultaneity

• **Synchronization of clocks.** Einstein prefaced his special theory of relativity with a note on synchronization of clocks. Each observer is provided with an identical clock to keep time. To compare measurements of time by the different observers, it is helpful if their clocks are synchronized, i.e., if they agree on the time of a single event.

• In Newtonian mechanics, if a flash of light goes off somewhere, then it arrives instantaneously at the locations of all observers, whose synchronized clocks all assign the same time to the event. This is a reasonable approximation if the speed of light is very large compared to that of observers and bodies under consideration.

• **Synchronization of clocks by the radar method.** In special relativity, the speed of light is finite. Einstein proposed a 'radar' scheme by which all inertial observers (at rest relative to one another) would assign the same time to any given event. Such a system of clocks is said to be synchronized and is of practical value, for instance in coordinating times across the internet.

• Suppose A and B are a pair of observers with identical clocks. Suppose A sends a light signal at time t_A (according to his clock) towards B. Suppose B receives this light signal at a time t_B on her clock. To synchronize A's clock with that of B, B immediately sends a light pulse back (with a message giving the time t_B), which is received by A at a time $t_A + 2\Delta$.

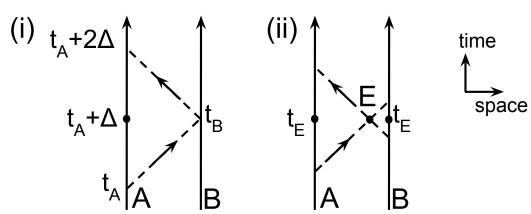


Figure 10: Space-time diagrams of two inertial observers A and B who are at rest relative to each other. Time increases upwards while the horizontal direction is space. The trajectories of A and B ('world lines') are upward-directed solid straight lines. World lines of light signals are indicated by dashed lines inclined at 45° since we work in units where $c = 1$. (i) A synchronizes his clock with that of B by sending and receiving a light signal and calibrating his clock to read t_B at the time $t_A + \Delta$. (ii) Having synchronized their clocks, both observers assign the same time to any event E.

• Now, the clocks are synchronized if A shifts the time displayed by his clock in such a way that $t_B = t_A + \Delta$ (see Fig. 10). Having synchronized them, both clocks will assign the same time to any event E, as shown in Fig. 10.

• Note that if B is moving relative to A then this synchronization cannot be maintained. More on this soon.

• Note: when we speak of light, we do not restrict to visible light. We may use electromagnetic waves of any frequency (X-rays, microwaves etc.) as they all travel at the speed of light. We use light instead of electrons or protons since light signals have a common speed for all inertial observers, which simplifies the discussion. Electrons of different kinetic energies have different speeds and their speed depends on the observer.

• **Relativity of simultaneity.** Two events E_1 and E_2 occur simultaneously for observer A if they occur at the same time on A's clock. The above radar method of synchronization ensures that if B is at rest relative to A, then E_1 and E_2 are simultaneous for B as well. However, the space-time diagram of Fig. 11 and the example that follows show that observers in relative motion need not agree on the simultaneity of events. Thus, synchronization of clocks of observes in relative motion is in general not possible.

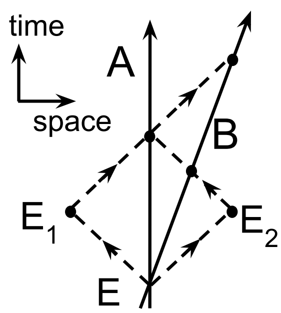
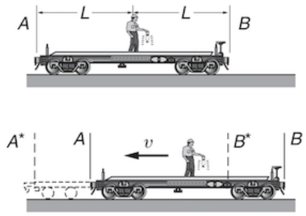


Figure 11: *Relativity of simultaneity.* In this figure, we use units where $c = 1$ so light rays travel along straight lines inclined at $45°$. Observer B moves at a uniform velocity $< c$ relative to inertial observer A, so their world lines (solid) are inclined at an angle less than $45°$. Time, as measured on their clocks increases upwards along their world lines. They meet at the event E, when both send light signals (dashed) to events E_1 and E_2, which are then sent back. Events E_1 and E_2 are simultaneous according to A (as he receives the returned signals at the same time on his clock) but not as reckoned by B. B says that E_2 precedes E_1 as she receives the reflected light signal from E_2 before the one from E_1.

• A guard stands in the middle of an empty railway wagon of length $2L$ and switches on his lantern sending out a light pulse in all directions propagating at speed c. In the rest frame of the wagon, the light pulse arrives simultaneously at the left and right ends A and B of the wagon after a time L/c.

• Now let us view these events from a frame that is moving to the right at speed v relative to the railway wagon. In this frame, the wagon moves to the left with speed v. Though the light pulse travels at the same speed c



in this frame, it has to cover a distance shorter than L to reach the right end and a distance longer than L to reach the left end. Thus, in this frame, the pulse does not arrive simultaneously at the two ends of the wagon!