A-LEVEL NOTES

MECHANICS
(for Kate)

March 2021
version 0.1

A-LEVEL MECHANICS
Mathematics and Further Mathematics
(March 2021)

This document is a self contained set of MECHANICS notes for A level Mathematics and Further Mathematics. The notes assume that the student is familiar with A-LEVEL calculus. Material on *vectors and matrices*, usually offered in a calculus course, is repeated here to provide continuity.

These notes are available as an *open source* document from:
https://hughmurrell.github.io/.

This work is largely derived from a set of student notes written up by **Dexter Chua** whilst an undergraduate at Cambridge. The courses in question were *Dynamics and Relativity* taught by **G. I. Ogilvie** in the Lent term of 2015 and *Vectors and Matrices* taught by **N. Peake** in the Michaelmas term of 2014.
The original versions of Dexter's course notes can be found here:
https://dec41.user.srcf.net/notes/

For this document, Dexter's notes have been augmented with worked examples from **Thomas Backman's** notes on mechanics. Thomas's problems have been selected and somewhat modified to fit the theme of the rest of the notes but the interested reader can view the originals here:
https://github.com/exscape/8.01x-notes

This collection also includes worked problems from past A-level and STEP (Sixth Term Examination Paper) papers. Further worked problems will be added in due course depending on reader engagement with the collection.

The intention is that this set of Mechanics notes will provide preparation material for the following A-LEVEL papers:

Mathematics	Further Mathematics	STEP
Paper 4 (Mechanics)	Paper 3 (Mechanics)	Section B (Mechanics)

Prospective A-level students and students planning to write the *Sixth Term Examination Paper* (STEP), are encouraged to make use of this text to supplement their A-level materials. Answers and hints to selected exercises are available in the appendix. Further assistance with challenging problems can be obtained via email. To obtain help, propose new problems or point out errors please feel free to email hugh.murrell@gmail.com.

Contents

Chapter 1

Vectors and Matrices

1.1 Introduction

Vectors and matrices is the language in which a lot of mathematics is written in. In physics, many variables such as position and momentum are expressed as vectors. Heisenberg also formulated quantum mechanics in terms of vectors and matrices. In statistics, one might pack all the results of all experiments into a single vector, and work with a large vector instead of many small quantities. In group theory, matrices are used to represent the symmetries of space (as well as many other groups).

So what is a vector? Vectors are very general objects, and can in theory represent very complex objects. However, in this course, our focus is on vectors in $\mathbb{R}^n$ or $\mathbb{C}^n$. We can think of each of these as an array of n real or complex numbers. For example, $(1, 6, 4)$ is a vector in $\mathbb{R}^3$. These vectors are added in the obvious way. For example, $(1, 6, 4)+(3, 5, 2) = (4, 11, 6)$. We can also multiply vectors by numbers, say $2(1, 6, 4) = (2, 12, 8)$. Often, these vectors represent points in an n-dimensional space.

Matrices, on the other hand, represent ***functions*** between vectors, i.e. a function that takes in a vector and outputs another vector. These, however, are not arbitrary functions. Instead matrices represent ***linear functions***. These are functions that satisfy the equality $f(\lambda\mathbf{x} + \mu\mathbf{y}) = \lambda f(\mathbf{x}) + \mu f(\mathbf{y})$ for arbitrary numbers λ, μ and vectors $\mathbf{x}, \mathbf{y}$. It is important to note that the function $\mathbf{x} \mapsto \mathbf{x} + \mathbf{c}$ for some constant vector $\mathbf{c}$ is ***not*** linear according to this definition, even though it might look linear.

It turns out that for each linear function from $\mathbb{R}^n$ to $\mathbb{R}^m$, we can represent the function uniquely by an $m \times n$ array of numbers, which is what we call the ***matrix***. Expressing a linear function as a matrix allows us to conveniently study many of its properties, which is why we usually talk about matrices instead of the function itself.

1.2 Complex numbers

In $\mathbb{R}$, not every polynomial equation has a solution. For example, there does not exist any x such that $x^2+1=0$, since for any x, x^2 is non-negative, and x^2+1 can never be 0. To solve this problem, we introduce the "number" i that satisfies $i^2=-1$. Then i is a solution to the equation $x^2+1=0$. Similarly, $-i$ is also a solution to the equation.

We can add and multiply numbers with i. For example, we can obtain numbers $3+i$ or $1+3i$. These numbers are known as ***complex numbers***. It turns out that by adding this single number i, ***every*** polynomial equation will have a root. In fact, for an nth order polynomial equation, we will later see that there will always be n roots, if we account for multiplicity. We will go into details in Chapter 1.6.

Apart from solving equations, complex numbers have a lot of rather important applications. For example, they are used in electronics to represent alternating currents, and form an integral part in the formulation of quantum mechanics.

Basic properties

Definition (Complex number)**.** A ***complex number*** is a number $z\in\mathbb{C}$ of the form $z=a+ib$ with $a,b\in\mathbb{R}$, where $i^2=-1$. We write $a=\mathrm{Re}(z)$ and $b=\mathrm{Im}(z)$.

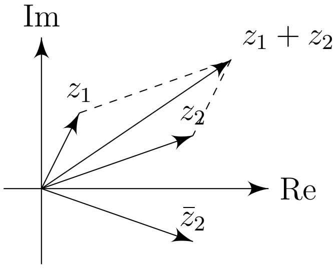
We have

$$\begin{aligned}
z_1\pm z_2 &= (a_1+ib_1)\pm(a_2+ib_2)\\
&= (a_1\pm a_2)+i(b_1\pm b_2)\\
z_1z_2 &= (a_1+ib_1)(a_2+ib_2)\\
&= (a_1a_2-b_1b_2)+i(b_1a_2+a_1b_2)\\
z^{-1} &= \frac{1}{a+ib}\\
&= \frac{a-ib}{a^2+b^2}
\end{aligned}$$

Definition (Complex conjugate)**.** The ***complex conjugate*** of $z=a+ib$ is $a-ib$. It is written as $\bar{z}$ or z^*.

It is often helpful to visualize complex numbers in a diagram:

Definition (Argand diagram)**.** An ***Argand diagram*** is a diagram in which a complex number $z=x+iy$ is represented by a vector $\mathbf{p}=\begin{pmatrix}x\\y\end{pmatrix}$. Addition of vectors corresponds to vector addition and $\bar{z}$ is the reflection of z in the x-axis.



Definition (Modulus and argument of complex number)**.** The ***modulus*** of $z = x + iy$ is $r = |z| = \sqrt{x^2 + y^2}$. The ***argument*** is $\theta = \arg z = \tan^{-1}(y/x)$. The modulus is the length of the vector in the Argand diagram, and the argument is the angle between z and the real axis. We have

$$z = r(\cos\theta + i\sin\theta)$$

Clearly the pair (r, θ) uniquely describes a complex number z, but each complex number $z \in \mathbb{C}$ can be described by many different θ since $\sin(2\pi+\theta) = \sin\theta$ and $\cos(2\pi+\theta) = \cos\theta$. Often we take the ***principle value*** $\theta \in (-\pi, \pi]$.

When writing $z_i = r_i(\cos\theta_i + i\sin\theta_i)$, we have

$$\begin{aligned} z_1 z_2 &= r_1 r_2[(\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2) + i(\sin\theta_1\cos\theta_2 + \sin\theta_2\cos\theta_1)] \\ &= r_1 r_2[\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)] \end{aligned}$$

In other words, when multiplying complex numbers, the moduli multiply and the arguments add.

Proposition. $z\bar{z} = a^2 + b^2 = |z|^2$.

Proposition. $z^{-1} = \bar{z}/|z|^2$.

Theorem (Triangle inequality)**.** For all $z_1, z_2 \in \mathbb{C}$, we have

$$|z_1 + z_2| \leq |z_1| + |z_2|.$$

Alternatively, we have $|z_1 - z_2| \geq ||z_1| - |z_2||$.

Complex exponential function

Exponentiation was originally defined for integer powers as repeated multiplication. This is then extended to rational powers using roots. We can also extend this to any real number since real numbers can be approximated arbitrarily accurately by rational numbers. However, what does it mean to take an exponent of a complex number?

To do so, we use the Taylor series definition of the exponential function:

Definition (Exponential function)**.** The ***exponential function*** is defined as

$$\exp(z) = e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

This automatically allows taking exponents of arbitrary complex numbers. Having defined exponentiation this way, we want to check that it satisfies the usual properties, such as $\exp(z + w) = \exp(z)\exp(w)$. To prove this, we will first need a helpful lemma.

Lemma.

$$\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} a_{mn} = \sum_{r=0}^{\infty}\sum_{m=0}^{r} a_{r-m,m}$$

Proof.

$$\begin{aligned}\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} a_{mn} &= a_{00} + a_{01} + a_{02} + \cdots \\ &\quad + a_{10} + a_{11} + a_{12} + \cdots \\ &\quad + a_{20} + a_{21} + a_{22} + \cdots \\ &= (a_{00}) + (a_{10} + a_{01}) + (a_{20} + a_{11} + a_{02}) + \cdots \\ &= \sum_{r=0}^{\infty}\sum_{m=0}^{r} a_{r-m,m}\end{aligned}$$ □

This is not exactly a rigorous proof, since we should not hand-wave about infinite sums so casually. But in fact, we did not even show that the definition of $\exp(z)$ is well defined for all numbers z, since the sum might diverge. All these will be done in that IA Analysis I course.

Theorem. $\exp(z_1)\exp(z_2) = \exp(z_1 + z_2)$

Proof.

$$\begin{aligned}\exp(z_1)\exp(z_2) &= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\frac{z_1^m}{m!}\frac{z_2^n}{n!} \\ &= \sum_{r=0}^{\infty}\sum_{m=0}^{r}\frac{z_1^{r-m}}{(r-m)!}\frac{z_2^m}{m!} \\ &= \sum_{r=0}^{\infty}\frac{1}{r!}\sum_{m=0}^{r}\frac{r!}{(r-m)!m!}z_1^{r-m}z_2^m \\ &= \sum_{r=0}^{\infty}\frac{(z_1+z_2)^r}{r!}\end{aligned}$$ □

Again, to define the sine and cosine functions, instead of referring to "angles" (since it doesn't make much sense to refer to complex "angles"), we again use a series definition.

Definition (Sine and cosine functions)**.** Define, for all $z \in \mathbb{C}$,

$$\begin{aligned}\sin z &= \sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)!}z^{2n+1} &&= z - \frac{1}{3!}z^3 + \frac{1}{5!}z^5 + \cdots \\ \cos z &= \sum_{n=0}^{\infty}\frac{(-1)^n}{(2n)!}z^{2n} &&= 1 - \frac{1}{2!}z^2 + \frac{1}{4!}z^4 + \cdots\end{aligned}$$

One very important result is the relationship between exp, sin and cos.

Theorem. $e^{iz} = \cos z + i \sin z$.

Alternatively, since $\sin(-z) = -\sin z$ and $\cos(-z) = \cos z$, we have

$$\cos z = \frac{e^{iz} + e^{-iz}}{2},$$
$$\sin z = \frac{e^{iz} - e^{-iz}}{2i}.$$

Proof.

$$\begin{aligned} e^{iz} &= \sum_{n=0}^{\infty} \frac{i^n}{n!} z^n \\ &= \sum_{n=0}^{\infty} \frac{i^{2n}}{(2n)!} z^{2n} + \sum_{n=0}^{\infty} \frac{i^{2n+1}}{(2n+1)!} z^{2n+1} \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} z^{2n} + i \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} z^{2n+1} \\ &= \cos z + i \sin z \end{aligned}$$ □

Thus we can write $z = r(\cos\theta + i\sin\theta) = re^{i\theta}$.

Roots of unity

Definition (Roots of unity)**.** The nth ***roots of unity*** are the roots to the equation $z^n = 1$ for $n \in \mathbb{N}$. Since this is a polynomial of order n, there are n roots of unity. In fact, the nth roots of unity are $\exp\left(2\pi i \frac{k}{n}\right)$ for $k = 0, 1, 2, 3 \cdots n-1$.

Proposition. If $\omega = \exp\left(\frac{2\pi i}{n}\right)$, then $1 + \omega + \omega^2 + \cdots + \omega^{n-1} = 0$

Proof. Two proofs are provided:

(i) Consider the equation $z^n = 1$. The coefficient of z^{n-1} is the sum of all roots. Since the coefficient of z^{n-1} is 0, then the sum of all roots $= 1 + \omega + \omega^2 + \cdots + \omega^{n-1} = 0$.

(ii) Since $\omega^n - 1 = (\omega - 1)(1 + \omega + \cdots + \omega^{n-1})$ and $\omega \neq 1$, dividing by $(\omega - 1)$, we have $1 + \omega + \cdots + \omega^{n-1} = (\omega^n - 1)/(\omega - 1) = 0$. □

Complex logarithm and power

Definition (Complex logarithm)**.** The ***complex logarithm*** $w = \log z$ is a solution to $e^\omega = z$, i.e. $\omega = \log z$. Writing $z = re^{i\theta}$, we have $\log z = \log(re^{i\theta}) = \log r + i\theta$. This can be multi-valued for different values of θ and, as above, we should select the θ that satisfies $-\pi < \theta \leq \pi$.

Example. $\log 2i = \log 2 + i\frac{\pi}{2}$

Definition (Complex power)**.** The ***complex power*** z^α for $z, \alpha \in \mathbb{C}$ is defined as $z^\alpha = e^{\alpha \log z}$. This, again, can be multi-valued, as $z^\alpha = e^{\alpha \log |z|} e^{i\alpha\theta} e^{2in\pi\alpha}$ (there are finitely many values if $\alpha \in \mathbb{Q}$, infinitely many otherwise). Nevertheless, we make z^α single-valued by insisting $-\pi < \theta \leq \pi$.

De Moivre's theorem

Theorem (De Moivre's theorem)**.**

$$\cos n\theta + i\sin n\theta = (\cos\theta + i\sin\theta)^n.$$

Proof. First prove for the $n \geq 0$ case by induction. The $n = 0$ case is true since it merely reads $1 = 1$. We then have

$$\begin{aligned}(\cos\theta + i\sin\theta)^{n+1} &= (\cos\theta + i\sin\theta)^n(\cos\theta + i\sin\theta)\\ &= (\cos n\theta + i\sin n\theta)(\cos\theta + i\sin\theta)\\ &= \cos(n+1)\theta + i\sin(n+1)\theta\end{aligned}$$

If $n < 0$, let $m = -n$. Then $m > 0$ and

$$\begin{aligned}(cos\theta + i\sin\theta)^{-m} &= (\cos m\theta + i\sin m\theta)^{-1}\\ &= \frac{\cos m\theta - i\sin m\theta}{(\cos m\theta + i\sin m\theta)(\cos m\theta - i\sin m\theta)}\\ &= \frac{\cos(-m\theta) + i\sin(-m\theta)}{\cos^2 m\theta + \sin^2 m\theta}\\ &= \cos(-m\theta) + i\sin(-m\theta)\\ &= \cos n\theta + i\sin n\theta\end{aligned}$$ □

Note that "$\cos n\theta + i\sin n\theta = e^{in\theta} = (e^{i\theta})^n = (\cos\theta + i\sin\theta)^n$" is ***not*** a valid proof of De Moivre's theorem, since we do not know yet that $e^{in\theta} = (e^{i\theta})^n$. In fact, De Moivre's theorem tells us that this is a valid rule to apply.

Example. We have $\cos 5\theta + i\sin 5\theta = (\cos\theta + i\sin\theta)^5$. By binomial expansion of the RHS and taking real and imaginary parts, we have

$$\begin{aligned}\cos 5\theta &= 5\cos\theta - 20\cos^3\theta + 16\cos^5\theta\\ \sin 5\theta &= 5\sin\theta - 20\sin^3\theta + 16\sin^5\theta\end{aligned}$$

Lines and circles in $\mathbb{C}$

Since complex numbers can be regarded as points on the 2D plane, we can often use complex numbers to represent two dimensional objects.

Suppose that we want to represent a straight line through $z_0 \in \mathbb{C}$ parallel to $w \in \mathbb{C}$. The obvious way to do so is to let $z = z_0 + \lambda w$ where λ can take any real value. However, this is not an optimal way of doing so, since we are not using the power of complex numbers fully. This is just the same as the vector equation for straight lines, which you may or may not know from your A levels.

Instead, we arrange the equation to give $\lambda = \frac{z - z_0}{w}$. We take the complex conjugate of this expression to obtain $\bar{\lambda} = \frac{\bar{z} - \bar{z}_0}{\bar{w}}$. The trick here is to realize that λ is a real number. So we must have $\lambda = \bar{\lambda}$. This means that we must have

$$\begin{aligned}\frac{z - z_0}{w} &= \frac{\bar{z} - \bar{z}_0}{\bar{w}}\\ z\bar{w} - \bar{z}w &= z_0\bar{w} - \bar{z}_0 w.\end{aligned}$$

Theorem (Equation of straight line)**.** The equation of a straight line through z_0 and parallel to w is given by

$$z\bar{w} - \bar{z}w = z_0\bar{w} - \bar{z}_0 w.$$

The equation of a circle, on the other hand, is rather straightforward. Suppose that we want a circle with center $c \in \mathbb{C}$ and radius $\rho \in \mathbb{R}^+$. By definition of a circle, a point z is on the circle iff its distance to c is ρ, i.e. $|z - c| = \rho$. Recalling that $|z|^2 = z\bar{z}$, we obtain,

$$\begin{aligned} |z - c| &= \rho \\ |z - c|^2 &= \rho^2 \\ (z - c)(\bar{z} - \bar{c}) &= \rho^2 \\ z\bar{z} - \bar{c}z - c\bar{z} &= \rho^2 - c\bar{c} \end{aligned}$$

Theorem. The general equation of a circle with center $c \in \mathbb{C}$ and radius $\rho \in \mathbb{R}^+$ can be given by

$$z\bar{z} - \bar{c}z - c\bar{z} = \rho^2 - c\bar{c}.$$

1.3 Vectors

We might have first learned vectors as arrays of numbers, and then defined addition and multiplication in terms of the individual numbers in the vector. This however, is not what we are going to do here. The array of numbers is just a ***representation*** of the vector, instead of the vector itself.

Here, we will define vectors in terms of what they are, and then the various operations are defined axiomatically according to their properties.

Definition and basic properties

Definition (Vector)**.** A ***vector space*** over $\mathbb{R}$ or $\mathbb{C}$ is a collection of vectors $\mathbf{v} \in V$, together with two operations: addition of two vectors and multiplication of a vector with a scalar (i.e. a number from $\mathbb{R}$ or $\mathbb{C}$, respectively).

Vector addition has to satisfy the following axioms:

(i) $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$ (commutativity)

(ii) $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$ (associativity)

(iii) There is a vector $\mathbf{0}$ such that $\mathbf{a} + \mathbf{0} = \mathbf{a}$. (identity)

(iv) For all vectors $\mathbf{a}$, there is a vector $(-\mathbf{a})$ such that $\mathbf{a} + (-\mathbf{a}) = \mathbf{0}$ (inverse)

Scalar multiplication has to satisfy the following axioms:

(i) $\lambda(\mathbf{a} + \mathbf{b}) = \lambda\mathbf{a} + \lambda\mathbf{b}$.

(ii) $(\lambda + \mu)\mathbf{a} = \lambda\mathbf{a} + \mu\mathbf{a}$.

(iii) $\lambda(\mu\mathbf{a}) = (\lambda\mu)\mathbf{a}$.

(iv) $1\mathbf{a} = \mathbf{a}$.

Often, vectors have a length and direction. The length is denoted by $|\mathbf{v}|$. In this case, we can think of a vector as an "arrow" in space. Note that $\lambda\mathbf{a}$ is either parallel $(\lambda \geq 0)$ to or anti-parallel $(\lambda \leq 0)$ to $\mathbf{a}$.

Definition (Unit vector)**.** A ***unit vector*** is a vector with length 1. We write a unit vector as $\hat{\mathbf{v}}$.

Example. $\mathbb{R}^n$ is a vector space with component-wise addition and scalar multiplication. Note that the vector space $\mathbb{R}$ is a line, but not all lines are vector spaces. For example, $x + y = 1$ is not a vector space since it does not contain $\mathbf{0}$.

Scalar product

In a vector space, we can define the ***scalar product*** of two vectors, which returns a scalar (i.e. a real or complex number). We will first look at the usual scalar product defined for $\mathbb{R}^n$, and then define the scalar product axiomatically.

Geometric picture ($\mathbb{R}^2$ and $\mathbb{R}^3$ only)

Definition (Scalar/dot product)**.** $\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos\theta$, where θ is the angle between $\mathbf{a}$ and $\mathbf{b}$. It satisfies the following properties:

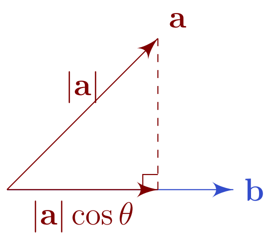
(i) $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$

(ii) $\mathbf{a} \cdot \mathbf{a} = |\mathbf{a}|^2 \geq 0$

(iii) $\mathbf{a} \cdot \mathbf{a} = 0$ iff $\mathbf{a} = \mathbf{0}$

(iv) If $\mathbf{a} \cdot \mathbf{b} = 0$ and $\mathbf{a}, \mathbf{b} \neq \mathbf{0}$, then $\mathbf{a}$ and $\mathbf{b}$ are perpendicular.

Intuitively, this is the product of the parts of $\mathbf{a}$ and $\mathbf{b}$ that are parallel.



Using the dot product, we can write the projection of $\mathbf{b}$ onto $\mathbf{a}$ as $(|\mathbf{b}| \cos\theta)\hat{\mathbf{a}} = (\hat{\mathbf{a}} \cdot \mathbf{b})\hat{\mathbf{a}}$.
The cosine rule can be derived as follows:

$$\begin{aligned}
|\overrightarrow{BC}|^2 &= |\overrightarrow{AC} - \overrightarrow{AB}|^2 \\
&= (\overrightarrow{AC} - \overrightarrow{AB}) \cdot (\overrightarrow{AC} - \overrightarrow{AB}) \\
&= |\overrightarrow{AB}|^2 + |\overrightarrow{AC}|^2 - 2|\overrightarrow{AB}||\overrightarrow{AC}| \cos\theta
\end{aligned}$$

We will later come up with a convenient algebraic way to evaluate this scalar product.

General algebraic definition

Definition (Inner/scalar product)**.** In a real vector space V, an ***inner product*** or ***scalar product*** is a map $V \times V \to \mathbb{R}$ that satisfies the following axioms. It is written as $\mathbf{x} \cdot \mathbf{y}$ or $\langle \mathbf{x} \mid \mathbf{y} \rangle$.

(i) $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$ (symmetry)

(ii) $\mathbf{x} \cdot (\lambda\mathbf{y} + \mu\mathbf{z}) = \lambda\mathbf{x} \cdot \mathbf{y} + \mu\mathbf{x} \cdot \mathbf{z}$ (linearity in 2nd argument)

(iii) $\mathbf{x} \cdot \mathbf{x} \geq 0$ with equality iff $\mathbf{x} = \mathbf{0}$ (positive definite)

Note that this is a definition only for ***real*** vector spaces, where the scalars are real. We will have a different set of definitions for complex vector spaces.

In particular, here we can use (i) and (ii) together to show linearity in 1st argument. However, this is generally not true for complex vector spaces.

Definition. The ***norm*** of a vector, written as $|\mathbf{a}|$ or $\|\mathbf{a}\|$, is defined as

$$|\mathbf{a}| = \sqrt{\mathbf{a}\cdot\mathbf{a}}.$$

Example. Instead of the usual $\mathbb{R}^n$ vector space, we can consider the set of all real (integrable) functions as a vector space. We can define the following inner product:

$$\langle f \mid g\rangle = \int_0^1 f(x)g(x)\, \mathrm{d}x.$$

Cauchy-Schwarz inequality

Theorem (Cauchy-Schwarz inequality)**.** For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$|\mathbf{x}\cdot\mathbf{y}| \le |\mathbf{x}||\mathbf{y}|.$$

Proof. Consider the expression $|\mathbf{x} - \lambda\mathbf{y}|^2$. We must have

$$\begin{aligned}
|\mathbf{x} - \lambda\mathbf{y}|^2 &\ge 0\\
(\mathbf{x} - \lambda\mathbf{y})\cdot(\mathbf{x} - \lambda\mathbf{y}) &\ge 0\\
\lambda^2|\mathbf{y}|^2 - \lambda(2\mathbf{x}\cdot\mathbf{y}) + |\mathbf{x}|^2 &\ge 0.
\end{aligned}$$

Viewing this as a quadratic in λ, we see that the quadratic is non-negative and thus cannot have 2 real roots. Thus the discriminant $\Delta \le 0$. So

$$\begin{aligned}
4(\mathbf{x}\cdot\mathbf{y})^2 &\le 4|\mathbf{y}|^2|\mathbf{x}|^2\\
(\mathbf{x}\cdot\mathbf{y})^2 &\le |\mathbf{x}|^2|\mathbf{y}|^2\\
|\mathbf{x}\cdot\mathbf{y}| &\le |\mathbf{x}||\mathbf{y}|.
\end{aligned}$$ □

Note that we proved this using the axioms of the scalar product. So this result holds for ***all*** possible scalar products on ***any*** (real) vector space.

Example. Let $\mathbf{x} = (\alpha, \beta, \gamma)$ and $\mathbf{y} = (1, 1, 1)$. Then by the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\alpha + \beta + \gamma &\le \sqrt{3}\sqrt{\alpha^2 + \beta^2 + \gamma^2}\\
\alpha^2 + \beta^2 + \gamma^2 &\ge \alpha\beta + \beta\gamma + \gamma\alpha,
\end{aligned}$$

with equality if $\alpha = \beta = \gamma$.

Corollary (Triangle inequality)**.**

$$|\mathbf{x} + \mathbf{y}| \le |\mathbf{x}| + |\mathbf{y}|.$$

Proof.

$$\begin{aligned}
|\mathbf{x} + \mathbf{y}|^2 &= (\mathbf{x} + \mathbf{y})\cdot(\mathbf{x} + \mathbf{y})\\
&= |\mathbf{x}|^2 + 2\mathbf{x}\cdot\mathbf{y} + |\mathbf{y}|^2\\
&\le |\mathbf{x}|^2 + 2|\mathbf{x}||\mathbf{y}| + |\mathbf{y}|^2\\
&= (|\mathbf{x}| + |\mathbf{y}|)^2.
\end{aligned}$$

So

$$|\mathbf{x} + \mathbf{y}| \le |\mathbf{x}| + |\mathbf{y}|.$$ □

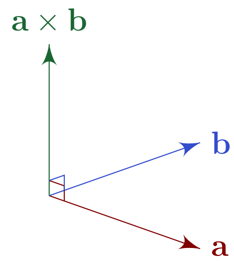
Vector product

Apart from the scalar product, we can also define the ***vector product***. However, this is defined only for $\mathbb{R}^3$ space, but not spaces in general.

Definition (Vector/cross product)**.** Consider $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$. Define the ***vector product***

$$\mathbf{a} \times \mathbf{b} = |\mathbf{a}||\mathbf{b}| \sin\theta \hat{\mathbf{n}},$$

where $\hat{\mathbf{n}}$ is a unit vector perpendicular to both $\mathbf{a}$ and $\mathbf{b}$. Since there are two (opposite) unit vectors that are perpendicular to both of them, we pick $\hat{\mathbf{n}}$ to be the one that is perpendicular to $\mathbf{a}, \mathbf{b}$ in a ***right-handed*** sense.



The vector product satisfies the following properties:

(i) $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$.

(ii) $\mathbf{a} \times \mathbf{a} = \mathbf{0}$.

(iii) $\mathbf{a} \times \mathbf{b} = \mathbf{0} \Rightarrow \mathbf{a} = \lambda\mathbf{b}$ for some $\lambda \in \mathbb{R}$ (or $\mathbf{b} = \mathbf{0}$).

(iv) $\mathbf{a} \times (\lambda\mathbf{b}) = \lambda(\mathbf{a} \times \mathbf{b})$.

(v) $\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}$.

If we have a triangle OAB, its area is given by $\frac{1}{2}|\overrightarrow{OA}||\overrightarrow{OB}| \sin\theta = \frac{1}{2}|\overrightarrow{OA} \times \overrightarrow{OB}|$. We define the vector area as $\frac{1}{2}\overrightarrow{OA} \times \overrightarrow{OB}$, which is often a helpful notion when we want to do calculus with surfaces.

There is a convenient way of calculating vector products:

Proposition.

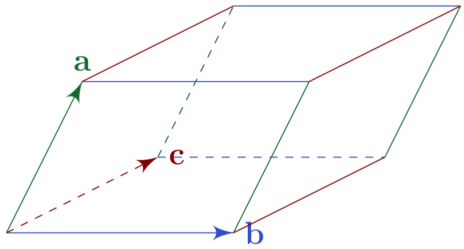
$$\begin{aligned}
\mathbf{a} \times \mathbf{b} &= (a_1\hat{\mathbf{i}} + a_2\hat{\mathbf{j}} + a_3\hat{\mathbf{k}}) \times (b_1\hat{\mathbf{i}} + b_2\hat{\mathbf{j}} + b_3\hat{\mathbf{k}}) \\
&= (a_2b_3 - a_3b_2)\hat{\mathbf{i}} + \cdots \\
&= \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}
\end{aligned}$$

Scalar triple product

Definition (Scalar triple product)**.** The ***scalar triple product*** is defined as

$$[\mathbf{a}, \mathbf{b}, \mathbf{c}] = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}).$$

Proposition. If a parallelepiped has sides represented by vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ that form a right-handed system, then the volume of the parallelepiped is given by $[\mathbf{a}, \mathbf{b}, \mathbf{c}]$.



Proof. The area of the base of the parallelepiped is given by $|\mathbf{b}||\mathbf{c}|\sin\theta = |\mathbf{b} \times \mathbf{c}|$. Thus the volume$= |\mathbf{b} \times \mathbf{c}||\mathbf{a}|\cos\phi = |\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})|$, where ϕ is the angle between $\mathbf{a}$ and the normal to $\mathbf{b}$ and $\mathbf{c}$. However, since $\mathbf{a}, \mathbf{b}, \mathbf{c}$ form a right-handed system, we have $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) \geq 0$. Therefore the volume is $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$. □

Since the order of $\mathbf{a}, \mathbf{b}, \mathbf{c}$ doesn't affect the volume, we know that

$$[\mathbf{a}, \mathbf{b}, \mathbf{c}] = [\mathbf{b}, \mathbf{c}, \mathbf{a}] = [\mathbf{c}, \mathbf{a}, \mathbf{b}] = -[\mathbf{b}, \mathbf{a}, \mathbf{c}] = -[\mathbf{a}, \mathbf{c}, \mathbf{b}] = -[\mathbf{c}, \mathbf{b}, \mathbf{a}].$$

Theorem. $\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}$.

Proof. Let $\mathbf{d} = \mathbf{a} \times (\mathbf{b} + \mathbf{c}) - \mathbf{a} \times \mathbf{b} - \mathbf{a} \times \mathbf{c}$. We have

$$\begin{aligned}
\mathbf{d} \cdot \mathbf{d} &= \mathbf{d} \cdot [\mathbf{a} \times (\mathbf{b} + \mathbf{c})] - \mathbf{d} \cdot (\mathbf{a} \times \mathbf{b}) - \mathbf{d} \cdot (\mathbf{a} \times \mathbf{c}) \\
&= (\mathbf{b} + \mathbf{c}) \cdot (\mathbf{d} \times \mathbf{a}) - \mathbf{b} \cdot (\mathbf{d} \times \mathbf{a}) - \mathbf{c} \cdot (\mathbf{d} \times \mathbf{a}) \\
&= 0
\end{aligned}$$

Thus $\mathbf{d} = \mathbf{0}$. □

Spanning sets and bases

2D space

Definition (Spanning set)**.** A set of vectors $\{\mathbf{a}, \mathbf{b}\}$ ***spans*** $\mathbb{R}^2$ if for all vectors $\mathbf{r} \in \mathbb{R}^2$, there exist some $\lambda, \mu \in \mathbb{R}$ such that $\mathbf{r} = \lambda\mathbf{a} + \mu\mathbf{b}$.

In $\mathbb{R}^2$, two vectors span the space if $\mathbf{a} \times \mathbf{b} \neq \mathbf{0}$.

Theorem. The coefficients λ, μ are unique.

Proof. Suppose that $\mathbf{r} = \lambda\mathbf{a} + \mu\mathbf{b} = \lambda'\mathbf{a} + \mu'\mathbf{b}$. Take the vector product with $\mathbf{a}$ on both sides to get $(\mu - \mu')\mathbf{a} \times \mathbf{b} = \mathbf{0}$. Since $\mathbf{a} \times \mathbf{b} \neq \mathbf{0}$, then $\mu = \mu'$. Similarly, $\lambda = \lambda'$. □

Definition (Linearly independent vectors in $\mathbb{R}^2$)**.** Two vectors $\mathbf{a}$ and $\mathbf{b}$ are ***linearly independent*** if for $\alpha, \beta \in \mathbb{R}$, $\alpha\mathbf{a} + \beta\mathbf{b} = \mathbf{0}$ iff $\alpha = \beta = 0$. In $\mathbb{R}^2$, $\mathbf{a}$ and $\mathbf{b}$ are linearly independent if $\mathbf{a} \times \mathbf{b} \neq \mathbf{0}$.

Definition (Basis of $\mathbb{R}^2$)**.** A set of vectors is a ***basis*** of $\mathbb{R}^2$ if it spans $\mathbb{R}^2$ and are linearly independent.

Example. $\{\hat{\mathbf{i}}, \hat{\mathbf{j}}\} = \{(1, 0), (0, 1)\}$ is a basis of $\mathbb{R}^2$. They are the standard basis of $\mathbb{R}^2$.

3D space

We can extend the above definitions of spanning set and linear independent set to $\mathbb{R}^3$. Here we have

Theorem. If $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$ are non-coplanar, i.e. $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) \neq 0$, then they form a basis of $\mathbb{R}^3$.

Proof. For any $\mathbf{r}$, write $\mathbf{r} = \lambda\mathbf{a} + \mu\mathbf{b} + \nu\mathbf{c}$. Performing the scalar product with $\mathbf{b} \times \mathbf{c}$ on both sides, one obtains $\mathbf{r} \cdot (\mathbf{b} \times \mathbf{c}) = \lambda\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) + \mu\mathbf{b} \cdot (\mathbf{b} \times \mathbf{c}) + \nu\mathbf{c} \cdot (\mathbf{b} \times \mathbf{c}) = \lambda[\mathbf{a}, \mathbf{b}, \mathbf{c}]$. Thus $\lambda = [\mathbf{r}, \mathbf{b}, \mathbf{c}]/[\mathbf{a}, \mathbf{b}, \mathbf{c}]$. The values of μ and ν can be found similarly. Thus each $\mathbf{r}$ can be written as a linear combination of $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$.

By the formula derived above, it follows that if $\alpha\mathbf{a} + \beta\mathbf{b} + \gamma\mathbf{c} = \mathbf{0}$, then $\alpha = \beta = \gamma = 0$. Thus they are linearly independent. □

Note that while we came up with formulas for λ, μ and ν, we did not actually prove that these coefficients indeed work. This is rather unsatisfactory. We could, of course, expand everything out and show that this indeed works, but in IB Linear Algebra, we will prove a much more general result, saying that if we have an n-dimensional space and a set of n linear independent vectors, then they form a basis.

In $\mathbb{R}^3$, the standard basis is $\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}$, or $(1, 0, 0), (0, 1, 0)$ and $(0, 0, 1)$.

$\mathbb{R}^n$ space

In general, we can define

Definition (Linearly independent vectors)**.** A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \cdots \mathbf{v}_m\}$ is ***linearly independent*** if

$$\sum_{i=1}^{m} \lambda_i \mathbf{v}_i = \mathbf{0} \Rightarrow (\forall i)\, \lambda_i = 0.$$

Definition (Spanning set)**.** A set of vectors $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3 \cdots \mathbf{u}_m\} \subseteq \mathbb{R}^n$ is a ***spanning set*** of $\mathbb{R}^n$ if

$$(\forall \mathbf{x} \in \mathbb{R}^n)(\exists \lambda_i) \sum_{i=1}^{m} \lambda_i \mathbf{u}_i = \mathbf{x}$$

Definition (Basis vectors)**.** A ***basis*** of $\mathbb{R}^n$ is a linearly independent spanning set. The standard basis of $\mathbb{R}^n$ is $\mathbf{e}_1 = (1, 0, 0, \cdots 0), \mathbf{e}_2 = (0, 1, 0, \cdots 0), \cdots \mathbf{e}_n = (0, 0, 0, \cdots, 1)$.

Definition (Orthonormal basis)**.** A basis $\{\mathbf{e}_i\}$ is ***orthonormal*** if $\mathbf{e}_i \cdot \mathbf{e}_j = 0$ if $i \neq j$ and $\mathbf{e}_i \cdot \mathbf{e}_i = 1$ for all i, j.

Using the Kronecker Delta symbol, which we will define later, we can write this condition as $\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$.

Definition (Dimension of vector space)**.** The ***dimension*** of a vector space is the number of vectors in its basis. (Exercise: show that this is well-defined)

We usually denote the components of a vector $\mathbf{x}$ by x_i. So we have $\mathbf{x} = (x_1, x_2, \cdots, x_n)$.

Definition (Scalar product)**.** The ***scalar product*** of $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is defined as $\mathbf{x} \cdot \mathbf{y} = \sum x_i y_i$.

The reader should check that this definition coincides with the $|\mathbf{x}||\mathbf{y}|\cos\theta$ definition in the case of $\mathbb{R}^2$ and $\mathbb{R}^3$.

$\mathbb{C}^n$ space

$\mathbb{C}^n$ is very similar to $\mathbb{R}^n$, except that we have complex numbers. As a result, we need a different definition of the scalar product. If we still defined $\mathbf{u} \cdot \mathbf{v} = \sum u_i v_i$, then if we let $\mathbf{u} = (0, i)$, then $\mathbf{u} \cdot \mathbf{u} = -1 < 0$. This would be bad if we want to use the scalar product to define a norm.

Definition ($\mathbb{C}^n$)**.** $\mathbb{C}^n = \{(z_1, z_2, \cdots, z_n) : z_i \in \mathbb{C}\}$. It has the same standard basis as $\mathbb{R}^n$ but the scalar product is defined differently. For $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$, $\mathbf{u} \cdot \mathbf{v} = \sum u_i^* v_i$. The scalar product has the following properties:

(i) $\mathbf{u} \cdot \mathbf{v} = (\mathbf{v} \cdot \mathbf{u})^*$

(ii) $\mathbf{u} \cdot (\lambda \mathbf{v} + \mu \mathbf{w}) = \lambda(\mathbf{u} \cdot \mathbf{v}) + \mu(\mathbf{u} \cdot \mathbf{w})$

(iii) $\mathbf{u} \cdot \mathbf{u} \geq 0$ and $\mathbf{u} \cdot \mathbf{u} = 0$ iff $\mathbf{u} = \mathbf{0}$

Instead of linearity in the first argument, here we have $(\lambda \mathbf{u} + \mu \mathbf{v}) \cdot \mathbf{w} = \lambda^* \mathbf{u} \cdot \mathbf{w} + \mu^* \mathbf{v} \cdot \mathbf{w}$.

Example.

$$\begin{aligned}
&\sum_{k=1}^{4} (-i)^k |\mathbf{x} + i^k \mathbf{y}|^2 \\
&= \sum (-i)^k \langle \mathbf{x} + i^k \mathbf{y} \mid \mathbf{x} + i^k \mathbf{y} \rangle \\
&= \sum (-i)^k (\langle \mathbf{x} + i^k \mathbf{y} \mid \mathbf{x} \rangle + i^k \langle \mathbf{x} + i^k \mathbf{y} \mid \mathbf{y} \rangle) \\
&= \sum (-i)^k (\langle \mathbf{x} \mid \mathbf{x} \rangle + (-i)^k \langle \mathbf{y} \mid \mathbf{x} \rangle + i^k \langle \mathbf{x} \mid \mathbf{y} \rangle + i^k (-i)^k \langle \mathbf{y} \mid \mathbf{y} \rangle) \\
&= \sum (-i)^k [(|\mathbf{x}|^2 + |\mathbf{y}|^2) + (-1)^k \langle \mathbf{y} \mid \mathbf{x} \rangle + \langle \mathbf{x} \mid \mathbf{y} \rangle] \\
&= (|\mathbf{x}|^2 + |\mathbf{y}|^2) \sum (-i)^k + \langle \mathbf{y} \mid \mathbf{x} \rangle \sum (-1)^k + \langle \mathbf{x} \mid \mathbf{y} \rangle \sum 1 \\
&= 4 \langle \mathbf{x} \mid \mathbf{y} \rangle.
\end{aligned}$$

We can prove the Cauchy-Schwarz inequality for complex vector spaces using the same proof as the real case, except that this time we have to first multiply $\mathbf{y}$ by some $e^{i\theta}$ so that $\mathbf{x}\cdot(e^{i\theta}\mathbf{y})$ is a real number. The factor of $e^{i\theta}$ will drop off at the end when we take the modulus signs.

Vector subspaces

Definition (Vector subspace)**.** A ***vector subspace*** of a vector space V is a subset of V that is also a vector space under the same operations. Both V and $\{\mathbf{0}\}$ are subspaces of V. All others are proper subspaces.

A useful criterion is that a subset $U \subseteq V$ is a subspace iff

(i) $\mathbf{x}, \mathbf{y} \in U \Rightarrow (\mathbf{x}+\mathbf{y}) \in U$.

(ii) $\mathbf{x} \in U \Rightarrow \lambda\mathbf{x} \in U$ for all scalars λ.

(iii) $\mathbf{0} \in U$.

This can be more concisely written as "U is non-empty and for all $\mathbf{x}, \mathbf{y} \in U$, $(\lambda\mathbf{x}+\mu\mathbf{y}) \in U$".

Example.

(i) If $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ is a basis of $\mathbb{R}^3$, then $\{\mathbf{a}+\mathbf{c}, \mathbf{b}+\mathbf{c}\}$ is a basis of a 2D subspace.

Suppose $\mathbf{x}, \mathbf{y} \in \operatorname{span}\{\mathbf{a}+\mathbf{c}, \mathbf{b}+\mathbf{c}\}$. Let

$$\begin{aligned}\mathbf{x} &= \alpha_1(\mathbf{a}+\mathbf{c}) + \beta_1(\mathbf{b}+\mathbf{c});\\ \mathbf{y} &= \alpha_2(\mathbf{a}+\mathbf{c}) + \beta_2(\mathbf{b}+\mathbf{c}).\end{aligned}$$

Then

$$\lambda\mathbf{x} + \mu\mathbf{y} = (\lambda\alpha_1 + \mu\alpha_2)(\mathbf{a}+\mathbf{c}) + (\lambda\beta_1 + \mu\beta_2)(\mathbf{b}+\mathbf{c}) \in \operatorname{span}\{\mathbf{a}+\mathbf{c}, \mathbf{b}+\mathbf{c}\}.$$

Thus this is a subspace of $\mathbb{R}^3$.

Now check that $\mathbf{a}+\mathbf{c}, \mathbf{b}+\mathbf{c}$ is a basis. We only need to check linear independence. If $\alpha(\mathbf{a}+\mathbf{c})+\beta(\mathbf{b}+\mathbf{c}) = \mathbf{0}$, then $\alpha\mathbf{a}+\beta\mathbf{b}+(\alpha+\beta)\mathbf{c} = \mathbf{0}$. Since $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ is a basis of $\mathbb{R}^3$, therefore $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are linearly independent and $\alpha = \beta = 0$. Therefore $\mathbf{a}+\mathbf{c}, \mathbf{b}+\mathbf{c}$ is a basis and the subspace has dimension 2.

(ii) Given a set of numbers α_i, let $U = \{\mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n \alpha_i x_i = 0\}$. We show that this is a vector subspace of $\mathbb{R}^n$: Take $\mathbf{x}, \mathbf{y} \in U$, then consider $\lambda\mathbf{x}+\mu\mathbf{y}$. We have $\sum \alpha_i(\lambda x_i + \mu y_i) = \lambda\sum\alpha_i x_i + \mu\sum\alpha_i y_i = 0$. Thus $\lambda\mathbf{x}+\mu\mathbf{y} \in U$.

The dimension of the subspace is $n-1$ as we can freely choose x_i for $i = 1, \cdots, n-1$ and then x_n is uniquely determined by the previous x_i's.

(iii) Let $W = \{\mathbf{x} \in \mathbb{R}^n : \sum\alpha_i x_i = 1\}$. Then $\sum\alpha_i(\lambda x_i + \mu y_i) = \lambda + \mu \neq 1$. Therefore W is not a vector subspace.

Suffix notation

Here we are going to introduce a powerful notation that can help us simplify a lot of things.

First of all, let $\mathbf{v} \in \mathbb{R}^3$. We can write $\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3 = (v_1, v_2, v_3)$. So in general, the ith component of $\mathbf{v}$ is written as v_i. We can thus write vector equations in component form. For example, $\mathbf{a} = \mathbf{b} \to a_i = b_i$ or $\mathbf{c} = \alpha\mathbf{a} + \beta\mathbf{b} \to c_i = \alpha a_i + \beta b_i$. A vector has one ***free*** suffix, i, while a scalar has none.

Notation (Einstein's summation convention)**.** Consider a sum $\mathbf{x} \cdot \mathbf{y} = \sum x_i y_i$. The ***summation convention*** says that we can drop the $\sum$ symbol and simply write $\mathbf{x} \cdot \mathbf{y} = x_i y_i$. If suffixes are repeated once, summation is understood.

Note that i is a dummy suffix and doesn't matter what it's called, i.e. $x_i y_i = x_j y_j = x_k y_k$ etc.

The rules of this convention are:

(i) Suffix appears once in a term: free suffix

(ii) Suffix appears twice in a term: dummy suffix and is summed over

(iii) Suffix appears three times or more: WRONG!

Example. $[(\mathbf{a} \cdot \mathbf{b})\mathbf{c} - (\mathbf{a} \cdot \mathbf{c})\mathbf{b}]_i = a_j b_j c_i - a_j c_j b_i$ summing over j understood.

It is possible for an item to have more than one index. These objects are known as ***tensors***, which will be studied in depth in the IA Vector Calculus course.

Here we will define two important tensors:

Definition (Kronecker delta)**.**

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

We have

$$\begin{pmatrix} \delta_{11} & \delta_{12} & \delta_{13} \\ \delta_{21} & \delta_{22} & \delta_{23} \\ \delta_{31} & \delta_{32} & \delta_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \mathbf{I}.$$

So the Kronecker delta represents an identity matrix.

Example.

(i) $a_i \delta_{i1} = a_1$. In general, $a_i \delta_{ij} = a_j$ (i is dummy, j is free).

(ii) $\delta_{ij}\delta_{jk} = \delta_{ik}$

(iii) $\delta_{ii} = n$ if we are in $\mathbb{R}^n$.

(iv) $a_p \delta_{pq} b_q = a_p b_p$ with p, q both dummy suffices and summed over.

Definition (Alternating symbol ε_{ijk})**.** Consider rearrangements of $1, 2, 3$. We can divide them into even and odd permutations. Even permutations include $(1, 2, 3)$, $(2, 3, 1)$ and $(3, 1, 2)$. These are permutations obtained by performing two (or no) swaps of the elements of $(1, 2, 3)$. (Alternatively, it is any "rotation" of $(1, 2, 3)$)

The odd permutations are $(2, 1, 3)$, $(1, 3, 2)$ and $(3, 2, 1)$. They are the permutations obtained by one swap only.

Define

$$\varepsilon_{ijk} = \begin{cases} +1 & ijk \text{ is even permutation} \\ -1 & ijk \text{ is odd permutation} \\ 0 & \text{otherwise (i.e. repeated suffices)} \end{cases}$$

ε_{ijk} has 3 free suffices.

We have $\varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} = +1$ and $\varepsilon_{213} = \varepsilon_{132} = \varepsilon_{321} = -1$. $\varepsilon_{112} = \varepsilon_{111} = \cdots = 0$.

We have

(i) $\varepsilon_{ijk}\delta_{jk} = \varepsilon_{ijj} = 0$

(ii) If $a_{jk} = a_{kj}$ (i.e. a_{ij} is symmetric), then $\varepsilon_{ijk}a_{jk} = \varepsilon_{ijk}a_{kj} = -\varepsilon_{ikj}a_{kj}$. Since $\varepsilon_{ijk}a_{jk} = \varepsilon_{ikj}a_{kj}$ (we simply renamed dummy suffices), we have $\varepsilon_{ijk}a_{jk} = 0$.

Proposition. $(\mathbf{a} \times \mathbf{b})_i = \varepsilon_{ijk}a_j b_k$

Proof. By expansion of formula □

Theorem. $\varepsilon_{ijk}\varepsilon_{ipq} = \delta_{jp}\delta_{kq} - \delta_{jq}\delta_{kp}$

Proof. Proof by exhaustion:

$$\text{RHS} = \begin{cases} +1 & \text{if } j = p \text{ and } k = q \\ -1 & \text{if } j = q \text{ and } k = p \\ 0 & \text{otherwise} \end{cases}$$

LHS: Summing over i, the only non-zero terms are when $j, k \neq i$ and $p, q \neq i$. If $j = p$ and $k = q$, LHS is $(-1)^2$ or $(+1)^2 = 1$. If $j = q$ and $k = p$, LHS is $(+1)(-1)$ or $(-1)(+1) = -1$. All other possibilities result in 0. □

Equally, we have $\varepsilon_{ijk}\varepsilon_{pqk} = \delta_{ip}\delta_{jq} - \delta_{jp}\delta_{iq}$ and $\varepsilon_{ijk}\varepsilon_{pjq} = \delta_{ip}\delta_{kq} - \delta_{iq}\delta_{kp}$.

Proposition.

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a})$$

Proof. In suffix notation, we have

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = a_i(\mathbf{b} \times \mathbf{c})_i = \varepsilon_{ijk}b_j c_k a_i = \varepsilon_{jki}b_j c_k a_i = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}).$$

□

Theorem (Vector triple product)**.**

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}.$$

Proof.

$$\begin{aligned}
[\mathbf{a}\times(\mathbf{b}\times\mathbf{c})]_i &= \varepsilon_{ijk}a_j(b\times c)_k\\
&= \varepsilon_{ijk}\varepsilon_{kpq}a_jb_pc_q\\
&= \varepsilon_{ijk}\varepsilon_{pqk}a_jb_pc_q\\
&= (\delta_{ip}\delta_{jq}-\delta_{iq}\delta_{jp})a_jb_pc_q\\
&= a_jb_ic_j - a_jc_ib_j\\
&= (\mathbf{a}\cdot\mathbf{c})b_i - (\mathbf{a}\cdot\mathbf{b})c_i
\end{aligned}$$ □

Similarly, $(\mathbf{a}\times\mathbf{b})\times\mathbf{c} = (\mathbf{a}\cdot\mathbf{c})\mathbf{b} - (\mathbf{b}\cdot\mathbf{c})\mathbf{a}$.

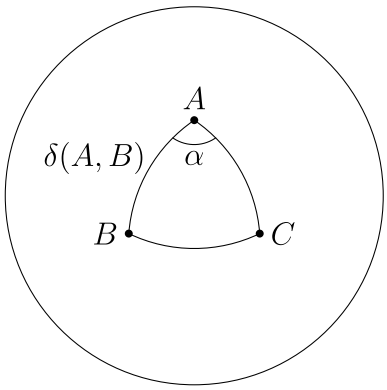
Spherical trigonometry

Proposition. $(\mathbf{a}\times\mathbf{b})\cdot(\mathbf{a}\times\mathbf{c}) = (\mathbf{a}\cdot\mathbf{a})(\mathbf{b}\cdot\mathbf{c}) - (\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{c})$.

Proof.

$$\begin{aligned}
\text{LHS} &= (\mathbf{a}\times\mathbf{b})_i(\mathbf{a}\times\mathbf{c})_i\\
&= \varepsilon_{ijk}a_jb_k\varepsilon_{ipq}a_pc_q\\
&= (\delta_{jp}\delta_{kq}-\delta_{jq}\delta_{kp})a_jb_ka_pc_q\\
&= a_jb_ka_jc_k - a_jb_ka_kc_j\\
&= (\mathbf{a}\cdot\mathbf{a})(\mathbf{b}\cdot\mathbf{c}) - (\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{c})
\end{aligned}$$ □

Consider the unit sphere, center O, with $\mathbf{a},\mathbf{b},\mathbf{c}$ on the surface.



Suppose we are living on the surface of the sphere. So the distance from A to B is the arc length on the sphere. We can imagine this to be along the circumference of the circle through A and B with center O. So the distance is $\angle AOB$, which we shall denote by $\delta(A,B)$. So $\mathbf{a}\cdot\mathbf{b} = \cos\angle AOB = \cos\delta(A,B)$. We obtain similar expressions for other dot products. Similarly, we get $|\mathbf{a}\times\mathbf{b}| = \sin\delta(A,B)$.

$$\begin{aligned}
\cos\alpha &= \frac{(\mathbf{a}\times\mathbf{b})\cdot(\mathbf{a}\times\mathbf{c})}{|\mathbf{a}\times\mathbf{b}||\mathbf{a}\times\mathbf{c}|}\\
&= \frac{\mathbf{b}\cdot\mathbf{c}-(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{c})}{|\mathbf{a}\times\mathbf{b}||\mathbf{a}\times\mathbf{c}|}
\end{aligned}$$

Putting in our expressions for the dot and cross products, we obtain

$$\cos\alpha \sin\delta(A,B)\sin\delta(A,C) = \cos\delta(B,C) - \cos\delta(A,B)\cos\delta(A,C).$$

This is the spherical cosine rule that applies when we live on the surface of a sphere. What does this spherical geometry look like?
Consider a spherical equilateral triangle. Using the spherical cosine rule,

$$\cos\alpha = \frac{\cos\delta - \cos^2\delta}{\sin^2\delta} = 1 - \frac{1}{1+\cos\delta}.$$

Since $\cos\delta \leq 1$, we have $\cos\alpha \leq \frac{1}{2}$ and $\alpha \geq 60^\circ$. Equality holds iff $\delta = 0$, i.e. the triangle is simply a point. So on a sphere, each angle of an equilateral triangle is greater than 60°, and the angle sum of a triangle is greater than 180°.

Geometry

Lines

Any line through $\mathbf{a}$ and parallel to $\mathbf{t}$ can be written as

$$\mathbf{x} = \mathbf{a} + \lambda\mathbf{t}.$$

By crossing both sides of the equation with $\mathbf{t}$, we have

Theorem. The equation of a straight line through $\mathbf{a}$ and parallel to $\mathbf{t}$ is

$$(\mathbf{x} - \mathbf{a}) \times \mathbf{t} = \mathbf{0} \text{ or } \mathbf{x} \times \mathbf{t} = \mathbf{a} \times \mathbf{t}.$$

Plane

To define a plane Π, we need a normal $\mathbf{n}$ to the plane and a fixed point $\mathbf{b}$. For any $\mathbf{x} \in \Pi$, the vector $\mathbf{x} - \mathbf{b}$ is contained in the plane and is thus normal to $\mathbf{n}$, i.e. $(\mathbf{x} - \mathbf{b}) \cdot \mathbf{n} = 0$.

Theorem. The equation of a plane through $\mathbf{b}$ with normal $\mathbf{n}$ is given by

$$\mathbf{x} \cdot \mathbf{n} = \mathbf{b} \cdot \mathbf{n}.$$

If $\mathbf{n} = \hat{\mathbf{n}}$ is a unit normal, then $d = \mathbf{x} \cdot \hat{\mathbf{n}} = \mathbf{b} \cdot \hat{\mathbf{n}}$ is the perpendicular distance from the origin to Π.
Alternatively, if $\mathbf{a}, \mathbf{b}, \mathbf{c}$ lie in the plane, then the equation of the plane is

$$(\mathbf{x} - \mathbf{a}) \cdot [(\mathbf{b} - \mathbf{a}) \times (\mathbf{c} - \mathbf{a})] = 0.$$

Example.

(i) Consider the intersection between a line $\mathbf{x} \times \mathbf{t} = \mathbf{a} \times \mathbf{t}$ with the plane $\mathbf{x} \cdot \mathbf{n} = \mathbf{b} \cdot \mathbf{n}$. Cross $\mathbf{n}$ on the right with the line equation to obtain

$$(\mathbf{x} \cdot \mathbf{n})\mathbf{t} - (\mathbf{t} \cdot \mathbf{n})\mathbf{x} = (\mathbf{a} \times \mathbf{t}) \times \mathbf{n}$$

Eliminate $\mathbf{x} \cdot \mathbf{n}$ using $\mathbf{x} \cdot \mathbf{n} = \mathbf{b} \cdot \mathbf{n}$

$$(\mathbf{t} \cdot \mathbf{n})\mathbf{x} = (\mathbf{b} \cdot \mathbf{n})\mathbf{t} - (\mathbf{a} \times \mathbf{t}) \times \mathbf{n}$$

Provided $\mathbf{t} \cdot \mathbf{n}$ is non-zero, the point of intersection is

$$\mathbf{x} = \frac{(\mathbf{b} \cdot \mathbf{n})\mathbf{t} - (\mathbf{a} \times \mathbf{t}) \times \mathbf{n}}{\mathbf{t} \cdot \mathbf{n}}.$$

Exercise: what if $\mathbf{t} \cdot \mathbf{n} = 0$?

(ii) Shortest distance between two lines. Let L_1 be $(\mathbf{x} - \mathbf{a}_1) \times \mathbf{t}_1 = \mathbf{0}$ and L_2 be $(\mathbf{x} - \mathbf{a}_2) \times \mathbf{t}_2 = \mathbf{0}$.

The distance of closest approach s is along a line perpendicular to both L_1 and L_2, i.e. the line of closest approach is perpendicular to both lines and thus parallel to $\mathbf{t}_1 \times \mathbf{t}_2$. The distance s can then be found by projecting $\mathbf{a}_1 - \mathbf{a}_2$ onto $\mathbf{t}_1 \times \mathbf{t}_2$. Thus $s = \left|(\mathbf{a}_1 - \mathbf{a}_2) \cdot \frac{\mathbf{t}_1 \times \mathbf{t}_2}{|\mathbf{t}_1 \times \mathbf{t}_2|}\right|$.

Vector equations

Example. $\mathbf{x} - (\mathbf{x} \times \mathbf{a}) \times \mathbf{b} = \mathbf{c}$. Strategy: take the dot or cross of the equation with suitable vectors. The equation can be expanded to form

$$\mathbf{x} - (\mathbf{x} \cdot \mathbf{b})\mathbf{a} + (\mathbf{a} \cdot \mathbf{b})\mathbf{x} = \mathbf{c}.$$

Dot this with $\mathbf{b}$ to obtain

$$\begin{aligned} \mathbf{x} \cdot \mathbf{b} - (\mathbf{x} \cdot \mathbf{b})(\mathbf{a} \cdot \mathbf{b}) + (\mathbf{a} \cdot \mathbf{b})(\mathbf{x} \cdot \mathbf{b}) &= \mathbf{c} \cdot \mathbf{b} \\ \mathbf{x} \cdot \mathbf{b} &= \mathbf{c} \cdot \mathbf{b}. \end{aligned}$$

Substituting this into the original equation, we have

$$\mathbf{x}(1 + \mathbf{a} \cdot \mathbf{b}) = \mathbf{c} + (\mathbf{c} \cdot \mathbf{b})\mathbf{a}$$

If $(1 + \mathbf{a} \cdot \mathbf{b})$ is non-zero, then

$$\mathbf{x} = \frac{\mathbf{c} + (\mathbf{c} \cdot \mathbf{b})\mathbf{a}}{1 + \mathbf{a} \cdot \mathbf{b}}$$

Otherwise, when $(1 + \mathbf{a} \cdot \mathbf{b}) = 0$, if $\mathbf{c} + (\mathbf{c} \cdot \mathbf{b})\mathbf{a} \neq \mathbf{0}$, then a contradiction is reached. Otherwise, $\mathbf{x} \cdot \mathbf{b} = \mathbf{c} \cdot \mathbf{b}$ is the most general solution, which is a plane of solutions.

1.4 Linear maps

A ***linear map*** is a special type of function between vector spaces. In fact, most of the time, these are the only functions we actually care about. They are maps that satisfy the property $f(\lambda\mathbf{a} + \mu\mathbf{b}) = \lambda f(\mathbf{a}) + \mu f(\mathbf{b})$.

We will first look at two important examples of linear maps — rotations and reflections, and then study their properties formally.

Examples

Rotation in $\mathbb{R}^3$

In $\mathbb{R}^3$, first consider the simple cases where we rotate about the z axis by θ. We call this rotation R and write $\mathbf{x}' = R(\mathbf{x})$.

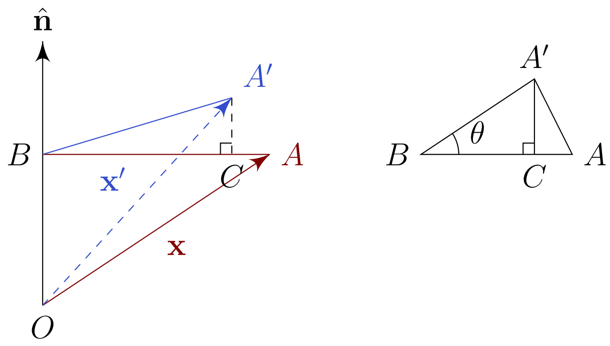
Suppose that initially, $\mathbf{x} = (x, y, z) = (r\cos\phi, r\sin\phi, z)$. Then after a rotation by θ, we get

$$\begin{aligned}\mathbf{x}' &= (r\cos(\phi+\theta), r\sin(\phi+\theta), z)\\ &= (r\cos\phi\cos\theta - r\sin\phi\sin\theta, r\sin\phi\cos\theta + r\cos\phi\sin\theta, z)\\ &= (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta, z).\end{aligned}$$

We can represent this by a matrix R such that $x_i' = R_{ij}x_j$. Using our formula above, we obtain

$$R = \begin{pmatrix}\cos\theta & -\sin\theta & 0\\ \sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{pmatrix}$$

Now consider the general case where we rotate by θ about $\hat{\mathbf{n}}$.



We have $\mathbf{x}' = \overrightarrow{OB} + \overrightarrow{BC} + \overrightarrow{CA'}$. We know that

$$\begin{aligned}\overrightarrow{OB} &= (\hat{\mathbf{n}}\cdot\mathbf{x})\hat{\mathbf{n}}\\ \overrightarrow{BC} &= \overrightarrow{BA}\cos\theta\\ &= (\overrightarrow{BO} + \overrightarrow{OA})\cos\theta\\ &= (-(\hat{\mathbf{n}}\cdot\mathbf{x})\hat{\mathbf{n}} + \mathbf{x})\cos\theta\end{aligned}$$

Finally, to get $\overrightarrow{CA}$, we know that $|\overrightarrow{CA'}| = |\overrightarrow{BA'}|\sin\theta = |\overrightarrow{BA}|\sin\theta = |\hat{\mathbf{n}}\times\mathbf{x}|\sin\theta$. Also, $\overrightarrow{CA'}$ is parallel to $\hat{\mathbf{n}}\times\mathbf{x}$. So we must have $\overrightarrow{CA'} = (\hat{\mathbf{n}}\times\mathbf{x})\sin\theta$.
Thus $\mathbf{x}' = \mathbf{x}\cos\theta + (1-\cos\theta)(\hat{\mathbf{n}}\cdot\mathbf{x})\hat{\mathbf{n}} + \hat{\mathbf{n}}\times\mathbf{x}\sin\theta$. In components,

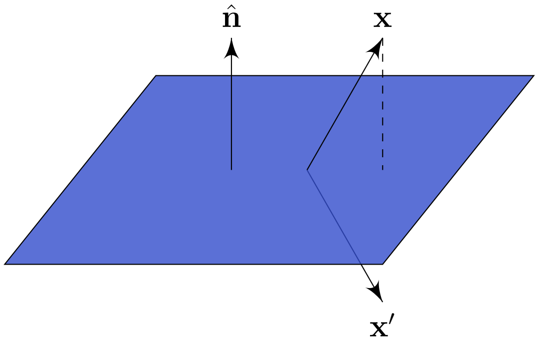
$$x_i' = x_i\cos\theta + (1-\cos\theta)n_jx_jn_i - \varepsilon_{ijk}x_jn_k\sin\theta.$$

We want to find an R such that $x_i' = R_{ij}x_j$. So

$$R_{ij} = \delta_{ij}\cos\theta + (1-\cos\theta)n_in_j - \varepsilon_{ijk}n_k\sin\theta.$$

Reflection in $\mathbb{R}^3$

Suppose we want to reflect through a plane through O with normal $\hat{\mathbf{n}}$. First of all the projection of $\mathbf{x}$ onto $\hat{\mathbf{n}}$ is given by $(\mathbf{x}\cdot\hat{\mathbf{n}})\hat{\mathbf{n}}$. So we get $\mathbf{x}' = \mathbf{x} - 2(\mathbf{x}\cdot\hat{\mathbf{n}})\hat{\mathbf{n}}$. In suffix notation, we have $x_i' = x_i - 2x_jn_jn_i$. So our reflection matrix is $R_{ij} = \delta_{ij} - 2n_in_j$.



Linear Maps

Definition (Domain, codomain and image of map)**.** Consider sets A and B and mapping $T : A \to B$ such that each $x \in A$ is mapped into a unique $x' = T(x) \in B$. A is the ***domain*** of T and B is the ***co-domain*** of T. Typically, we have $T : \mathbb{R}^n \to \mathbb{R}^m$ or $T : \mathbb{C}^n \to \mathbb{C}^m$.

Definition (Linear map)**.** Let V, W be real (or complex) vector spaces, and $T : V \to W$. Then T is a ***linear map*** if

(i) $T(\mathbf{a}+\mathbf{b}) = T(\mathbf{a}) + T(\mathbf{b})$ for all $\mathbf{a}, \mathbf{b} \in V$.

(ii) $T(\lambda\mathbf{a}) = \lambda T(\mathbf{a})$ for all $\lambda \in \mathbb{R}$ (or $\mathbb{C}$).

Equivalently, we have $T(\lambda\mathbf{a} + \mu\mathbf{b}) = \lambda T(\mathbf{a}) + \mu T(\mathbf{b})$.

Example.

(i) Consider a translation $T : \mathbb{R}^3 \to \mathbb{R}^3$ with $T(\mathbf{x}) = \mathbf{x} + \mathbf{a}$ for some fixed, given $\mathbf{a}$. This is ***not*** a linear map since $T(\lambda\mathbf{x} + \mu\mathbf{y}) \neq \lambda\mathbf{x} + \mu\mathbf{y} + (\lambda+\mu)\mathbf{a}$.

(ii) Rotation, reflection and projection are linear transformations.

Definition (Image and kernel of map)**.** The ***image*** of a map $f : U \to V$ is the subset of V $\{f(\mathbf{u}) : \mathbf{u} \in U\}$. The ***kernel*** is the subset of U $\{\mathbf{u} \in U : f(\mathbf{u}) = \mathbf{0}\}$.

Example.

(i) Consider $S : \mathbb{R}^3 \to \mathbb{R}^2$ with $S(x, y, z) = (x + y, 2x - z)$. Simple yet tedious algebra shows that this is linear. Now consider the effect of S on the standard basis. $S(1, 0, 0) = (1, 2)$, $S(0, 1, 0) = (1, 0)$ and $S(0, 0, 1) = (0, -1)$. Clearly these are linearly dependent, but they do span the whole of $\mathbb{R}^2$. We can say $S(\mathbb{R}^3) = \mathbb{R}^2$. So the image is $\mathbb{R}^2$.

Now solve $S(x, y, z) = \mathbf{0}$. We need $x + y = 0$ and $2x - z = 0$. Thus $\mathbf{x} = (x, -x, 2x)$, i.e. it is parallel to $(1, -1, 2)$. So the set $\{\lambda(1, -1, 2) : \lambda \in \mathbb{R}\}$ is the kernel of S.

(ii) Consider a rotation in $\mathbb{R}^3$. The kernel is the zero vector and the image is $\mathbb{R}^3$.

(iii) Consider a projection of $\mathbf{x}$ onto a plane with normal $\hat{\mathbf{n}}$. The image is the plane itself, and the kernel is any vector parallel to $\hat{\mathbf{n}}$

Theorem. Consider a linear map $f : U \to V$, where U, V are vector spaces. Then $\mathrm{im}(f)$ is a subspace of V, and $\ker(f)$ is a subspace of U.

Proof. Both are non-empty since $f(\mathbf{0}) = \mathbf{0}$.

If $\mathbf{x}, \mathbf{y} \in \mathrm{im}(f)$, then $\exists \mathbf{a}, \mathbf{b} \in U$ such that $\mathbf{x} = f(\mathbf{a}), \mathbf{y} = f(\mathbf{b})$. Then $\lambda\mathbf{x} + \mu\mathbf{y} = \lambda f(\mathbf{a}) + \mu f(\mathbf{b}) = f(\lambda\mathbf{a} + \mu\mathbf{b})$. Now $\lambda\mathbf{a} + \mu\mathbf{b} \in U$ since U is a vector space, so there is an element in U that maps to $\lambda\mathbf{x} + \mu\mathbf{y}$. So $\lambda\mathbf{x} + \mu\mathbf{y} \in \mathrm{im}(f)$ and $\mathrm{im}(f)$ is a subspace of V.

Suppose $\mathbf{x}, \mathbf{y} \in \ker(f)$, i.e. $f(\mathbf{x}) = f(\mathbf{y}) = \mathbf{0}$. Then $f(\lambda\mathbf{x} + \mu\mathbf{y}) = \lambda f(\mathbf{x}) + \mu f(\mathbf{y}) = \lambda\mathbf{0} + \mu\mathbf{0} = \mathbf{0}$. Therefore $\lambda\mathbf{x} + \mu\mathbf{y} \in \ker(f)$. □

Rank and nullity

Definition (Rank of linear map)**.** The ***rank*** of a linear map $f : U \to V$, denoted by $r(f)$, is the dimension of the image of f.

Definition (Nullity of linear map)**.** The ***nullity*** of f, denoted $n(f)$ is the dimension of the kernel of f.

Example. For the projection onto a plane in $\mathbb{R}^3$, the image is the whole plane and the rank is 2. The kernel is a line so the nullity is 1.

Theorem (Rank-nullity theorem)**.** For a linear map $f : U \to V$,

$$r(f) + n(f) = \dim(U).$$

Proof. (Non-examinable) Write $\dim(U) = n$ and $n(f) = m$. If $m = n$, then f is the zero map, and the proof is trivial, since $r(f) = 0$. Otherwise, assume $m < n$.

Suppose $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_m\}$ is a basis of $\ker f$, Extend this to a basis of the whole of U to get $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_m, \mathbf{e}_{m+1}, \cdots, \mathbf{e}_n\}$. To prove the theorem, we need to prove that $\{f(\mathbf{e}_{m+1}), f(\mathbf{e}_{m+2}), \cdots f(\mathbf{e}_n)\}$ is a basis of $\mathrm{im}(f)$.

(i) First show that it spans $\text{im}(f)$. Take $\mathbf{y} \in \text{im}(f)$. Thus $\exists \mathbf{x} \in U$ such that $\mathbf{y} = f(\mathbf{x})$. Then
$$\mathbf{y} = f(\alpha_1\mathbf{e}_1 + \alpha_2\mathbf{e}_2 + \cdots + \alpha_n\mathbf{e}_n),$$
since $\mathbf{e}_1, \cdots \mathbf{e}_n$ is a basis of U. Thus
$$\mathbf{y} = \alpha_1 f(\mathbf{e}_1) + \alpha_2 f(\mathbf{e}_2) + \cdots + \alpha_m f(\mathbf{e}_m) + \alpha_{m+1} f(\mathbf{e}_{m+1}) + \cdots + \alpha_n f(\mathbf{e}_n).$$
The first m terms map to $\mathbf{0}$, since $\mathbf{e_1}, \cdots \mathbf{e_m}$ is the basis of the kernel of f. Thus
$$\mathbf{y} = \alpha_{m+1} f(\mathbf{e}_{m+1}) + \cdots + \alpha_n f(\mathbf{e}_n).$$

(ii) To show that they are linearly independent, suppose
$$\alpha_{m+1} f(\mathbf{e}_{m+1}) + \cdots + \alpha_n f(\mathbf{e}_n) = \mathbf{0}.$$
Then
$$f(\alpha_{m+1}\mathbf{e}_{m+1} + \cdots + \alpha_n\mathbf{e}_n) = \mathbf{0}.$$
Thus $\alpha_{m+1}\mathbf{e}_{m+1} + \cdots + \alpha_n\mathbf{e}_n \in \ker(f)$. Since $\{\mathbf{e}_1, \cdots, \mathbf{e}_m\}$ span $\ker(f)$, there exist some $\alpha_1, \alpha_2, \cdots \alpha_m$ such that
$$\alpha_{m+1}\mathbf{e}_{m+1} + \cdots + \alpha_n\mathbf{e}_n = \alpha_1\mathbf{e_1} + \cdots + \alpha_m\mathbf{e}_m.$$
But $\mathbf{e}_1 \cdots \mathbf{e}_n$ is a basis of U and are linearly independent. So $\alpha_i = 0$ for all i. Then the only solution to the equation $\alpha_{m+1} f(\mathbf{e}_{m+1}) + \cdots + \alpha_n f(\mathbf{e}_n) = \mathbf{0}$ is $\alpha_i = 0$, and they are linearly independent by definition. $\square$

Example. Calculate the kernel and image of $f : \mathbb{R}^3 \to \mathbb{R}^3$, defined by $f(x, y, z) = (x + y + z, 2x - y + 5z, x + 2z)$.
First find the kernel: we've got the system of equations:
$$\begin{aligned} x + y + z &= 0 \\ 2x - y + 5z &= 0 \\ x + 2z &= 0 \end{aligned}$$
Note that the first and second equation add to give $3x + 6z = 0$, which is identical to the third. Then using the first and third equation, we have $y = -x - z = z$. So the kernel is any vector in the form $(-2z, z, z)$ and is the span of $(-2, 1, 1)$.
To find the image, extend the basis of $\ker(f)$ to a basis of the whole of $\mathbb{R}^3$: $\{(-2, 1, 1), (0, 1, 0), (0, 0, 1)\}$. Apply f to this basis to obtain $(0, 0, 0), (1, -1, 0)$ and $(1, 5, 2)$. From the proof of the rank-nullity theorem, we know that $f(0, 1, 0)$ and $f(0, 0, 1)$ is a basis of the image.
To get the standard form of the image, we know that the normal to the plane is parallel to $(1, -1, 0) \times (1, 5, 2) \parallel (1, 1, -3)$. Since $\mathbf{0} \in \text{im}(f)$, the equation of the plane is $x + y - 3z = 0$.

Matrices

In the examples above, we have represented our linear maps by some object R such that $x'_i = R_{ij}x_j$. We call R the ***matrix*** for the linear map. In general, let $\alpha : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map, and $\mathbf{x}' = \alpha(\mathbf{x})$.

Let $\{\mathbf{e}_i\}$ be a basis of $\mathbb{R}^n$. Then $\mathbf{x} = x_j\mathbf{e}_j$ for some x_j. Then we get

$$\mathbf{x}' = \alpha(x_j\mathbf{e}_j) = x_j\alpha(\mathbf{e}_j).$$

So we get that

$$x'_i = [\alpha(\mathbf{e}_j)]_i x_j.$$

We now define $A_{ij} = [\alpha(\mathbf{e}_j)]_i$. Then $x'_i = A_{ij}x_j$. We write

$$A = \{A_{ij}\} = \begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & A_{ij} & \vdots \\ A_{m1} & \cdots & A_{mn} \end{pmatrix}$$

Here A_{ij} is the entry in the ith row of the jth column. We say that A is an $m \times n$ matrix, and write $\mathbf{x}' = A\mathbf{x}$.

We see that the columns of the matrix are the images of the standard basis vectors under the mapping α.

Example.

Examples

(i) In $\mathbb{R}^2$, consider a reflection in a line with an angle θ to the x axis. We know that $\hat{\mathbf{i}} \mapsto \cos 2\theta\hat{\mathbf{i}} + \sin 2\theta\hat{\mathbf{j}}$, with $\hat{\mathbf{j}} \mapsto -\cos 2\theta\hat{\mathbf{j}} + \sin 2\theta\hat{\mathbf{i}}$. Then the matrix is $\begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}$.

(ii) In $\mathbb{R}^3$, as we've previously seen, a rotation by θ about the z axis is given by

$$R = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

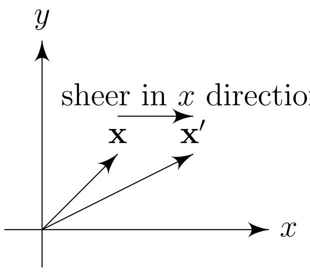
(iii) In $\mathbb{R}^3$, a reflection in plane with normal $\hat{\mathbf{n}}$ is given by $R_{ij} = \delta_{ij} - 2\hat{n}_i\hat{n}_j$. Written as a matrix, we have

$$\begin{pmatrix} 1-2\hat{n}_1^2 & -2\hat{n}_1\hat{n}_2 & -2\hat{n}_1\hat{n}_3 \\ -2\hat{n}_2\hat{n}_1 & 1-2\hat{n}_2^2 & -2\hat{n}_2\hat{n}_3 \\ -2\hat{n}_3\hat{n}_1 & -2\hat{n}_3\hat{n}_2 & 1-2\hat{n}_3^2 \end{pmatrix}$$

(iv) Dilation ("stretching") $\alpha : \mathbb{R}^3 \to \mathbb{R}^3$ is given by a map $(x, y, z) \mapsto (\lambda x, \mu y, \nu z)$ for some λ, μ, ν. The matrix is

$$\begin{pmatrix} \lambda & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 0 & \nu \end{pmatrix}$$

(v) Shear: Consider $S : \mathbb{R}^3 \to \mathbb{R}^3$ that sheers in the x direction:



We have $(x, y, z) \mapsto (x + \lambda y, y, z)$. Then

$$S = \begin{pmatrix} 1 & \lambda & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Matrix Algebra

This part is mostly on a whole lot of definitions, saying what we can do with matrices and classifying them into different types.

Definition (Addition of matrices)**.** Consider two linear maps $\alpha, \beta : \mathbb{R}^n \to \mathbb{R}^m$. The sum of α and β is defined by

$$(\alpha + \beta)(\mathbf{x}) = \alpha(\mathbf{x}) + \beta(\mathbf{x})$$

In terms of the matrix, we have

$$(A + B)_{ij} x_j = A_{ij} x_j + B_{ij} x_j,$$

or

$$(A + B)_{ij} = A_{ij} + B_{ij}.$$

Definition (Scalar multiplication of matrices)**.** Define $(\lambda\alpha)\mathbf{x} = \lambda[\alpha(\mathbf{x})]$. So $(\lambda A)_{ij} = \lambda A_{ij}$.

Definition (Matrix multiplication)**.** Consider maps $\alpha : \mathbb{R}^\ell \to \mathbb{R}^n$ and $\beta : \mathbb{R}^n \to \mathbb{R}^m$. The composition is $\beta\alpha : \mathbb{R}^\ell \to \mathbb{R}^m$. Take $\mathbf{x} \in \mathbb{R}^\ell \mapsto \mathbf{x}'' \in \mathbb{R}^m$. Then $\mathbf{x}'' = (BA)\mathbf{x} = B\mathbf{x}'$, where $\mathbf{x}' = A\mathbf{x}$. Using suffix notation, we have $x_i'' = (B\mathbf{x}')_i = b_{ik} x_k' = B_{ik} A_{kj} x_j$. But $x_i'' = (BA)_{ij} x_j$. So

$$(BA)_{ij} = B_{ik} A_{kj}.$$

Generally, an $m \times n$ matrix multiplied by an $n \times \ell$ matrix gives an $m \times \ell$ matrix. $(BA)_{ij}$ is given by the ith row of B dotted with the jth column of A.

Note that the number of columns of B has to be equal to the number of rows of A for multiplication to be defined. If $\ell = m$ as well, then both BA and AB make sense, but $AB \neq BA$ in general. In fact, they don't even have to have the same dimensions.

Also, since function composition is associative, we get $A(BC) = (AB)C$.

Definition (Transpose of matrix)**.** If A is an $m \times n$ matrix, the ***transpose*** A^T is an $n \times m$ matrix defined by $(A^T)_{ij} = A_{ji}$.

Proposition.

(i) $(A^T)^T = A$.

(ii) If $\mathbf{x}$ is a column vector $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, $\mathbf{x}^T$ is a row vector $(x_1\ x_2 \cdots x_n)$.

(iii) $(AB)^T = B^T A^T$ since $(AB)^T_{ij} = (AB)_{ji} = A_{jk}B_{ki} = B_{ki}A_{jk}$
$= (B^T)_{ik}(A^T)_{kj} = (B^T A^T)_{ij}$.

Definition (Hermitian conjugate)**.** Define $A^\dagger = (A^T)^*$. Similarly, $(AB)^\dagger = B^\dagger A^\dagger$.

Definition (Symmetric matrix)**.** A matrix is ***symmetric*** if $A^T = A$.

Definition (Hermitian matrix)**.** A matrix is ***Hermitian*** if $A^\dagger = A$. (The diagonal of a Hermitian matrix must be real).

Definition (Anti/skew symmetric matrix)**.** A matrix is ***anti-symmetric*** or ***skew symmetric*** if $A^T = -A$. The diagonals are all zero.

Definition (Skew-Hermitian matrix)**.** A matrix is ***skew-Hermitian*** if $A^\dagger = -A$. The diagonals are pure imaginary.

Definition (Trace of matrix)**.** The ***trace*** of an $n \times n$ matrix A is the sum of the diagonal. $\mathrm{tr}(A) = A_{ii}$.

Example. Consider the reflection matrix $R_{ij} = \delta_{ij} - 2\hat{n}_i\hat{n}_j$. We have $\mathrm{tr}(A) = R_{ii} = 3 - 2\hat{n}\cdot\hat{n} = 3 - 2 = 1$.

Proposition. $\mathrm{tr}(BC) = \mathrm{tr}(CB)$

Proof. $\mathrm{tr}(BC) = B_{ik}C_{ki} = C_{ki}B_{ik} = (CB)_{kk} = \mathrm{tr}(CB)$ □

Definition (Identity matrix)**.** $I = \delta_{ij}$.

Decomposition of an $n \times n$ matrix

Any $n \times n$ matrix B can be split as a sum of symmetric and antisymmetric parts. Write

$$B_{ij} = \underbrace{\frac{1}{2}(B_{ij} + B_{ji})}_{S_{ij}} + \underbrace{\frac{1}{2}(B_{ij} - B_{ji})}_{A_{ij}}.$$

We have $S_{ij} = S_{ji}$, so S is symmetric, while $A_{ji} = -A_{ij}$, and A is antisymmetric. So $B = S + A$.

Furthermore , we can decompose S into an isotropic part (a scalar multiple of the identity) plus a trace-less part (i.e. sum of diagonal = 0). Write

$$S_{ij} = \underbrace{\frac{1}{n}\operatorname{tr}(S)\delta_{ij}}_{\text{isotropic part}} + \underbrace{(S_{ij} - \frac{1}{n}\operatorname{tr}(S)\delta_{ij})}_{T_{ij}}.$$

We have $\operatorname{tr}(T) = T_{ii} = S_{ii} - \frac{1}{n}\operatorname{tr}(S)\delta_{ii} = \operatorname{tr}(S) - \frac{1}{n}\operatorname{tr}(S)(n) = 0$.
Putting all these together,

$$B = \frac{1}{n}\operatorname{tr}(B)I + \left\{\frac{1}{2}(B + B^T) - \frac{1}{n}\operatorname{tr}(B)I\right\} + \frac{1}{2}(B - B^T).$$

In three dimensions, we can write the antisymmetric part A in terms of a single vector: we have

$$A = \begin{pmatrix} 0 & a & -b \\ -a & 0 & c \\ b & -c & 0 \end{pmatrix}$$

and we can consider

$$\varepsilon_{ijk}\omega_k = \begin{pmatrix} 0 & \omega_3 & -\omega_2 \\ -\omega_3 & 0 & \omega_1 \\ \omega_2 & -\omega_1 & 0 \end{pmatrix}$$

So if we have $\omega = (c, b, a)$, then $A_{ij} = \varepsilon_{ijk}\omega_k$.
This decomposition can be useful in certain physical applications. For example, if the matrix represents the stress of a system, different parts of the decomposition will correspond to different types of stresses.

Matrix inverse

Definition (Inverse of matrix)**.** Consider an $m \times n$ matrix A and $n \times m$ matrices B and C. If $BA = I$, then we say B is the ***left inverse*** of A. If $AC = I$, then we say C is the ***right inverse*** of A. If A is square ($n \times n$), then $B = B(AC) = (BA)C = C$, i.e. the left and right inverses coincide. Both are denoted by A^{-1}, the ***inverse*** of A. Therefore we have

$$AA^{-1} = A^{-1}A = I.$$

Note that not all square matrices have inverses. For example, the zero matrix clearly has no inverse.

Definition (Invertible matrix)**.** If A has an inverse, then A is ***invertible***.

Proposition. $(AB)^{-1} = B^{-1}A^{-1}$

Proof. $(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}B = I$. □

Definition (Orthogonal and unitary matrices)**.** A real $n \times n$ matrix is ***orthogonal*** if $A^TA = AA^T = I$, i.e. $A^T = A^{-1}$. A complex $n \times n$ matrix is ***unitary*** if $U^\dagger U = UU^\dagger = I$, i.e. $U^\dagger = U^{-1}$.

Note that an orthogonal matrix A satisfies $A_{ik}(A^T_{kj}) = \delta_{ij}$, i.e. $A_{ik}A_{jk} = \delta_{ij}$. We can see this as saying "the scalar product of two distinct rows is 0, and the scalar product of a row with itself is 1". Alternatively, the rows (and columns — by considering A^T) of an orthogonal matrix form an orthonormal set.

Similarly, for a unitary matrix, $U_{ik}U^\dagger_{kj} = \delta_{ij}$, i.e. $u_{ik}u^*_{jk} = u^*_{ik}u_{jk} = \delta_{ij}$. i.e. the rows are orthonormal, using the definition of complex scalar product.

Example.

(i) The reflection in a plane is an orthogonal matrix. Since $R_{ij} = \delta_{ij} - 2n_i n_j$, We have

$$\begin{aligned} R_{ik}R_{jk} &= (\delta_{ik} - 2n_i n_k)(\delta_{jk} - 2n_j n_k) \\ &= \delta_{ik}\delta_{jk} - 2\delta_{jk}n_i n_k - 2\delta_{ik}n_j n_k + 2n_i n_k n_j n_k \\ &= \delta_{ij} - 2n_i n_j - 2n_j n_i + 4n_i n_j(n_k n_k) \\ &= \delta_{ij} \end{aligned}$$

(ii) The rotation is an orthogonal matrix. We could multiply out using suffix notation, but it would be cumbersome to do so. Alternatively, denote rotation matrix by θ about $\hat{\mathbf{n}}$ as $R(\theta, \hat{\mathbf{n}})$. Clearly, $R(\theta, \hat{\mathbf{n}})^{-1} = R(-\theta, \hat{\mathbf{n}})$. We have

$$\begin{aligned} R_{ij}(-\theta, \hat{\mathbf{n}}) &= (\cos\theta)\delta_{ij} + n_i n_j(1 - \cos\theta) + \varepsilon_{ijk}n_k \sin\theta \\ &= (\cos\theta)\delta_{ji} + n_j n_i(1 - \cos\theta) - \varepsilon_{jik}n_k \sin\theta \\ &= R_{ji}(\theta, \hat{\mathbf{n}}) \end{aligned}$$

In other words, $R(-\theta, \hat{\mathbf{n}}) = R(\theta, \hat{\mathbf{n}})^T$. So $R(\theta, \hat{\mathbf{n}})^{-1} = R(\theta, \hat{\mathbf{n}})^T$.

Determinants

Consider a linear map $\alpha : \mathbb{R}^3 \to \mathbb{R}^3$. The standard basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ is mapped to $\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3$ with $\mathbf{e}'_i = A\mathbf{e}_i$. Thus the unit cube formed by $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ is mapped to the parallelepiped with volume

$$\begin{aligned} [\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3] &= \varepsilon_{ijk}(e'_1)_i(e'_2)_j(e'_3)_k \\ &= \varepsilon_{ijk}A_{i\ell}\underbrace{(e_1)_\ell}_{\delta_{1\ell}} A_{jm}\underbrace{(e_2)_m}_{\delta_{2m}} A_{kn}\underbrace{(e_3)_n}_{\delta_{3n}} \\ &= \varepsilon_{ijk}A_{i1}A_{j2}A_{k3} \end{aligned}$$

We call this the determinant and write as

$$\det(A) = \begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{vmatrix}$$

Permutations

To define the determinant for square matrices of arbitrary size, we first have to consider ***permutations***.

Definition (Permutation)**.** A ***permutation*** of a set S is a bijection $\varepsilon : S \to S$.

Notation. Consider the set S_n of all permutations of $1, 2, 3, \cdots, n$. S_n contains $n!$ elements. Consider $\rho \in S_n$ with $i \mapsto \rho(i)$. We write

$$\rho = \begin{pmatrix} 1 & 2 & \cdots & n \\ \rho(1) & \rho(2) & \cdots & \rho(n) \end{pmatrix}.$$

Definition (Fixed point)**.** A ***fixed point*** of ρ is a k such that $\rho(k) = k$. e.g. in $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 3 & 2 \end{pmatrix}$, 3 is the fixed point. By convention, we can omit the fixed point and write as $\begin{pmatrix} 1 & 2 & 4 \\ 4 & 1 & 2 \end{pmatrix}$.

Definition (Disjoint permutation)**.** Two permutations are ***disjoint*** if numbers moved by one are fixed by the other, and vice versa. e.g. $\begin{pmatrix} 1 & 2 & 4 & 5 & 6 \\ 5 & 6 & 1 & 4 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 6 \\ 6 & 2 \end{pmatrix}\begin{pmatrix} 1 & 4 & 5 \\ 5 & 1 & 4 \end{pmatrix}$, and the two cycles on the right hand side are disjoint. Disjoint permutations commute, but in general non-disjoint permutations do not.

Definition (Transposition and k-cycle)**.** $\begin{pmatrix} 2 & 6 \\ 6 & 2 \end{pmatrix}$ is a ***2-cycle*** or a ***transposition***, and we can simply write $(2\ 6)$. $\begin{pmatrix} 1 & 4 & 5 \\ 5 & 1 & 4 \end{pmatrix}$ is a 3-cycle, and we can simply write $(1\ 5\ 4)$. (1 is mapped to 5; 5 is mapped to 4; 4 is mapped to 1)

Proposition. Any q-cycle can be written as a product of 2-cycles.

Proof. $(1\ 2\ 3\ \cdots\ n) = (1\ 2)(2\ 3)(3\ 4)\cdots(n-1\ n)$. □

Definition (Sign of permutation)**.** The ***sign*** of a permutation $\varepsilon(\rho)$ is $(-1)^r$, where r is the number of 2-cycles when ρ is written as a product of 2-cycles. If $\varepsilon(\rho) = +1$, it is an even permutation. Otherwise, it is an odd permutation. Note that $\varepsilon(\rho\sigma) = \varepsilon(\rho)\varepsilon(\sigma)$ and $\varepsilon(\rho^{-1}) = \varepsilon(\rho)$.

The proof that this is well-defined can be found in IA Groups.

Definition (Levi-Civita symbol)**.** The ***Levi-Civita*** symbol is defined by

$$\varepsilon_{j_1 j_2 \cdots j_n} = \begin{cases} +1 & \text{if } j_1 j_2 j_3 \cdots j_n \text{ is an even permutation of } 1, 2, \cdots n \\ -1 & \text{if it is an odd permutation} \\ 0 & \text{if any 2 of them are equal} \end{cases}$$

Clearly, $\varepsilon_{\rho(1)\rho(2)\cdots\rho(n)} = \varepsilon(\rho)$.

Definition (Determinant)**.** The ***determinant*** of an $n \times n$ matrix A is defined as:

$$\det(A) = \sum_{\sigma \in S_n} \varepsilon(\sigma) A_{\sigma(1)1} A_{\sigma(2)2} \cdots A_{\sigma(n)n},$$

or equivalently,

$$\det(A) = \varepsilon_{j_1 j_2 \cdots j_n} A_{j_1 1} A_{j_2 2} \cdots A_{j_n n}.$$

Proposition.

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

Properties of determinants

Proposition. $\det(A) = \det(A^T)$.

Proof. Take a single term $A_{\sigma(1)1} A_{\sigma(2)2} \cdots A_{\sigma(n)n}$ and let ρ be another permutation in S_n. We have

$$A_{\sigma(1)1} A_{\sigma(2)2} \cdots A_{\sigma(n)n} = A_{\sigma(\rho(1))\rho(1)} A_{\sigma(\rho(2))\rho(2)} \cdots A_{\sigma(\rho(n))\rho(n)}$$

since the right hand side is just re-ordering the order of multiplication. Choose $\rho = \sigma^{-1}$ and note that $\varepsilon(\sigma) = \varepsilon(\rho)$. Then

$$\det(A) = \sum_{\rho \in S_n} \varepsilon(\rho) A_{1\rho(1)} A_{2\rho(2)} \cdots A_{n\rho(n)} = \det(A^T). \qquad \square$$

Proposition. If matrix B is formed by multiplying every element in a single row of A by a scalar λ, then $\det(B) = \lambda \det(A)$. Consequently, $\det(\lambda A) = \lambda^n \det(A)$.

Proof. Each term in the sum is multiplied by λ, so the whole sum is multiplied by λ^n. $\square$

Proposition. If 2 rows (or 2 columns) of A are identical, the determinant is 0.

Proof. wlog, suppose columns 1 and 2 are the same. Then

$$\det(A) = \sum_{\sigma \in S_n} \varepsilon(\sigma) A_{\sigma(1)1} A_{\sigma(2)2} \cdots A_{\sigma(n)n}.$$

Now write an arbitrary σ in the form $\sigma = \rho(1\ 2)$. Then $\varepsilon(\sigma) = \varepsilon(\rho)\varepsilon((1\ 2)) = -\varepsilon(\rho)$. So

$$\det(A) = \sum_{\rho \in S_n} -\varepsilon(\rho) A_{\rho(2)1} A_{\rho(1)2} A_{\rho(3)3} \cdots A_{\rho(n)n}.$$

But columns 1 and 2 are identical, so $A_{\rho(2)1} = A_{\rho(2)2}$ and $A_{\rho(1)2} = A_{\rho(1)1}$. So $\det(A) = -\det(A)$ and $\det(A) = 0$. $\square$

Proposition. If 2 rows or 2 columns of a matrix are linearly dependent, then the determinant is zero.

Proof. Suppose in A, (column r) + λ(column s) = 0. Define

$$B_{ij} = \begin{cases} A_{ij} & j \neq r \\ A_{ij} + \lambda A_{is} & j = r \end{cases}.$$

Then $\det(B) = \det(A) + \lambda \det(\text{matrix with column } r = \text{column } s) = \det(A)$. Then we can see that the rth column of B is all zeroes. So each term in the sum contains one zero and $\det(A) = \det(B) = 0$. □

Even if we don't have linearly dependent rows or columns, we can still run the exact same proof as above, and still get that $\det(B) = \det(A)$. Linear dependence is only required to show that $\det(B) = 0$. So in general, we can add a linear multiple of a column (or row) onto another column (or row) without changing the determinant.

Proposition. Given a matrix A, if B is a matrix obtained by adding a multiple of a column (or row) of A to another column (or row) of A, then $\det A = \det B$.

Corollary. Swapping two rows or columns of a matrix negates the determinant.

Proof. We do the column case only. Let $A = (\mathbf{a}_1 \cdots \mathbf{a}_i \cdots \mathbf{a}_j \cdots \mathbf{a}_n)$. Then

$$\begin{aligned}
\det(\mathbf{a}_1 \cdots \mathbf{a}_i \cdots \mathbf{a}_j \cdots \mathbf{a}_n) &= \det(\mathbf{a}_1 \cdots \mathbf{a}_i + \mathbf{a}_j \cdots \mathbf{a}_j \cdots \mathbf{a}_n) \\
&= \det(\mathbf{a}_1 \cdots \mathbf{a}_i + \mathbf{a}_j \cdots \mathbf{a}_j - (\mathbf{a}_i + \mathbf{a}_j) \cdots \mathbf{a}_n) \\
&= \det(\mathbf{a}_1 \cdots \mathbf{a}_i + \mathbf{a}_j \cdots - \mathbf{a}_i \cdots \mathbf{a}_n) \\
&= \det(\mathbf{a}_1 \cdots \mathbf{a}_j \cdots - \mathbf{a}_i \cdots \mathbf{a}_n) \\
&= -\det(\mathbf{a}_1 \cdots \mathbf{a}_j \cdots \mathbf{a}_i \cdots \mathbf{a}_n)
\end{aligned}$$

Alternatively, we can prove this from the definition directly, using the fact that the sign of a transposition is -1 (and that the sign is multiplicative). □

Proposition. $\det(AB) = \det(A)\det(B)$.

Proof. First note that $\sum_\sigma \varepsilon(\sigma) A_{\sigma(1)\rho(1)} A_{\sigma(2)\rho(2)} = \varepsilon(\rho)\det(A)$, i.e. swapping columns (or rows) an even/odd number of times gives a factor ± 1 respectively. We can prove this by writing $\sigma = \mu\rho$.

Now

$$\begin{aligned}
\det AB &= \sum_\sigma \varepsilon(\sigma)(AB)_{\sigma(1)1}(AB)_{\sigma(2)2} \cdots (AB)_{\sigma(n)n} \\
&= \sum_\sigma \varepsilon(\sigma) \sum_{k_1, k_2, \cdots, k_n}^{n} A_{\sigma(1)k_1} B_{k_1 1} \cdots A_{\sigma(n)k_n} B_{k_n n} \\
&= \sum_{k_1, \cdots, k_n} B_{k_1 1} \cdots B_{k_n n} \underbrace{\sum_\sigma \varepsilon(\sigma) A_{\sigma(1)k_1} A_{\sigma(2)k_2} \cdots A_{\sigma(n)k_n}}_{S}
\end{aligned}$$

Now consider the many different S's. If in S, two of k_1 and k_n are equal, then S is a determinant of a matrix with two columns the same, i.e. $S = 0$. So we only have to consider the sum over distinct k_is. Thus the k_is are are a permutation of $1, \cdots n$, say $k_i = \rho(i)$. Then we can write

$$\begin{aligned}
\det AB &= \sum_\rho B_{\rho(1)1} \cdots B_{\rho(n)n} \sum_\sigma \varepsilon(\sigma) A_{\sigma(1)\rho(1)} \cdots A_{\sigma(n)\rho(n)} \\
&= \sum_\rho B_{\rho(1)1} \cdots B_{\rho(n)n} (\varepsilon(\rho) \det A) \\
&= \det A \sum_\rho \varepsilon(\rho) B_{\rho(1)1} \cdots B_{\rho(n)n} \\
&= \det A \det B
\end{aligned}$$

□

Corollary. If A is orthogonal, $\det A = \pm 1$.

Proof.

$$\begin{aligned}
AA^T &= I \\
\det AA^T &= \det I \\
\det A \det A^T &= 1 \\
(\det A)^2 &= 1 \\
\det A &= \pm 1
\end{aligned}$$

□

Corollary. If U is unitary, $|\det U| = 1$.

Proof. We have $\det U^\dagger = (\det U^T)^* = \det(U)^*$. Since $UU^\dagger = I$, we have $\det(U)\det(U)^* = 1$. □

Proposition. In $\mathbb{R}^3$, orthogonal matrices represent either a rotation ($\det = 1$) or a reflection ($\det = -1$).

Minors and Cofactors

Definition (Minor and cofactor)**.** For an $n \times n$ matrix A, define A^{ij} to be the $(n-1)\times(n-1)$ matrix in which row i and column j of A have been removed.

The ***minor*** of the ijth element of A is $M_{ij} = \det A^{ij}$

The ***cofactor*** of the ijth element of A is $\Delta_{ij} = (-1)^{i+j} M_{ij}$.

Notation. We use $\bar{\ }$ to denote a symbol which has been missed out of a natural sequence.

Example. $1, 2, 3, 5 = 1, 2, 3, \bar{4}, 5$.

The significance of these definitions is that we can use them to provide a systematic way of evaluating determinants. We will also use them to find inverses of matrices.

Theorem (Laplace expansion formula)**.** For any particular fixed i,

$$\det A = \sum_{j=1}^{n} A_{ji}\Delta_{ji}.$$

Proof.

$$\det A = \sum_{j_i=1}^{n} A_{j_i i} \sum_{j_1,\cdots,\overline{j_i},\cdots j_n}^{n} \varepsilon_{j_1 j_2 \cdots j_n} A_{j_1 1} A_{j_2 2} \cdots \overline{A_{j_i i}} \cdots A_{j_n n}$$

Let $\sigma \in S_n$ be the permutation which moves j_i to the ith position, and leave everything else in its natural order, i.e.

$$\sigma = \begin{pmatrix} 1 & \cdots & i & i+1 & i+2 & \cdots & j_i - 1 & j_i & j_i + 1 & \cdots & n \\ 1 & \cdots & j_i & i & i+1 & \cdots & j_i - 2 & j_i - 1 & j_i + 1 & \cdots & n \end{pmatrix}$$

if $j_i > i$, and similarly for other cases. To perform this permutation, $|i - j_i|$ transpositions are made. So $\varepsilon(\sigma) = (-1)^{i-j_i}$.

Now consider the permutation $\rho \in S_n$

$$\rho = \begin{pmatrix} 1 & \cdots & \cdots & \bar{j}_i & \cdots & n \\ j_1 & \cdots & \bar{j}_i & \cdots & \cdots & j_n \end{pmatrix}$$

The composition $\rho\sigma$ reorders $(1, \cdots, n)$ to $(j_1, j_2, \cdots, j_n)$. So $\varepsilon(\rho\sigma) = \varepsilon_{j_1\cdots j_n} = \varepsilon(\rho)\varepsilon(\sigma) = (-1)^{i-j_i}\varepsilon_{j_1\cdots\bar{j}_i\cdots j_n}$. Hence the original equation becomes

$$\begin{aligned}
\det A &= \sum_{j_i=1}^{n} A_{j_i i} \sum_{j_1\cdots\bar{j}_i\cdots j_n} (-1)^{i-j_i}\varepsilon_{j_1\cdots\bar{j}_i\cdots j_n} A_{j_1 1} \cdots \overline{A_{j_i i}} \cdots A_{j_n n} \\
&= \sum_{j_i=1}^{n} A_{j_i i}(-1)^{i-j_i} M_{j_i i} \\
&= \sum_{j_i=1}^{n} A_{j_i i}\Delta_{j_i i} \\
&= \sum_{j=1}^{n} A_{ji}\Delta_{ji}
\end{aligned}$$

□

Example. $\det A = \begin{vmatrix} 2 & 4 & 2 \\ 3 & 2 & 1 \\ 2 & 0 & 1 \end{vmatrix}$. We can pick the first row and have

$$\begin{aligned}
\det A &= 2\begin{vmatrix} 2 & 1 \\ 0 & 1 \end{vmatrix} - 4\begin{vmatrix} 3 & 1 \\ 2 & 1 \end{vmatrix} + 2\begin{vmatrix} 3 & 2 \\ 2 & 0 \end{vmatrix} \\
&= 2(2-0) - 4(3-2) + 2(0-4) \\
&= -8.
\end{aligned}$$

Alternatively, we can pick the second column and have

$$\begin{aligned}
\det A &= -4\begin{vmatrix} 3 & 1 \\ 2 & 1 \end{vmatrix} + 2\begin{vmatrix} 2 & 2 \\ 2 & 1 \end{vmatrix} - 0\begin{vmatrix} 2 & 2 \\ 3 & 1 \end{vmatrix} \\
&= -4(3-2) + 2(2-4) - 0 \\
&= -8.
\end{aligned}$$

In practical terms, we use a combination of properties of determinants with a sensible choice of i to evaluate $\det(A)$.

Example. Consider $\begin{vmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{vmatrix}$. Row 1 - row 2 gives

$$\begin{vmatrix} 0 & a-b & a^2-b^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{vmatrix} = (a-b)\begin{vmatrix} 0 & 1 & a+b \\ 1 & b & b^2 \\ 1 & c & c^2 \end{vmatrix}.$$

Do row 2 - row 3. We obtain

$$(a-b)(b-c)\begin{vmatrix} 0 & 1 & a+b \\ 0 & 1 & b+c \\ 1 & c & c^2 \end{vmatrix}.$$

Row 1 - row 2 gives

$$(a-b)(b-c)(a-c)\begin{vmatrix} 0 & 0 & 1 \\ 0 & 1 & b+c \\ 1 & c & c^2 \end{vmatrix} = (a-b)(b-c)(a-c).$$

1.5 Matrices and linear equations

Simple example, 2×2

Consider the system of equations

$$A_{11}x_1 + A_{12}x_2 = d_1 \tag{a}$$

$$A_{21}x_1 + A_{22}x_2 = d_2. \tag{b}$$

We can write this as

$$A\mathbf{x} = \mathbf{d}.$$

If we do (a)$\times A_{22}-$(b)$\times A_{12}$ and similarly the other way round, we obtain

$$\begin{aligned}(A_{11}A_{22} - A_{12}A_{21})x_1 &= A_{22}d_1 - A_{12}d_2\\ \underbrace{(A_{11}A_{22} - A_{12}A_{21})}_{\det A} x_2 &= A_{11}d_2 - A_{21}d_1\end{aligned}$$

Dividing by $\det A$ and writing in matrix form, we have

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{\det A}\begin{pmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{pmatrix}\begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$$

On the other hand, given the equation $A\mathbf{x} = \mathbf{d}$, if A^{-1} exists, then by multiplying both sides on the left by A^{-1}, we obtain $\mathbf{x} = A^{-1}\mathbf{d}$.
Hence, we have constructed A^{-1} in the 2×2 case, and shown that the condition for its existence is $\det A \neq 0$, with

$$A^{-1} = \frac{1}{\det A}\begin{pmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{pmatrix}$$

Inverse of an $n \times n$ matrix

For larger matrices, the formula for the inverse is similar, but slightly more complicated (and costly to evaluate). The key to finding the inverse is the following:

Lemma. $\sum A_{ik}\Delta_{jk} = \delta_{ij} \det A$.

Proof. If $i \neq j$, then consider an $n \times n$ matrix B, which is identical to A except the jth row is replaced by the ith row of A. So Δ_{jk} of $B = \Delta_{jk}$ of A, since Δ_{jk} does not depend on the elements in row j. Since B has a duplicate row, we know that

$$0 = \det B = \sum_{k=1}^{n} B_{jk}\Delta_{jk} = \sum_{k=1}^{n} A_{ik}\Delta_{jk}.$$

If $i = j$, then the expression is $\det A$ by the Laplace expansion formula. $\square$

Theorem. If $\det A \neq 0$, then A^{-1} exists and is given by

$$(A^{-1})_{ij} = \frac{\Delta_{ji}}{\det A}.$$

Proof.

$$(A^{-1})_{ik}A_{kj} = \frac{\Delta_{ki}}{\det A}A_{kj} = \frac{\delta_{ij}\det A}{\det A} = \delta_{ij}.$$

So $A^{-1}A = I$. □

The other direction is easy to prove. If $\det A = 0$, then it has no inverse, since for any matrix B, $\det AB = 0$, and hence AB cannot be the identity.

Example. Consider the shear matrix $S_\lambda = \begin{pmatrix} 1 & \lambda & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. We have $\det S_\lambda = 1$. The cofactors are

$$\begin{array}{lll} \Delta_{11} = 1 & \Delta_{12} = 0 & \Delta_{13} = 0 \\ \Delta_{21} - \lambda & \Delta_{22} = 1 & \Delta_{23} = 0 \\ \Delta_{31} = 0 & \Delta_{32} = 0 & \Delta_{33} = 1 \end{array}$$

So $S_\lambda^{-1} = \begin{pmatrix} 1 & -\lambda & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

How many arithmetic operations are involved in calculating the inverse of an $n \times n$ matrix? We just count multiplication operations since they are the most time-consuming. Suppose that calculating $\det A$ takes f_n multiplications. This involves n $(n-1) \times (n-1)$ determinants, and you need n more multiplications to put them together. So $f_n = nf_{n-1} + n$. So $f_n = O(n!)$ (in fact $f_n \approx (1+e)n!$).

To find the inverse, we need to calculate n^2 cofactors. Each is a $n-1$ determinant, and each takes $O((n-1)!)$. So the time complexity is $O(n^2(n-1)!) = O(n \cdot n!)$.

This is incredibly slow. Hence while it is theoretically possible to solve systems of linear equations by inverting a matrix, sane people do not do so in general. Instead, we develop certain better methods to solve the equations. In fact, the "usual" method people use to solve equations by hand only has complexity $O(n^3)$, which is a much better complexity.

Homogeneous and inhomogeneous equations

Consider $A\mathbf{x} = \mathbf{b}$ where A is an $n \times n$ matrix, $\mathbf{x}$ and $\mathbf{b}$ are $n \times 1$ column vectors.

Definition (Homogeneous equation)**.** If $\mathbf{b} = \mathbf{0}$, then the system is ***homogeneous***. Otherwise, it's ***inhomogeneous***.

Suppose $\det A \neq 0$. Then there is a unique solution $\mathbf{x} = A^{-1}\mathbf{b}$ ($\mathbf{x} = \mathbf{0}$ for homogeneous).

How can we understand this result? Recall that $\det A \neq 0$ means that the columns of A are linearly independent. The columns are the images of the standard basis, $\mathbf{e}'_i = A\mathbf{e_i}$. So $\det A \neq 0$ means that $\mathbf{e}'_i$ are linearly independent and form a basis of $\mathbb{R}^n$. Therefore the image is the whole of $\mathbb{R}^n$. This automatically ensures that $\mathbf{b}$ is in the image, i.e. there is a solution.

To show that there is exactly one solution, suppose $\mathbf{x}$ and $\mathbf{x}'$ are both solutions. Then $A\mathbf{x} = A\mathbf{x}' = \mathbf{b}$. So $A(\mathbf{x} - \mathbf{x}') = \mathbf{0}$. So $\mathbf{x} - \mathbf{x}'$ is in the kernel of A. But since the rank of A is n, by the rank-nullity theorem, the nullity is 0. So the kernel is trivial. So $\mathbf{x} - \mathbf{x}' = \mathbf{0}$, i.e. $\mathbf{x} = \mathbf{x}'$.

Gaussian elimination

Consider a general solution

$$\begin{aligned}
A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n &= d_1\\
A_{21}x_1 + A_{22}x_2 + \cdots + A_{2n}x_n &= d_2\\
&\vdots\\
A_{m1}x_1 + A_{m2}x_2 + \cdots + A_{mn}x_n &= d_m
\end{aligned}$$

So we have m equations and n unknowns.

Assume $A_{11} \neq 0$ (if not, we can re-order the equations). We can use the first equation to eliminate x_1 from the remaining $(m-1)$ equations. Then use the second equation to eliminate x_2 from the remaining $(m-2)$ equations (if anything goes wrong, just re-order until things work). Repeat.

We are left with

$$\begin{aligned}
A_{11}x_1 + A_{12}x_2 + A_{13}x_3 + \cdots + A_{1n}x_n &= d_1\\
A_{22}^{(2)}x_2 + A_{23}^{(2)}x_3 + \cdots + A_{2n}^{(2)}x_n &= d_2\\
&\vdots\\
A_{rr}^{(r)}x_r + \cdots + A_{rn}^{(r)}x_n &= d_r\\
0 &= d_{r+1}^{(r)}\\
&\vdots\\
0 &= d_m^{(r)}
\end{aligned}$$

Here $A_{ii}^{(i)} \neq 0$ (which we can achieve by re-ordering), and the superfix (i) refers to the "version number" of the coefficient, e.g. $A_{22}^{(2)}$ is the second version of the coefficient of x_2 in the second row.

Let's consider the different possibilities:

(i) $r < m$ and at least one of $d_{r+1}^{(r)}, \cdots d_m^{(r)} \neq 0$. Then a contradiction is reached. The system is inconsistent and has no solution. We say it is ***overdetermined***.

Example. Consider the system

$$\begin{aligned} 3x_1 + 2x_2 + x_3 &= 3 \\ 6x_1 + 3x_2 + 3x_3 &= 0 \\ 6x_1 + 2x_2 + 4x_3 &= 6 \end{aligned}$$

This becomes

$$\begin{aligned} 3x_1 + 2x_2 + x_3 &= 3 \\ 0 - x_2 + x_3 &= -6 \\ 0 - 2x_2 + 2x_3 &= 0 \end{aligned}$$

And then

$$\begin{aligned} 3x_1 + 2x_2 + x_3 &= 3 \\ 0 - x_2 + x_3 &= -6 \\ 0 &= 12 \end{aligned}$$

We have $d_3^{(3)} = 12 = 0$ and there is no solution.

(ii) If $r = n \leq m$, and all $d_{r+i}^{(r)} = 0$. Then from the nth equation, there is a unique solution for $x_n = d_n^{(n)}/A_{nn}^{(n)}$, and hence for all x_i by back substitution. This system is ***determined***.

Example.

$$\begin{aligned} 2x_1 + 5x_2 &= 2 \\ 4x_1 + 3x_2 &= 11 \end{aligned}$$

This becomes

$$\begin{aligned} 2x_1 + 5x_2 &= 2 \\ -7x_2 &= 7 \end{aligned}$$

So $x_2 = -1$ and thus $x_1 = 7/2$.

(iii) If $r < n$ and $d_{r+i}^{(r)} = 0$, then $x_{r+1}, \cdots x_n$ can be freely chosen, and there are infinitely many solutions. System is ***under-determined***. e.g.

$$\begin{aligned} x_1 + x_2 &= 1 \\ 2x_1 + 2x_2 &= 2 \end{aligned}$$

Which gives

$$\begin{aligned} x_1 + x_2 &= 1 \\ 0 &= 0 \end{aligned}$$

So $x_1 = 1 - x_2$ is a solution for any x_2.

In the $n = m$ case, there are $O(n^3)$ operations involved, which is much less than inverting the matrix. So this is an efficient way of solving equations.

This is also be related to the determinant. Consider the case where $m = n$ and A is square. Since row operations do not change the determinant and swapping rows give a factor of (-1). So

$$\det A = (-1)^k \begin{vmatrix} A_{11} & A_{12} & \cdots & \cdots & \cdots & A_{1n} \\ 0 & A_{22}^{(2)} & \cdots & \cdots & \cdots & A_{2n}^{(n)} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & A_{rr}^{(r)} & \cdots & A_{rn}^{(n)} \\ 0 & 0 & \cdots & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{vmatrix}$$

This determinant is an ***upper triangular*** one (all elements below diagonal are 0) and the determinant is the product of its diagonal elements.

Hence if $r < n$ (and $d_i^{(r)} = 0$ for $i > r$), then we have case (ii) and the $\det A = 0$. If $r = n$, then $\det A = (-1)^k A_{11} A_{22}^{(2)} \cdots A_{nn}^{(n)} \neq 0$.

Matrix rank

Consider a linear map $\alpha : \mathbb{R}^n \to \mathbb{R}^m$. Recall the rank $r(\alpha)$ is the dimension of the image. Suppose that the matrix A is associated with the linear map. We also call $r(A)$ the ***rank*** of A.

Recall that if the standard basis is $\mathbf{e}_1, \cdots \mathbf{e}_n$, then $A\mathbf{e}_1, \cdots, A\mathbf{e}_n$ span the image (but not necessarily linearly independent).

Further, $A\mathbf{e}_1, \cdots, A\mathbf{e}_n$ are the columns of the matrix A. Hence $r(A)$ is the number of linearly independent columns.

Definition (Column and row rank of linear map)**.** The ***column rank*** of a matrix is the maximum number of linearly independent columns.

The ***row rank*** of a matrix is the maximum number of linearly independent rows.

Theorem. The column rank and row rank are equal for any $m \times n$ matrix.

Proof. Let r be the row rank of A. Write the biggest set of linearly independent rows as $\mathbf{v}_1^T, \mathbf{v}_2^T, \cdots \mathbf{v}_r^T$ or in component form $\mathbf{v}_k^T = (v_{k1}, v_{k2}, \cdots, v_{kn})$ for $k = 1, 2, \cdots, r$.

Now denote the ith row of A as $\mathbf{r}_i^T = (A_{i1}, A_{i2}, \cdots A_{in})$.

Note that every row of A can be written as a linear combination of the $\mathbf{v}$'s. (If $\mathbf{r_i}$ cannot be written as a linear combination of the $\mathbf{v}$'s, then it is independent of the $\mathbf{v}$'s and $\mathbf{v}$ is not the maximum collection of linearly independent rows) Write

$$\mathbf{r}_i^T = \sum_{k=1}^{r} C_{ik} \mathbf{v}_k^T.$$

For some coefficients C_{ik} with $1 \leq i \leq m$ and $1 \leq k \leq r$.

Now the elements of A are

$$A_{ij} = (\mathbf{r}_i)^T_j = \sum_{k=1}^{r} C_{ik}(\mathbf{v}_k)_j,$$

or

$$\begin{pmatrix} A_{1j} \\ A_{2j} \\ \vdots \\ A_{mj} \end{pmatrix} = \sum_{k=1}^{r} \mathbf{v}_{kj} \begin{pmatrix} C_{1k} \\ C_{2k} \\ \vdots \\ C_{mk} \end{pmatrix}$$

So every column of A can be written as a linear combination of the r column vectors $\mathbf{c}_k$. Then the column rank of $A \leq r$, the row rank of A.

Apply the same argument to A^T to see that the row rank is $\leq$ the column rank. $\square$

Homogeneous problem $A\mathbf{x} = \mathbf{0}$

We restrict our attention to the square case, i.e. number of unknowns = number of equations. Here A is an $n \times n$ matrix. We want to solve $A\mathbf{x} = \mathbf{0}$.

First of all, if $\det A \neq 0$, then A^{-1} exists and $\mathbf{x}^{-1} = A^{-1}\mathbf{0} = \mathbf{0}$, which is the unique solution. Hence if $A\mathbf{x} = \mathbf{0}$ with $\mathbf{x} \neq \mathbf{0}$, then $\det A = 0$.

Geometrical interpretation

We consider a 3×3 matrix

$$A = \begin{pmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \mathbf{r}_3^T \end{pmatrix}$$

$A\mathbf{x} = \mathbf{0}$ means that $\mathbf{r}_i \cdot \mathbf{x} = 0$ for all i. Each equation $\mathbf{r}_i \cdot \mathbf{x} = 0$ represents a plane through the origin. So the solution is the intersection of the three planes.

There are three possibilities:

(i) If $\det A = [\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3] \neq 0$, $\operatorname{span}\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3\} = \mathbb{R}^3$ and thus $r(A) = 3$. By the rank-nullity theorem, $n(A) = 0$ and the kernel is $\{\mathbf{0}\}$. So $\mathbf{x} = \mathbf{0}$ is the unique solution.

(ii) If $\det A = 0$, then $\dim(\operatorname{span}\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3\}) = 1$ or 2.

   (a) If rank $= 2$, wlog assume $\mathbf{r}_1, \mathbf{r}_2$ are linearly independent. So $\mathbf{x}$ lies on the intersection of two planes $\mathbf{x} \cdot \mathbf{r}_1 = 0$ and $\mathbf{x} \cdot \mathbf{r}_2 = 0$, which is the line $\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} = \lambda \mathbf{r}_1 \times \mathbf{r}_2\}$ (Since $\mathbf{x}$ lies on the intersection of the two planes, it has to be normal to the normals of both planes). All such points on this line also satisfy $\mathbf{x} \cdot \mathbf{r}_3 = 0$ since $\mathbf{r}_3$ is a linear combination of $\mathbf{r}_1$ and $\mathbf{r}_2$. The kernel is a line, $n(A) = 1$.

   (b) If rank $= 1$, then $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ are parallel. So $\mathbf{x} \cdot \mathbf{r}_1 = 0 \Rightarrow \mathbf{x} \cdot \mathbf{r}_2 = \mathbf{x} \cdot \mathbf{r}_3 = 0$. So all $\mathbf{x}$ that satisfy $\mathbf{x} \cdot \mathbf{r}_1 = 0$ are in the kernel, and the kernel now is a plane. $n(A) = 2$.

(We also have the trivial case where $r(A) = 0$, we have the zero mapping and the kernel is $\mathbb{R}^3$)

Linear mapping view of $A\mathbf{x} = \mathbf{0}$

In the general case, consider a linear map $\alpha : \mathbb{R}^n \to \mathbb{R}^n$ $\mathbf{x} \mapsto \mathbf{x}' = A\mathbf{x}$. The kernel $k(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\}$ has dimension $n(A)$.

(i) If $n(A) = 0$, then $A(\mathbf{e}_1), A(\mathbf{e}_2), \cdots, A(\mathbf{e}_n)$ is a linearly independent set, and $r(A) = n$.

(ii) If $n(A) > 0$, then the image is not the whole of $\mathbb{R}^n$. Let $\{\mathbf{u}_i\}, i = 1, \cdots, n(A)$ be a basis of the kernel, i.e. so given any solution to $A\mathbf{x} = \mathbf{0}$, $\mathbf{x} = \sum_{i=1}^{n(A)} \lambda_i \mathbf{u}_i$ for some λ_i. Extend $\{\mathbf{u}_i\}$ to be a basis of $\mathbb{R}^n$ by introducing extra vectors $\mathbf{u}_i$ for $i = n(A)+1, \cdots, n$. The vectors $A(\mathbf{u}_i)$ for $i = n(A)+1, \cdots, n$ form a basis of the image.

General solution of $A\mathbf{x} = \mathbf{d}$

Finally consider the general equation $A\mathbf{x} = \mathbf{d}$, where A is an $n \times n$ matrix and $\mathbf{x}, \mathbf{d}$ are $n \times 1$ column vectors. We can separate into two main cases.

(i) $\det(A) \neq 0$. So A^{-1} exists and $n(A) = 0$, $r(A) = n$. Then for any $\mathbf{d} \in \mathbb{R}^n$, a unique solution must exists and it is $\mathbf{x} = A^{-1}\mathbf{d}$.

(ii) $\det(A) = 0$. Then A^{-1} does not exist, and $n(A) > 0$, $r(A) < n$. So the image of A is not the whole of $\mathbb{R}^n$.

   (a) If $\mathbf{d} \notin \operatorname{im} A$, then there is no solution (by definition of the image)

   (b) If $\mathbf{d} \in \operatorname{im} A$, then by definition there exists at least one $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{d}$. The general solution of $A\mathbf{x} = \mathbf{d}$ can be written as $\mathbf{x} = \mathbf{x}_0 + \mathbf{y}$, where $\mathbf{x}_0$ is a particular solution (i.e. $A\mathbf{x}_0 = \mathbf{d}$), and $\mathbf{y}$ is any vector in $\ker A$ (i.e. $A\mathbf{y} = \mathbf{0}$). (cf. Isomorphism theorem)
   If $n(A) = 0$, then $\mathbf{y} = \mathbf{0}$ only, and then the solution is unique (i.e. case (i)). If $n(A) > 0$, then $\{\mathbf{u}_i\}, i = 1, \cdots, n(A)$ is a basis of the kernel. Hence

$$\mathbf{y} = \sum_{j=1}^{n(A)} \mu_j \mathbf{u}_j,$$

   so

$$\mathbf{x} = \mathbf{x}_0 + \sum_{j=1}^{n(A)} \mu_j \mathbf{u}_j$$

   for any μ_j, i.e. there are infinitely many solutions.

Example.

$$\begin{pmatrix} 1 & 1 \\ a & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ b \end{pmatrix}$$

We have $\det A = 1 - a$. If $a \neq 1$, then A^{-1} exists and

$$A^{-1} = \frac{1}{1-a} = \frac{1}{1-a}\begin{pmatrix} 1 & -1 \\ -a & 1 \end{pmatrix}.$$

Then

$$\mathbf{x} = \frac{1}{1-a}\begin{pmatrix} 1-b \\ -a+b \end{pmatrix}.$$

If $a = 1$, then

$$A\mathbf{x} = \begin{pmatrix} x_1 + x_2 \\ x_1 + x_2 \end{pmatrix} = (x_1 + x_2)\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

So $\operatorname{im} A = \operatorname{span}\left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\}$ and $\ker A = \operatorname{span}\left\{\begin{pmatrix} 1 \\ -1 \end{pmatrix}\right\}$. If $b \neq 1$, then $\begin{pmatrix} 1 \\ b \end{pmatrix} \notin \operatorname{im} A$ and there is no solution. If $b = 1$, then $\begin{pmatrix} 1 \\ b \end{pmatrix} \in \operatorname{im} A$.

We find a particular solution of $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. So The general solution is

$$\mathbf{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \lambda\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Example. Find the general solution of

$$\begin{pmatrix} a & a & b \\ b & a & a \\ a & b & a \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ c \\ 1 \end{pmatrix}$$

We have $\det A = (a-b)^2(2a+b)$. If $a \neq b$ and $b \neq -2a$, then the inverse exists and there is a unique solution for any c. Otherwise, the possible cases are

(i) $a = b, b \neq -2a$. So $a \neq 0$. The kernel is the plane $x + y + z = 0$ which is $\operatorname{span}\left\{\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}\right\}$ We extend this basis to $\mathbb{R}^3$ by adding $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$.

So the image is the span of $\begin{pmatrix} a \\ a \\ a \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. Hence if $c \neq 1$, then $\begin{pmatrix} 1 \\ c \\ 1 \end{pmatrix}$ is not in the image and there is no solution. If $c = 1$, then a particular solution is $\begin{pmatrix} \frac{1}{a} \\ 0 \\ 0 \end{pmatrix}$ and the general solution is

$$\mathbf{x} = \begin{pmatrix} \frac{1}{a} \\ 0 \\ 0 \end{pmatrix} + \lambda\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + \mu\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

(ii) If $a \neq b$ and $b = -2a$, then $a \neq 0$. The kernel satisfies

$$\begin{aligned} x + y - 2z &= 0 \\ -2x + y + z &= 0 \\ x - 2y + z &= 0 \end{aligned}$$

This can be solved to give $x = y = z$, and the kernel is span $\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$. We add $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ to form a basis of $\mathbb{R}^3$. So the image is the span of $\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}$.

If $\begin{pmatrix} 1 \\ c \\ 1 \end{pmatrix}$ is in the image, then

$$\begin{pmatrix} 1 \\ c \\ 1 \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} + \mu \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}.$$

Then the only solution is $\mu = 0, \lambda = 1, c = -2$. Thus there is no solution if $c \neq -2$, and when $c = -2$, pick a particular solution $\begin{pmatrix} \frac{1}{a} \\ 0 \\ 0 \end{pmatrix}$ and the general solution is

$$\mathbf{x} = \begin{pmatrix} \frac{1}{a} \\ 0 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

(iii) If $a = b$ and $b = -2a$, then $a = b = 0$ and $\ker A = \mathbb{R}^3$. So there is no solution for any c.

1.6 Eigenvalues and eigenvectors

Given a matrix A, an eigenvector is a vector $\mathbf{x}$ that satisfies $A\mathbf{x} = \lambda\mathbf{x}$ for some λ. We call λ the associated eigenvalue. In some sense, these vectors are not modified by the matrix, and are just scaled up by the matrix. We will look at the properties of eigenvectors and eigenvalues, and see their importance in diagonalizing matrices.

Preliminaries and definitions

Theorem (Fundamental theorem of algebra)**.** Let $p(z)$ be a polynomial of degree $m \geq 1$, i.e.

$$p(z) = \sum_{j=0}^{m} c_j z^j,$$

where $c_j \in \mathbb{C}$ and $c_m \neq 0$.

Then $p(z) = 0$ has precisely m (not necessarily distinct) roots in the complex plane, accounting for multiplicity.

Note that we have the disclaimer "accounting for multiplicity". For example, $x^2 - 2x + 1 = 0$ has only one distinct root, 1, but we say that this root has multiplicity 2, and is thus counted twice. Formally, multiplicity is defined as follows:

Definition (Multiplicity of root)**.** The root $z = \omega$ has ***multiplicity*** k if $(z - \omega)^k$ is a factor of $p(z)$ but $(z - \omega)^{k+1}$ is not.

Example. Let $p(z) = z^3 - z^2 - z + 1 = (z - 1)^2(z + 1)$. So $p(z) = 0$ has roots $1, 1, -1$, where $z = 1$ has multiplicity 2.

Definition (Eigenvector and eigenvalue)**.** Let $\alpha : \mathbb{C}^n \to \mathbb{C}^n$ be a linear map with associated matrix A. Then $\mathbf{x} \neq \mathbf{0}$ is an ***eigenvector*** of A if

$$A\mathbf{x} = \lambda\mathbf{x}$$

for some λ. λ is the associated ***eigenvalue***. This means that the direction of the eigenvector is preserved by the mapping, but is scaled up by λ.

There is a rather easy way of finding eigenvalues:

Theorem. λ is an eigenvalue of A iff

$$\det(A - \lambda I) = 0.$$

Proof. ($\Rightarrow$) Suppose that λ is an eigenvalue and $\mathbf{x}$ is the associated eigenvector. We can rearrange the equation in the definition above to

$$(A - \lambda I)\mathbf{x} = \mathbf{0}$$

and thus

$$\mathbf{x} \in \ker(A - \lambda I)$$

But $\mathbf{x} \neq \mathbf{0}$. So $\ker(A - \lambda I)$ is non-trivial and $\det(A - \lambda I) = 0$. The ($\Leftarrow$) direction is similar. □

Definition (Characteristic equation of matrix)**.** The ***characteristic equation*** of A is

$$\det(A - \lambda I) = 0.$$

Definition (Characteristic polynomial of matrix)**.** The ***characteristic polynomial*** of A is

$$p_A(\lambda) = \det(A - \lambda I).$$

From the definition of the determinant,

$$\begin{aligned} p_A(\lambda) &= \det(A - \lambda I) \\ &= \varepsilon_{j_1 j_2 \cdots j_n}(A_{j_1 1} - \lambda \delta_{j_1 1}) \cdots (A_{j_n n} - \lambda \delta_{j_n n}) \\ &= c_0 + c_1 \lambda + \cdots + c_n \lambda^n \end{aligned}$$

for some constants $c_0, \cdots, c_n$. From this, we see that

(i) $p_A(\lambda)$ has degree n and has n roots. So an $n \times n$ matrix has n eigenvalues (accounting for multiplicity).

(ii) If A is real, then all $c_i \in \mathbb{R}$. So eigenvalues are either real or come in complex conjugate pairs.

(iii) $c_n = (-1)^n$ and $c_{n-1} = (-1)^{n-1}(A_{11} + A_{22} + \cdots + A_{nn}) = (-1)^{n-1}\operatorname{tr}(A)$. But c_{n-1} is the sum of roots, i.e. $c_{n-1} = (-1)^{n-1}(\lambda_1 + \lambda_2 + \cdots \lambda_n)$, so

$$\operatorname{tr}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n.$$

Finally, $c_0 = p_A(0) = \det(A)$. Also c_0 is the product of all roots, i.e. $c_0 = \lambda_1 \lambda_2 \cdots \lambda_n$. So

$$\det A = \lambda_1 \lambda_2 \cdots \lambda_n.$$

The kernel of the matrix $A - \lambda I$ is the set $\{\mathbf{x} : A\mathbf{x} = \lambda \mathbf{x}\}$. This is a vector subspace because the kernel of any map is always a subspace.

Definition (Eigenspace)**.** The ***eigenspace*** denoted by E_λ is the kernel of the matrix $A - \lambda I$, i.e. the set of eigenvectors with eigenvalue λ.

Definition (Algebraic multiplicity of eigenvalue)**.** The ***algebraic multiplicity*** $M(\lambda)$ or M_λ of an eigenvalue λ is the multiplicity of λ in $p_A(\lambda) = 0$. By the fundamental theorem of algebra,

$$\sum_\lambda M(\lambda) = n.$$

If $M(\lambda) > 1$, then the eigenvalue is ***degenerate***.

Definition (Geometric multiplicity of eigenvalue)**.** The ***geometric multiplicity*** $m(\lambda)$ or m_λ of an eigenvalue λ is the dimension of the eigenspace, i.e. the maximum number of linearly independent eigenvectors with eigenvalue λ.

Definition (Defect of eigenvalue)**.** The ***defect*** Δ_λ of eigenvalue λ is

$$\Delta_\lambda = M(\lambda) - m(\lambda).$$

It can be proven that $\Delta_\lambda \geq 0$, i.e. the geometric multiplicity is never greater than the algebraic multiplicity.

Linearly independent eigenvectors

Theorem. Suppose $n \times n$ matrix A has ***distinct*** eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$. Then the corresponding eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ are linearly independent.

Proof. Proof by contradiction: Suppose $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ are linearly dependent. Then we can find non-zero constants d_i for $i = 1, 2, \cdots, r$, such that

$$d_1\mathbf{x}_1 + d_2\mathbf{x}_2 + \cdots + d_r\mathbf{x}_r = \mathbf{0}.$$

Suppose that this is the shortest non-trivial linear combination that gives $\mathbf{0}$ (we may need to re-order $\mathbf{x}_i$).
Now apply $(A - \lambda_1 I)$ to the whole equation to obtain

$$d_1(\lambda_1 - \lambda_1)\mathbf{x}_1 + d_2(\lambda_2 - \lambda_1)\mathbf{x}_2 + \cdots + d_r(\lambda_r - \lambda_1)\mathbf{x}_r = \mathbf{0}.$$

We know that the first term is $\mathbf{0}$, while the others are not (since we assumed $\lambda_i \neq \lambda_j$ for $i \neq j$). So

$$d_2(\lambda_2 - \lambda_1)\mathbf{x}_2 + \cdots + d_r(\lambda_r - \lambda_1)\mathbf{x}_r = \mathbf{0},$$

and we have found a shorter linear combination that gives $\mathbf{0}$. Contradiction. $\square$

Example.

(i) $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Then $p_A(\lambda) = \lambda^2 + 1 = 0$. So $\lambda_1 = i$ and $\lambda_2 = -i$.

To solve $(A - \lambda_1 I)\mathbf{x} = \mathbf{0}$, we obtain

$$\begin{pmatrix} -i & 1 \\ -1 & -i \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{0}.$$

So we obtain

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ i \end{pmatrix}$$

to be an eigenvector. Clearly any scalar multiple of $\begin{pmatrix} 1 \\ i \end{pmatrix}$ is also a solution, but still in the same eigenspace $E_i = \operatorname{span}\begin{pmatrix} 1 \\ i \end{pmatrix}$

Solving $(A - \lambda_2 I)\mathbf{x} = \mathbf{0}$ gives

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -i \end{pmatrix}.$$

So $E_{-i} = \operatorname{span}\begin{pmatrix} 1 \\ -i \end{pmatrix}$.

Note that $M(\pm i) = m(\pm i) = 1$, so $\Delta_{\pm i} = 0$. Also note that the two eigenvectors are linearly independent and form a basis of $\mathbb{C}^2$.

(ii) Consider

$$A = \begin{pmatrix} -2 & 2 & -3 \\ 2 & 1 & -6 \\ -1 & -2 & 0 \end{pmatrix}$$

Then $\det(A - \lambda I) = 0$ gives $45 + 21\lambda - \lambda^2 - \lambda^3$. So $\lambda_1 = 5, \lambda_2 = \lambda_3 = -3$.

The eigenvector with eigenvalue 5 is

$$\mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$$

We can find that the eigenvectors with eigenvalue -3 are

$$\mathbf{x} = \begin{pmatrix} -2x_2 + 3x_3 \\ x_2 \\ x_3 \end{pmatrix}$$

for any x_2, x_3. This gives two linearly independent eigenvectors, say $\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}$.

So $M(5) = m(5) = 1$ and $M(-3) = m(-3) = 2$, and there is no defect for both of them. Note that these three eigenvectors form a basis of $\mathbb{C}^3$.

(iii) Let

$$A = \begin{pmatrix} -3 & -1 & 1 \\ -1 & -3 & 1 \\ -2 & -2 & 0 \end{pmatrix}$$

Then $0 = p_A(\lambda) = -(\lambda + 2)^4$. So $\lambda = -2, -2, -2$. To find the eigenvectors, we have

$$(A + 2I)\mathbf{x} = \begin{pmatrix} -1 & -1 & 1 \\ -1 & -1 & 1 \\ -2 & -2 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \mathbf{0}$$

The general solution is thus $x_1 + x_2 - x_3 = 0$, and the general solution is thus $x = \begin{pmatrix} x_1 \\ x_2 \\ x_1 + x_2 \end{pmatrix}$. The eigenspace $E_{-2} = \operatorname{span}\left\{\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}\right\}$.

Hence $M(-2) = 3$ and $m(-2) = 2$. Thus the defect $\Delta_{-2} = 1$. So the eigenvectors do not form a basis of $\mathbb{C}^3$.

(iv) Consider the reflection R in the plane with normal $\mathbf{n}$. Clearly $R\mathbf{n} = -\mathbf{n}$. The eigenvalue is -1 and the eigenvector is $\mathbf{n}$. Then $E_1 = \operatorname{span}\{\mathbf{n}\}$. So $M(-1) = m(-1) = 1$.

If $\mathbf{p}$ is any vector in the plane, $R\mathbf{p} = \mathbf{p}$. So this has an eigenvalue of 1 and eigenvectors being any vector in the plane. So $M(1) = m(1) = 2$.

So the eigenvectors form a basis of $\mathbb{R}^3$.

(v) Consider a rotation R by θ about $\mathbf{n}$. Since $R\mathbf{n} = \mathbf{n}$, we have an eigenvalue of 1 and eigenspace $E_1 = \text{span}\{\mathbf{n}\}$.

We know that there are no other real eigenvalues since rotation changes the direction of any other vector. The other eigenvalues turn out to be $e^{\pm i\theta}$. If $\theta \neq 0$, there are 3 distinct eigenvalues and the eigenvectors form a basis of $\mathbb{C}^3$.

(vi) Consider a shear

$$A = \begin{pmatrix} 1 & \mu \\ 0 & 1 \end{pmatrix}$$

The characteristic equation is $(1-\lambda)^2 = 0$ and $\lambda = 1$. The eigenvectors corresponding to $\lambda = 1$ is $\mathbf{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. We have $M(1) = 2$ and $m(1) = 1$. So $\Delta_1 = 1$.

If $n \times n$ matrix A has n distinct eigenvalues, and hence has n linearly independent eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_n$, then ***with respect to this eigenvector basis***, A is diagonal. In this basis, $v_1 = (1, 0, \cdots, 0)$ etc. We know that $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$ (no summation). So the image of the ith basis vector is λ_i times the ith basis. Since the columns of A are simply the images of the basis,

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

The fact that A can be diagonalized by changing the basis is an important observation. We will now look at how we can change bases and see how we can make use of this.

Transformation matrices

How do the components of a vector or a matrix change when we change the basis?

Let $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$ and $\{\tilde{\mathbf{e}}_1, \tilde{\mathbf{e}}_2, \cdots, \tilde{\mathbf{e}}_n\}$ be 2 different bases of $\mathbb{R}^n$ or $\mathbb{C}^n$. Then we can write

$$\tilde{\mathbf{e}_j} = \sum_{i=1}^{n} P_{ij} \mathbf{e}_i$$

i.e. P_{ij} is the ith component of $\tilde{\mathbf{e}_j}$ with respect to the basis $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$. Note that the sum is made as $P_{ij}\mathbf{e}_i$, not $P_{ij}\mathbf{e}_j$. This is different from the formula for matrix multiplication. Matrix P has as its columns the vectors $\tilde{\mathbf{e}_j}$ relative to $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$. So $P = (\tilde{\mathbf{e}_1}\ \tilde{\mathbf{e}_2}\ \cdots\ \tilde{\mathbf{e}_n})$ and

$$P(\mathbf{e}_i) = \tilde{\mathbf{e}_i}$$

Similarly, we can write

$$\mathbf{e}_i = \sum_{k=1}^{n} Q_{ki} \tilde{\mathbf{e}_k}$$

with $Q = (\mathbf{e}_1\ \mathbf{e}_2\ \cdots\ \mathbf{e}_n)$.

Substituting this into the equation for $\tilde{\mathbf{e}}_j$, we have

$$\begin{aligned}\tilde{\mathbf{e}}_j &= \sum_{i=1}^{n}\left(\sum_{k=1}^{n} Q_{ki}\tilde{\mathbf{e}}_k\right) P_{ij}\\ &= \sum_{k=1}^{n}\tilde{\mathbf{e}}_k\left(\sum_{i=1}^{n} Q_{ki}P_{ij}\right)\end{aligned}$$

But $\tilde{\mathbf{e}}_1, \tilde{\mathbf{e}}_2, \cdots, \tilde{\mathbf{e}}_n$ are linearly independent, so this is only possible if

$$\sum_{i=1}^{n} Q_{ki}P_{ij} = \delta_{kj},$$

which is just a fancy way of saying $QP = I$, or $Q = P^{-1}$.

Transformation law for vectors

With respect to basis $\{\mathbf{e}_i\}$, $\mathbf{u} = \sum_{i=1}^{n} u_i\mathbf{e}_i$. With respect to basis $\{\tilde{\mathbf{e}}_i\}$, $\mathbf{u} = \sum_{i=1}^{n}\tilde{u}_i\tilde{\mathbf{e}}_i$. Note that this is the ***same*** vector $\mathbf{u}$ but has different components with respect to different bases. Using the transformation matrix above for the basis, we have

$$\begin{aligned}\mathbf{u} &= \sum_{j=1}^{n}\tilde{u}_j\sum_{i=1}^{n} P_{ij}\mathbf{e}_i\\ &= \sum_{i=1}^{n}\left(\sum_{j=1}^{n} P_{ij}\tilde{u}_j\right)\mathbf{e}_i\end{aligned}$$

By comparison, we know that

$$u_i = \sum_{j=1}^{n} P_{ij}\tilde{u}_j$$

Theorem. Denote vector as $\mathbf{u}$ with respect to $\{\mathbf{e}_i\}$ and $\tilde{\mathbf{u}}$ with respect to $\{\tilde{\mathbf{e}}_i\}$. Then

$$\mathbf{u} = P\tilde{\mathbf{u}} \text{ and } \tilde{\mathbf{u}} = P^{-1}\mathbf{u}$$

Example. Take the first basis as $\{\mathbf{e}_1 = (1,0), \mathbf{e}_2 = (0,1)\}$ and the second as $\{\tilde{\mathbf{e}}_1 = (1,1), \tilde{\mathbf{e}}_2 = (-1,1)\}$.
So $\tilde{\mathbf{e}}_1 = \mathbf{e}_1 + \mathbf{e}_2$ and $\tilde{\mathbf{e}}_2 = -\mathbf{e}_1 + \mathbf{e}_2$. We have

$$\mathbf{P} = \begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix}.$$

Then for an arbitrary vector $\mathbf{u}$, we have

$$\begin{aligned}\mathbf{u} &= u_1\mathbf{e}_1 + u_2\mathbf{e}_2\\ &= u_1\frac{1}{2}(\tilde{\mathbf{e}}_1 - \tilde{\mathbf{e}}_2) + u_2\frac{1}{2}(\tilde{\mathbf{e}}_1 + \tilde{\mathbf{e}}_2)\\ &= \frac{1}{2}(u_1 + u_2)\tilde{\mathbf{e}}_1 + \frac{1}{2}(-u_1 + u_2)\tilde{\mathbf{e}}_2.\end{aligned}$$

Alternatively, using the formula above, we obtain

$$\begin{aligned}\tilde{\mathbf{u}} &= P^{-1}\mathbf{u}\\ &= \frac{1}{2}\begin{pmatrix}1 & 1\\ -1 & 1\end{pmatrix}\begin{pmatrix}u_1\\ u_2\end{pmatrix}\\ &= \begin{pmatrix}\frac{1}{2}(u_1+u_2)\\ \frac{1}{2}(-u_1+u_2)\end{pmatrix}\end{aligned}$$

Which agrees with the above direct expansion.

Transformation law for matrix

Consider a linear map $\alpha : \mathbb{C}^n \to \mathbb{C}^n$ with associated $n \times n$ matrix A. We have

$$\mathbf{u}' = \alpha(\mathbf{u}) = A\mathbf{u}.$$

Denote $\mathbf{u}$ and $\mathbf{u}'$ as being with respect to basis $\{\mathbf{e}_i\}$ (i.e. same basis in both spaces), and $\tilde{\mathbf{u}}, \tilde{\mathbf{u}}'$ with respect to $\{\tilde{\mathbf{e}_i}\}$.

Using what we've got above, we have

$$\begin{aligned}\mathbf{u}' &= A\mathbf{u}\\ P\tilde{\mathbf{u}}' &= AP\tilde{\mathbf{u}}\\ \tilde{\mathbf{u}}' &= P^{-1}AP\tilde{\mathbf{u}}\\ &= \tilde{A}\tilde{\mathbf{u}}\end{aligned}$$

So

Theorem.

$$\tilde{A} = P^{-1}AP.$$

Example. Consider the shear $S_\lambda = \begin{pmatrix}1 & \lambda & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}$ with respect to the standard basis. Choose a new set of basis vectors by rotating by θ about the $\mathbf{e}_3$ axis:

$$\begin{aligned}\tilde{\mathbf{e}_1} &= \cos\theta\mathbf{e}_1 + \sin\theta\mathbf{e}_2\\ \tilde{\mathbf{e}_2} &= -\sin\theta\mathbf{e}_1 + \cos\theta\mathbf{e}_2\\ \tilde{\mathbf{e}_3} &= \mathbf{e}_3\end{aligned}$$

So we have

$$P = \begin{pmatrix}\cos\theta & -\sin\theta & 0\\ \sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{pmatrix}, P^{-1} = \begin{pmatrix}\cos\theta & \sin\theta & 0\\ -\sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{pmatrix}$$

Now use the basis transformation laws to obtain

$$\tilde{S}_\lambda = \begin{pmatrix}1+\lambda\sin\theta\cos\theta & \lambda\cos^2\theta & 0\\ -\lambda\sin^2\theta & 1-\lambda\sin\theta\cos\theta & 0\\ 0 & 0 & 1\end{pmatrix}$$

Clearly this is much more complicated than our original basis. This shows that choosing a sensible basis is important.

More generally, given $\alpha : \mathbb{C}^m \to \mathbb{C}^n$, given $\mathbf{x} \in \mathbb{C}^m$, $\mathbf{x}' \in \mathbb{C}^n$ with $\mathbf{x}' = A\mathbf{x}$. We know that A is an $n \times m$ matrix.
Suppose $\mathbb{C}^m$ has a basis $\{\mathbf{e}_i\}$ and $\mathbb{C}^n$ has a basis $\{\mathbf{f}_i\}$. Now change bases to $\{\tilde{\mathbf{e}}_i\}$ and $\{\tilde{\mathbf{f}}_i\}$. We know that $\mathbf{x} = P\tilde{\mathbf{x}}$ with P being an $m \times m$ matrix, with $\mathbf{x}' = R\tilde{\mathbf{x}}'$ with R being an $n \times n$ matrix.
Combining both of these, we have

$$\begin{aligned} R\tilde{\mathbf{x}}' &= AP\tilde{\mathbf{x}} \\ \tilde{\mathbf{x}}' &= R^{-1}AP\tilde{\mathbf{x}} \end{aligned}$$

Therefore $\tilde{A} = R^{-1}AP$.

Example. Consider $\alpha : \mathbb{R}^3 \to \mathbb{R}^2$, with respect to the standard bases in both spaces,

$$A = \begin{pmatrix} 2 & 3 & 4 \\ 1 & 6 & 3 \end{pmatrix}$$

Use a new basis $\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 5 \end{pmatrix}$ in $\mathbb{R}^2$ and keep the standard basis in $\mathbb{R}^3$. The basis change matrix in $\mathbb{R}^3$ is simply I, while

$$R = \begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}, R^{-1} = \frac{1}{9}\begin{pmatrix} 5 & -1 \\ -1 & 2 \end{pmatrix}$$

is the transformation matrix for $\mathbb{R}^2$. So

$$\begin{aligned} \tilde{A} &= \begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}\begin{pmatrix} 2 & 3 & 4 \\ 1 & 6 & 3 \end{pmatrix} I \\ &= \frac{1}{9}\begin{pmatrix} 5 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 2 & 3 & 4 \\ 1 & 6 & 3 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 1 & 17/9 \\ 0 & 1 & 2/9 \end{pmatrix} \end{aligned}$$

We can alternatively do it this way: we know that $\tilde{\mathbf{f}_1} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \tilde{\mathbf{f}_2} = \begin{pmatrix} 1 \\ 5 \end{pmatrix}$ Then we know that

$$\begin{aligned} \tilde{\mathbf{e}_1} &= \mathbf{e}_1 \mapsto 2\mathbf{f}_1 + \mathbf{f}_2 = \mathbf{f}_1 \\ \tilde{\mathbf{e}_2} &= \mathbf{e}_2 \mapsto 3\mathbf{f}_1 + 6\mathbf{f}_2 = \tilde{\mathbf{f}_1} + \tilde{\mathbf{f}_2} \\ \tilde{\mathbf{e}_3} &= \mathbf{e}_3 \mapsto 4\mathbf{f}_1 + 3\mathbf{f}_2 = \frac{17}{9}\tilde{\mathbf{f}_1} + \frac{2}{9}\tilde{\mathbf{f}_2} \end{aligned}$$

and we can construct the matrix correspondingly.

Similar matrices

Definition (Similar matrices)**.** Two $n \times n$ matrices A and B are ***similar*** if there exists an invertible matrix P such that

$$B = P^{-1}AP,$$

i.e. they represent the same map under different bases. Alternatively, using the language from IA Groups, we say that they are in the same conjugacy class.

Proposition. Similar matrices have the following properties:

(i) Similar matrices have the same determinant.

(ii) Similar matrices have the same trace.

(iii) Similar matrices have the same characteristic polynomial.

Note that (iii) implies (i) and (ii) since the determinant and trace are the coefficients of the characteristic polynomial

Proof. They are proven as follows:

(i) $\det B = \det(P^{-1}AP) = (\det A)(\det P)^{-1}(\det P) = \det A$

(ii)

$$\begin{aligned} \operatorname{tr} B &= B_{ii} \\ &= P^{-1}_{ij} A_{jk} P_{ki} \\ &= A_{jk} P_{ki} P^{-1}_{ij} \\ &= A_{jk} (PP^{-1})_{kj} \\ &= A_{jk} \delta_{kj} \\ &= A_{jj} \\ &= \operatorname{tr} A \end{aligned}$$

(iii)

$$\begin{aligned} p_B(\lambda) &= \det(B - \lambda I) \\ &= \det(P^{-1}AP - \lambda I) \\ &= \det(P^{-1}AP - \lambda P^{-1}IP) \\ &= \det(P^{-1}(A - \lambda I)P) \\ &= \det(A - \lambda I) \\ &= p_A(\lambda) \end{aligned}$$

□

Diagonalizable matrices

Definition (Diagonalizable matrices)**.** An $n \times n$ matrix A is ***diagonalizable*** if it is similar to a diagonal matrix. We showed above that this is equivalent to saying the eigenvectors form a basis of $\mathbb{C}^n$.

The requirement that matrix A has n distinct eigenvalues is a ***sufficient*** condition for diagonalizability as shown above. However, it is ***not*** necessary.
Consider the second example in Section 5.2,

$$A = \begin{pmatrix} -2 & 2 & -3 \\ 2 & 1 & -6 \\ -1 & -2 & 0 \end{pmatrix}$$

We found three linear eigenvectors

$$\tilde{\mathbf{e}_1} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \tilde{\mathbf{e}_2} = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \tilde{\mathbf{e}_3} = \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}$$

If we let

$$P = \begin{pmatrix} 1 & -2 & 3 \\ 2 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, P^{-1} = \frac{1}{8}\begin{pmatrix} 1 & 2 & -3 \\ -2 & 4 & 6 \\ 1 & 2 & 5 \end{pmatrix},$$

then

$$\tilde{A} = P^{-1}AP = \begin{pmatrix} 5 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -3 \end{pmatrix},$$

so A is diagonalizable.

Theorem. Let $\lambda_1, \lambda_2, \cdots, \lambda_r$, with $r \leq n$ be the distinct eigenvalues of A. Let $B_1, B_2, \cdots B_r$ be the bases of the eigenspaces $E_{\lambda_1}, E_{\lambda_2}, \cdots, E_{\lambda_r}$ correspondingly. Then the set $B = \bigcup_{i=1}^{r} B_i$ is linearly independent.

This is similar to the proof we had for the case where the eigenvalues are distinct. However, we are going to do it much concisely, and the actual meat of the proof is actually just a single line.

Proof. Write $B_1 = \{\mathbf{x}_1^{(1)}, \mathbf{x}_2^{(1)}, \cdots \mathbf{x}_{m(\lambda_1)}^{(1)}\}$. Then $m(\lambda_1) = \dim(E_{\lambda_1})$, and similarly for all B_i.
Consider the following general linear combination of all elements in B. Consider the equation

$$\sum_{i=1}^{r}\sum_{j=1}^{m(\lambda_i)} \alpha_{ij}\mathbf{x}_j^{(i)} = 0.$$

The first sum is summing over all eigenspaces, and the second sum sums over the basis vectors in B_i. Now apply the matrix

$$\prod_{k=1,2,\cdots,\bar{K},\cdots,r}(A-\lambda_k I)$$

to the above sum, for some arbitrary K. We obtain

$$\sum_{j=1}^{m(\lambda_K)}\alpha_{Kj}\left[\prod_{k=1,2,\cdots,\bar{K},\cdots,r}(\lambda_K-\lambda_k)\right]\mathbf{x}_j^{(K)}=0.$$

Since the $\mathbf{x}_j^{(K)}$ are linearly independent (B_K is a basis), $\alpha_{Kj}=0$ for all j. Since K was arbitrary, all α_{ij} must be zero. So B is linearly independent. $\square$

Proposition. A is diagonalizable iff all its eigenvalues have zero defect.

Canonical (Jordan normal) form

Given a matrix A, if its eigenvalues all have non-zero defect, then we can find a basis in which it is diagonal. However, if some eigenvalue ***does*** have defect, we can still put it into an almost-diagonal form. This is known as the ***Jordan normal form***.

Theorem. Any 2×2 complex matrix A is similar to exactly one of

$$\begin{pmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{pmatrix},\begin{pmatrix}\lambda & 0\\ 0 & \lambda\end{pmatrix},\begin{pmatrix}\lambda & 1\\ 0 & \lambda\end{pmatrix}$$

Proof. For each case:

(i) If A has two distinct eigenvalues, then eigenvectors are linearly independent. Then we can use P formed from eigenvectors as its columns

(ii) If $\lambda_1=\lambda_2=\lambda$ and $\dim E_\lambda=2$, then write $E_\lambda=\operatorname{span}\{\mathbf{u},\mathbf{v}\}$, with $\mathbf{u},\mathbf{v}$ linearly independent. Now use $\{\mathbf{u},\mathbf{v}\}$ as a new basis of $\mathbb{C}^2$ and $\tilde{A}=P^{-1}AP=\begin{pmatrix}\lambda & 0\\ 0 & \lambda\end{pmatrix}=\lambda I$

Note that since $P^{-1}AP=\lambda I$, we have $A=P(\lambda I)P^{-1}=\lambda I$. So A is ***isotropic***, i.e. the same with respect to any basis.

(iii) If $\lambda_1=\lambda_2=\lambda$ and $\dim(E_\lambda)=1$, then $E_\lambda=\operatorname{span}\{\mathbf{v}\}$. Now choose basis of $\mathbb{C}^2$ as $\{\mathbf{v},\mathbf{w}\}$, where $\mathbf{w}\in\mathbb{C}^2\setminus E_\lambda$.

We know that $A\mathbf{w}\in\mathbb{C}^2$. So $A\mathbf{w}=\alpha\mathbf{v}+\beta\mathbf{w}$. Hence, if we change basis to $\{\mathbf{v},\mathbf{w}\}$, then $\tilde{A}=P^{-1}AP=\begin{pmatrix}\lambda & \alpha\\ 0 & \beta\end{pmatrix}$.

However, A and $\tilde{A}$ both have eigenvalue λ with algebraic multiplicity 2. So we must have $\beta=\lambda$. To make $\alpha=1$, let $\mathbf{u}=(\tilde{A}-\lambda I)\mathbf{w}$. We know $\mathbf{u}\neq\mathbf{0}$ since $\mathbf{w}$ is not in the eigenspace. Then

$$(\tilde{A}-\lambda I)\mathbf{u}=(\tilde{A}-\lambda I)^2\mathbf{w}=\begin{pmatrix}0 & \alpha\\ 0 & 0\end{pmatrix}\begin{pmatrix}0 & \alpha\\ 0 & 0\end{pmatrix}\mathbf{w}=\mathbf{0}.$$

So $\mathbf{u}$ is an eigenvector of $\tilde{A}$ with eigenvalue λ.

We have $\mathbf{u} = \tilde{A}\mathbf{w} - \lambda\mathbf{w}$. So $\tilde{A}\mathbf{w} = \mathbf{u} + \lambda\mathbf{w}$.

Change basis to $\{\mathbf{u}, \mathbf{w}\}$. Then A with respect to this basis is $\begin{pmatrix}\lambda & 1\\ 0 & \lambda\end{pmatrix}$.

This is a two-stage process: P sends basis to $\{\mathbf{v}, \mathbf{w}\}$ and then matrix Q sends to basis $\{\mathbf{u}, \mathbf{w}\}$. So the similarity transformation is $Q^{-1}(P^{-1}AP)Q = (PQ)^{-1}A(PQ)$. □

Proposition. (Without proof) The canonical form, or Jordan normal form, exists for any $n \times n$ matrix A. Specifically, there exists a similarity transform such that A is similar to a matrix to $\tilde{A}$ that satisfies the following properties:

(i) $\tilde{A}_{\alpha\alpha} = \lambda_\alpha$, i.e. the diagonal composes of the eigenvalues.

(ii) $\tilde{A}_{\alpha,\alpha+1} = 0$ or 1.

(iii) $\tilde{A}_{ij} = 0$ otherwise.

The actual theorem is actually stronger than this, and the Jordan normal form satisfies some additional properties in addition to the above. However, we shall not go into details, and this is left for the IB Linear Algebra course.

Example. Let

$$A = \begin{pmatrix}-3 & -1 & 1\\ -1 & -3 & 1\\ -2 & -2 & 0\end{pmatrix}$$

The eigenvalues are $-2, -2, -2$ and the eigenvectors are $\begin{pmatrix}-1\\ 1\\ 0\end{pmatrix}, \begin{pmatrix}1\\ 0\\ 1\end{pmatrix}$. Pick $\mathbf{w} = \begin{pmatrix}1\\ 0\\ 0\end{pmatrix}$.

Write $\mathbf{u} = (A - \lambda I)\mathbf{w} = \begin{pmatrix}-1 & -1 & 1\\ -1 & -1 & 1\\ -2 & -2 & 2\end{pmatrix}\begin{pmatrix}1\\ 0\\ 0\end{pmatrix} = \begin{pmatrix}-1\\ -1\\ -2\end{pmatrix}$. Note that $A\mathbf{u} = -2\mathbf{u}$. We also have $A\mathbf{w} = \mathbf{u} - 2\mathbf{w}$. Form a basis $\{\mathbf{u}, \mathbf{w}, \mathbf{v}\}$, where $\mathbf{v}$ is another eigenvector linearly independent from $\mathbf{u}$, say $\begin{pmatrix}1\\ 0\\ 1\end{pmatrix}$.

Now change to this basis with $P = \begin{pmatrix}-1 & 1 & 1\\ -1 & 0 & 0\\ -2 & 0 & 1\end{pmatrix}$. Then the Jordan normal form is

$$P^{-1}AP = \begin{pmatrix}-2 & 1 & 0\\ 0 & -2 & 0\\ 0 & 0 & -2\end{pmatrix}$$

Cayley-Hamilton Theorem

Theorem (Cayley-Hamilton theorem). Every $n \times n$ complex matrix satisfies its own characteristic equation.

Proof. We will only prove for diagonalizable matrices here. So suppose for our matrix A, there is some P such that $D = \operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) = P^{-1}AP$. Note that

$$D^i = (P^{-1}AP)(P^{-1}AP)\cdots(P^{-1}AP) = P^{-1}A^iP.$$

Hence

$$p_D(D) = p_D(P^{-1}AP) = P^{-1}[p_D(A)]P.$$

Since similar matrices have the same characteristic polynomial. So

$$p_A(D) = P^{-1}[p_A(A)]P.$$

However, we also know that $D^i = \operatorname{diag}(\lambda_1^i, \lambda_2^i, \cdots \lambda_n^i)$. So

$$p_A(D) = \operatorname{diag}(p_A(\lambda_1), p_A(\lambda_2), \cdots, p_A(\lambda_n)) = \operatorname{diag}(0, 0, \cdots, 0)$$

since the eigenvalues are roots of $p_A(\lambda) = 0$. So $0 = p_A(D) = P^{-1}p_A(A)P$ and thus $p_A(A) = 0$. □

There are a few things to note.

(i) If A^{-1} exists, then $A^{-1}p_A(A) = A^{-1}(c_0 + c_1A + c_2A^2 + \cdots + c_nA^n) = 0$. So $c_0A^{-1} + c_1 + c_2A + \cdots + c_nA^{n-1}$. Since A^{-1} exists, $c_0 = \pm\det A \neq 0$. So

$$A^{-1} = \frac{-1}{c_0}(c_1 + c_2A + \cdots + c_nA^{n-1}).$$

So we can calculate A^{-1} from positive powers of A.

(ii) We can define matrix exponentiation by

$$e^A = I + A + \frac{1}{2!}A^2 + \cdots + \frac{1}{n!}A^n + \cdots.$$

It is a fact that this always converges.

If A is diagonalizable with P with $D = P^{-1}AP = \operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$, then

$$\begin{aligned}
P^{-1}e^AP &= P^{-1}IP + P^{-1}AP + \frac{1}{2!}P^{-1}A^2P + \cdots \\
&= I + D + \frac{1}{2!}D^2 + \cdots \\
&= \operatorname{diag}(e^{\lambda_1}, e^{\lambda_2}, \cdots e^{\lambda_n})
\end{aligned}$$

So

$$e^A = P[\operatorname{diag}(e^{\lambda_1}, e^{\lambda_2}, \cdots, e^{\lambda_n})]P^{-1}.$$

(iii) For 2×2 matrices which are similar to $B = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$ We see that the characteristic polynomial $p_B(z) = \det(B - zI) = (\lambda - z)^2$. Then $p_B(B) = (\lambda I - B)^2 = \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix}^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$.

Since we have proved for the diagonalizable matrices above, we now know that ***any*** 2×2 matrix satisfies Cayley-Hamilton theorem.

In IB Linear Algebra, we will prove the Cayley Hamilton theorem properly for all matrices without assuming diagonalizability.

Eigenvalues and eigenvectors of a Hermitian matrix

Eigenvalues and eigenvectors

Theorem. The eigenvalues of a Hermitian matrix H are real.

Proof. Suppose that H has eigenvalue λ with eigenvector $\mathbf{v} \neq 0$. Then

$$H\mathbf{v} = \lambda\mathbf{v}.$$

We pre-multiply by $\mathbf{v}^\dagger$, a $1 \times n$ row vector, to obtain

$$\mathbf{v}^\dagger H\mathbf{v} = \lambda\mathbf{v}^\dagger\mathbf{v} \tag{$*$}$$

We take the Hermitian conjugate of both sides. The left hand side is

$$(\mathbf{v}^\dagger H\mathbf{v})^\dagger = \mathbf{v}^\dagger H^\dagger\mathbf{v} = \mathbf{v}^\dagger H\mathbf{v}$$

since H is Hermitian. The right hand side is

$$(\lambda\mathbf{v}^\dagger\mathbf{v})^\dagger = \lambda^*\mathbf{v}^\dagger\mathbf{v}$$

So we have

$$\mathbf{v}^\dagger H\mathbf{v} = \lambda^*\mathbf{v}^\dagger\mathbf{v}.$$

From $(*)$, we know that $\lambda\mathbf{v}^\dagger\mathbf{v} = \lambda^*\mathbf{v}^\dagger\mathbf{v}$. Since $\mathbf{v} \neq 0$, we know that $\mathbf{v}^\dagger\mathbf{v} = \mathbf{v}\cdot\mathbf{v} \neq 0$. So $\lambda = \lambda^*$ and λ is real. □

Theorem. The eigenvectors of a Hermitian matrix H corresponding to distinct eigenvalues are orthogonal.

Proof. Let

$$H\mathbf{v}_i = \lambda_i\mathbf{v}_i \tag{i}$$
$$H\mathbf{v}_j = \lambda_j\mathbf{v}_j. \tag{ii}$$

Pre-multiply (i) by $\mathbf{v}_j^\dagger$ to obtain

$$\mathbf{v}_j^\dagger H\mathbf{v}_i = \lambda_i\mathbf{v}_j^\dagger\mathbf{v}_i. \tag{iii}$$

Pre-multiply (ii) by $\mathbf{v}_i^\dagger$ and take the Hermitian conjugate to obtain

$$\mathbf{v}_j^\dagger H\mathbf{v}_i = \lambda_j\mathbf{v}_j^\dagger\mathbf{v}_i. \tag{iv}$$

Equating (iii) and (iv) yields

$$\lambda_i\mathbf{v}_j^\dagger\mathbf{v}_i = \lambda_j\mathbf{v}_j^\dagger\mathbf{v}_i.$$

Since $\lambda_i \neq \lambda_j$, we must have $\mathbf{v}_j^\dagger\mathbf{v}_i = 0$. So their inner product is zero and are orthogonal. □

So we know that if a Hermitian matrix has n distinct eigenvalues, then the eigenvectors form an orthonormal basis. However, if there are degenerate eigenvalues, it is more difficult, and requires the Gram-Schmidt process.

Gram-Schmidt orthogonalization (non-examinable)

Suppose we have a set $B = \{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_r\}$ of linearly independent vectors. We want to find an orthogonal set $\tilde{\mathbf{B}} = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r\}$.
Define the projection of $\mathbf{w}$ onto $\mathbf{v}$ by $\mathcal{P}_{\mathbf{v}}(\mathbf{w}) = \frac{\langle \mathbf{v}|\mathbf{w}\rangle}{\langle \mathbf{v}|\mathbf{v}\rangle}\mathbf{v}$. Now construct $\tilde{\mathbf{B}}$ iteratively:

(i) $\mathbf{v}_1 = \mathbf{w}_1$

(ii) $\mathbf{v}_2 = \mathbf{w}_2 - \mathcal{P}_{\mathbf{v}_1}(\mathbf{w})$

Then we get that $\langle \mathbf{v}_1 \mid \mathbf{v}_2\rangle = \langle \mathbf{v}_1 \mid \mathbf{w}_2\rangle - \left(\frac{\langle \mathbf{v}_1|\mathbf{w}_2\rangle}{\langle \mathbf{v}_1|\mathbf{v}_1\rangle}\right)\langle \mathbf{v}_1 \mid \mathbf{v}_1\rangle = 0$

(iii) $\mathbf{v}_3 = \mathbf{w}_3 - \mathcal{P}_{\mathbf{v}_1}(\mathbf{w}_3) - \mathcal{P}_{\mathbf{v}_2}(\mathbf{w}_3)$

(iv) $\vdots$

(v) $\mathbf{v}_r = \mathbf{w}_r - \sum_{j=1}^{r-1} \mathcal{P}_{\mathbf{v}_j}(\mathbf{w}_r)$

At each step, we subtract out the components of $\mathbf{v}_i$ that belong to the space of $\{\mathbf{v}_1, \cdots, \mathbf{v}_{k-1}\}$. This ensures that all the vectors are orthogonal. Finally, we normalize each basis vector individually to obtain an orthonormal basis.

Unitary transformation

Suppose U is the transformation between one orthonormal basis and a new orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$, i.e. $\langle \mathbf{u}_i \mid \mathbf{u}_j\rangle = \delta_{ij}$. Then

$$U = \begin{pmatrix} (\mathbf{u}_1)_1 & (\mathbf{u}_2)_1 & \cdots & (\mathbf{u}_n)_1 \\ (\mathbf{u}_1)_2 & (\mathbf{u}_2)_2 & \cdots & (\mathbf{u}_n)_2 \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{u}_1)_n & (\mathbf{u}_2)_n & \cdots & (\mathbf{u}_n)_n \end{pmatrix}$$

Then

$$\begin{aligned} (U^\dagger U)_{ij} &= (U^\dagger)_{ik} U_{kj} \\ &= U^*_{ki} U_{kj} \\ &= (\mathbf{u}_i)^*_k (\mathbf{u}_j)_k \\ &= \langle \mathbf{u}_i \mid \mathbf{u}_j\rangle \\ &= \delta_{ij} \end{aligned}$$

So U is a unitary matrix.

Diagonalization of $n \times n$ Hermitian matrices

Theorem. An $n \times n$ Hermitian matrix has precisely n orthogonal eigenvectors.

Proof. (Non-examinable) Let $\lambda_1, \lambda_2, \cdots, \lambda_r$ be the distinct eigenvalues of H ($r \leq n$), with a set of corresponding orthonormal eigenvectors $B = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r\}$. Extend to a basis of the whole of $\mathbb{C}^n$

$$B' = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r, \mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_{n-r}\}$$

Now use Gram-Schmidt to create an orthonormal basis

$$\tilde{B} = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r, \mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_{n-r}\}.$$

Now write

$$P = \begin{pmatrix} \uparrow & \uparrow & & \uparrow & \uparrow & & \uparrow \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_r & \mathbf{u}_1 & \cdots & \mathbf{u}_{n-r} \\ \downarrow & \downarrow & & \downarrow & \downarrow & & \downarrow \end{pmatrix}$$

We have shown above that this is a unitary matrix, i.e. $P^{-1} = P^\dagger$. So if we change basis, we have

$$\begin{aligned} P^{-1}HP &= P^\dagger HP \\ &= \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_r & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & c_{11} & c_{12} & \cdots & c_{1,n-r} \\ 0 & 0 & \cdots & 0 & c_{21} & c_{22} & \cdots & c_{2,n-r} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & c_{n-r,1} & c_{n-r,2} & \cdots & c_{n-r,n-r} \end{pmatrix} \end{aligned}$$

Here C is an $(n-r) \times (n-r)$ Hermitian matrix. The eigenvalues of C are also eigenvalues of H because $\det(H - \lambda I) = \det(P^\dagger HP - \lambda I) = (\lambda_1 - \lambda) \cdots (\lambda_r - \lambda) \det(C - \lambda I)$. So the eigenvalues of C are the eigenvalues of H.

We can keep repeating the process on C until we finish all rows. For example, if the eigenvalues of C are all distinct, there are $n - r$ orthonormal eigenvectors $\mathbf{w}_j$ (for $j = r+1, \cdots, n$) of C. Let

$$Q = \begin{pmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & \ddots & & & & \\ & & & 1 & & & \\ & & & & \uparrow & \uparrow & & \uparrow \\ & & & & \mathbf{w}_{r+1} & \mathbf{w}_{r+2} & \cdots & \mathbf{w}_n \\ & & & & \downarrow & \downarrow & & \downarrow \end{pmatrix}$$

with other entries 0. (where we have a $r \times r$ identity matrix block on the top left corner and a $(n-r) \times (n-r)$ with columns formed by $\mathbf{w}_j$)

Since the columns of Q are orthonormal, Q is unitary. So $Q^\dagger P^\dagger HPQ = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_r, \lambda_{r+1}, \cdots, \lambda_n)$, where the first r λs are distinct and the remaining ones are copies of previous ones.
The n linearly-independent eigenvectors are the columns of PQ.

□

So it now follows that H is diagonalizable via transformation $U(= PQ)$. U is a unitary matrix because P and Q are. We have

$$D = U^\dagger HU$$
$$H = UDU^\dagger$$

Note that a real symmetric matrix S is a special case of Hermitian matrices. So we have

$$D = Q^T SQ$$
$$S = QDQ^T$$

Example. Find the orthogonal matrix which diagonalizes the following real symmetric matrix: $S = \begin{pmatrix} 1 & \beta \\ \beta & 1 \end{pmatrix}$ with $\beta \neq 0 \in \mathbb{R}$.
We find the eigenvalues by solving the characteristic equation: $\det(S - \lambda I) = 0$, and obtain $\lambda = 1 \pm \beta$.
The corresponding eigenvectors satisfy $(S - \lambda I)\mathbf{x} = 0$, which gives $\mathbf{x} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm 1 \end{pmatrix}$
We change the basis from the standard basis to $\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ (which is just a rotation by $\pi/4$).
The transformation matrix is $Q = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$. Then we know that $S = QDQ^T$ with $D = \mathrm{diag}(1, -1)$

Normal matrices

We have seen that the eigenvalues and eigenvectors of Hermitian matrices satisfy some nice properties. More generally, we can define the following:

Definition (Normal matrix)**.** A ***normal matrix*** as a matrix that commutes with its own Hermitian conjugate, i.e.
$$NN^\dagger = N^\dagger N$$

Hermitian, real symmetric, skew-Hermitian, real anti-symmetric, orthogonal, unitary matrices are all special cases of normal matrices.
It can be shown that:

Proposition.

(i) If λ is an eigenvalue of N, then λ^* is an eigenvalue of $N^\dagger$.

(ii) The eigenvectors of distinct eigenvalues are orthogonal.

(iii) A normal matrix can always be diagonalized with an orthonormal basis of eigenvectors.

1.7 Quadratic forms and conics

We want to study quantities like $x_1^2 + x_2^2$ and $3x_1^2 + 2x_1x_2 + 4x_2^2$. For example, conic sections generally take this form. The common characteristic of these is that each term has degree 2. Consequently, we can write it in the form $\mathbf{x}^\dagger A\mathbf{x}$ for some matrix A.

Definition (Sesquilinear, Hermitian and quadratic forms)**.** A ***sesquilinear form*** is a quantity $F = \mathbf{x}^\dagger A\mathbf{x} = x_i^* A_{ij} x_j$. If A is Hermitian, then F is a ***Hermitian form***. If A is real symmetric, then F is a ***quadratic form***.

Theorem. Hermitian forms are real.

Proof. $(\mathbf{x}^\dagger H\mathbf{x})^* = (\mathbf{x}^\dagger H\mathbf{x})^\dagger = \mathbf{x}^\dagger H^\dagger\mathbf{x} = \mathbf{x}^\dagger H\mathbf{x}$. So $(\mathbf{x}^\dagger H\mathbf{x})^* = \mathbf{x}^\dagger H\mathbf{x}$ and it is real. □

We know that any Hermitian matrix can be diagonalized with a unitary transformation. So $F(\mathbf{x}) = \mathbf{x}^\dagger H\mathbf{x} = \mathbf{x}^\dagger UDU^\dagger\mathbf{x}$. Write $\mathbf{x}' = U^\dagger\mathbf{x}$. So $F = (\mathbf{x}')^\dagger D\mathbf{x}'$, where $D = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$.
We know that $\mathbf{x}'$ is the vector $\mathbf{x}$ relative to the eigenvector basis. So

$$F(\mathbf{x}) = \sum_{i=1}^n \lambda_i |x_i'|^2$$

The eigenvectors are known as the principal axes.

Example. Take $F = 2x^2 - 4xy + 5y^2 = \mathbf{x}^T S\mathbf{x}$, where $\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$ and $S = \begin{pmatrix} 2 & -2 \\ -2 & 5 \end{pmatrix}$.
Note that we can always choose the matrix to be symmetric. This is since for any antisymmetric A, we have $\mathbf{x}^\dagger A\mathbf{x} = 0$. So we can just take the symmetric part.
The eigenvalues are $1, 6$ with corresponding eigenvectors $\frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ -2 \end{pmatrix}$. Now change basis with

$$Q = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix}$$

Then $\mathbf{x}' = Q^T\mathbf{x} = \frac{1}{\sqrt{5}}\begin{pmatrix} 2x + y \\ x - 2y \end{pmatrix}$. Then $F = (x')^2 + 6(y')^2$.
So $F = c$ is an ellipse.

Quadrics and conics

Quadrics

Definition (Quadric)**.** A ***quadric*** is an n-dimensional surface defined by the zero of a real quadratic polynomial, i.e.

$$\mathbf{x}^T A\mathbf{x} + \mathbf{b}^T\mathbf{x} + c = 0,$$

where A is a real $n \times n$ matrix, $\mathbf{x}, \mathbf{b}$ are n-dimensional column vectors and c is a constant scalar.

As noted in example, anti-symmetric matrix has $\mathbf{x}^T A\mathbf{x} = 0$, so for any A, we can split it into symmetric and anti-symmetric parts, and just retain the symmetric part $S = (A + A^T)/2$. So we can have

$$\mathbf{x}^T S\mathbf{x} + \mathbf{b}^T\mathbf{x} + c = 0$$

with S symmetric.

Since S is real and symmetric, we can diagonalize it using $S = QDQ^T$ with D diagonal. We write $\mathbf{x}' = Q^T\mathbf{x}$ and $\mathbf{b}' = Q^T\mathbf{b}$. So we have

$$(\mathbf{x}')^T D\mathbf{x}' + (\mathbf{b}')^T\mathbf{x}' + c = 0.$$

If S is invertible, i.e. with no zero eigenvalues, then write $\mathbf{x}'' = \mathbf{x}' + \frac{1}{2}D^{-1}\mathbf{b}'$ which shifts the origin to eliminate the linear term $(\mathbf{b}')^T\mathbf{x}'$ and finally have (dropping the prime superfixes)

$$\mathbf{x}^T D\mathbf{x} = k.$$

So through two transformations, we have ended up with a simple quadratic form.

Conic sections $(n = 2)$

From the equation above, we obtain

$$\lambda_1 x_1^2 + \lambda_2 x_2^2 = k.$$

We have the following cases:

(i) $\lambda_1\lambda_2 > 0$: we have ellipses with axes coinciding with eigenvectors of S. (We require $\operatorname{sgn}(k) = \operatorname{sgn}(\lambda_1, \lambda_2)$, or else we would have no solutions at all)

(ii) $\lambda_1\lambda_2 < 0$: say $\lambda_1 = k/a^2 > 0$, $\lambda_2 = -k/b^2 < 0$. So we obtain

$$\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 1,$$

which is a hyperbola.

(iii) $\lambda_1\lambda_2 = 0$: Say $\lambda_2 = 0$, $\lambda_1 \neq 0$. Note that in this case, our symmetric matrix S is not invertible and we cannot shift our origin using as above.

From our initial equation, we have

$$\lambda_1(x_1')^2 + b_1'x_1' + b_2'x_2' + c = 0.$$

We perform the coordinate transform (which is simply completing the square!)

$$x_1'' = x_1' + \frac{b_1'}{2\lambda_1}$$
$$x_2'' = x_2' + \frac{c}{b_2'} - \frac{(b_1')^2}{4\lambda_1 b_2'}$$

to remove the x_1' and constant term. Dropping the primes, we have

$$\lambda_1 x_1^2 + b_2 x_2 = 0,$$

which is a parabola.

Note that above we assumed $b_2' \neq 0$. If $b_2' = 0$, we have $\lambda_1 (x_1')^2 + b_1' x_1' + c = 0$. If we solve this quadratic for x_1', we obtain 0, 1 or 2 solutions for x_1 (and x_2 can be any value). So we have 0, 1 or 2 straight lines.

These are known as conic sections. As you will see in IA Dynamics and Relativity, this are the trajectories of planets under the influence of gravity.

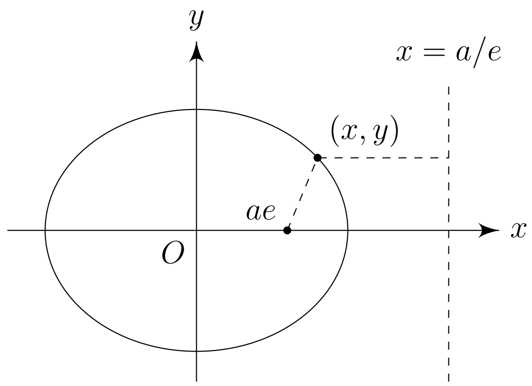
Focus-directrix property

Conic sections can be defined in a different way, in terms of

Definition (Conic sections)**.** The ***eccentricity*** and ***scale*** are properties of a conic section that satisfy the following:
Let the ***foci*** of a conic section be $(\pm ae, 0)$ and the ***directrices*** be $x = \pm a/e$.
A ***conic section*** is the set of points whose distance from focus is $e\times$ distance from directrix which is closer to that of focus (unless $e = 1$, where we take the distance to the other directrix).

Now consider the different cases of e:

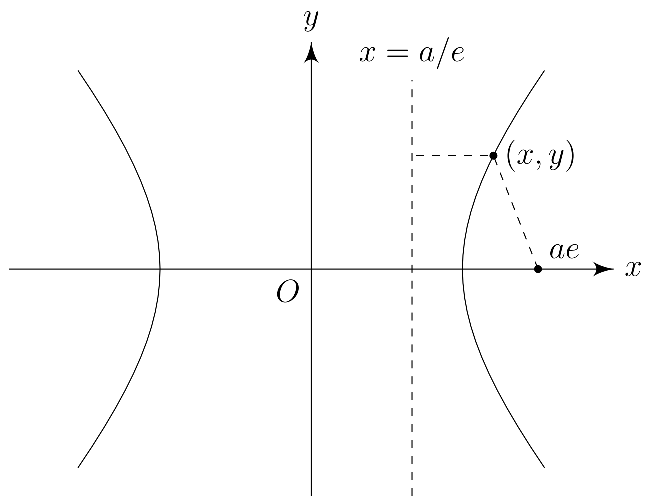
(i) $e < 1$. By definition,



$$\sqrt{(x - ae)^2 + y^2} = e\left(\frac{a}{e} - x\right)$$
$$\frac{x^2}{a^2} + \frac{y^2}{a^2(1 - e^2)} = 1$$

Which is an ellipse with semi-major axis a and semi-minor axis $a\sqrt{1 - e^2}$. (if $e = 0$, then we have a circle)

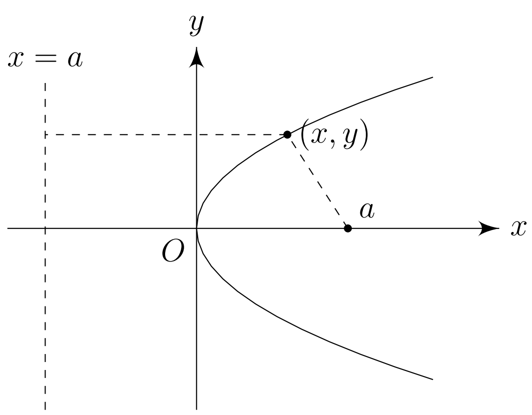
(ii) $e > 1$. So



$$\sqrt{(x-ae)^2+y^2} = e\left(x - \frac{a}{e}\right)$$
$$\frac{x^2}{a^2} - \frac{y^2}{a^2(e^2-1)} = 1$$

and we have a hyperbola.

(iii) $e = 1$: Then



$$\sqrt{(x-a)^2+y^2} = (x+1)$$
$$y^2 = 4ax$$

and we have a parabola.

Conics also work in polar coordinates. We introduce a new parameter l such that l/e is the distance from the focus to the directrix. So

$$l = a|1 - e^2|.$$

We use polar coordinates (r, θ) centered on a focus. So the focus-directrix property is

$$r = e\left(\frac{l}{e} - r\cos\theta\right)$$
$$r = \frac{l}{1 + e\cos\theta}$$

We see that $r \to \infty$ if $\theta \to \cos^{-1}(-1/e)$, which is only possible if $e \geq 1$, i.e. hyperbola or parabola. But ellipses have $e < 1$. So r is bounded, as expected.

1.8 Transformation groups

We have previously seen that orthogonal matrices are used to transform between orthonormal bases. Alternatively, we can see them as transformations of space itself that preserve distances, which is something we will prove shortly.

Using this as the definition of an orthogonal matrix, we see that our definition of orthogonal matrices is dependent on our choice of the notion of distance, or metric. In special relativity, we will need to use a different metric, which will lead to the ***Lorentz matrices***, the matrices that conserve distances in special relativity. We will have a brief look at these as well.

Groups of orthogonal matrices

Proposition. The set of all $n \times n$ orthogonal matrices P forms a group under matrix multiplication.

Proof.

0. If P, Q are orthogonal, then consider $R = PQ$. $RR^T = (PQ)(PQ)^T = P(QQ^T)P^T = PP^T = I$. So R is orthogonal.
1. I satisfies $II^T = I$. So I is orthogonal and is an identity of the group.
2. Inverse: if P is orthogonal, then $P^{-1} = P^T$ by definition, which is also orthogonal.
3. Matrix multiplication is associative since function composition is associative. □

Definition (Orthogonal group)**.** The ***orthogonal group*** $O(n)$ is the group of orthogonal matrices.

Definition (Special orthogonal group)**.** The ***special orthogonal group*** is the subgroup of $O(n)$ that consists of all orthogonal matrices with determinant 1.

In general, we can show that any matrix in $O(2)$ is of the form

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \text{ or } \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$$

Length preserving matrices

Theorem. Let $P \in O(n)$. Then the following are equivalent:

(i) P is orthogonal

(ii) $|P\mathbf{x}| = |\mathbf{x}|$

(iii) $(P\mathbf{x})^T(P\mathbf{y}) = \mathbf{x}^T\mathbf{y}$, i.e. $(P\mathbf{x}) \cdot (P\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$.

(iv) If $(\mathbf{v}_1, \mathbf{v_2}, \cdots, \mathbf{v}_n)$ are orthonormal, so are $(P\mathbf{v}_1, P\mathbf{v}_2, \cdots, P\mathbf{v}_n)$

(v) The columns of P are orthonormal.

Proof. We do them one by one:

(i) $\Rightarrow$ (ii): $|P\mathbf{x}|^2 = (P\mathbf{x})^T(P\mathbf{x}) = \mathbf{x}^T P^T P\mathbf{x} = \mathbf{x}^T\mathbf{x} = |\mathbf{x}|^2$

(ii) $\Rightarrow$ (iii): $|P(\mathbf{x}+\mathbf{y})|^2 = |\mathbf{x}+\mathbf{y}|^2$. The right hand side is

$$(\mathbf{x}^T + \mathbf{y}^T)(\mathbf{x}+\mathbf{y}) = \mathbf{x}^T\mathbf{x} + y^T\mathbf{y} + \mathbf{y}^T\mathbf{x} + \mathbf{x}^T\mathbf{y} = |\mathbf{x}|^2 + |\mathbf{y}|^2 + 2\mathbf{x}^T\mathbf{y}.$$

Similarly, the left hand side is

$$|P\mathbf{x} + P\mathbf{y}|^2 = |P\mathbf{x}|^2 + |P\mathbf{y}| + 2(P\mathbf{x})^T P\mathbf{y} = |\mathbf{x}|^2 + |\mathbf{y}|^2 + 2(P\mathbf{x})^T P\mathbf{y}.$$

So $(P\mathbf{x})^T P\mathbf{y} = \mathbf{x}^T\mathbf{y}$.

(iii) $\Rightarrow$ (iv): $(P\mathbf{v}_i)^T P\mathbf{v}_j = \mathbf{v}_i^T\mathbf{v}_j = \delta_{ij}$. So $P\mathbf{v}_i$'s are also orthonormal.

(iv) $\Rightarrow$ (v): Take the $\mathbf{v}_i$'s to be the standard basis. So the columns of P, being $P\mathbf{e}_i$, are orthonormal.

(v) $\Rightarrow$ (i): The columns of P are orthonormal. Then $(PP^T)_{ij} = P_{ik}P_{jk} = (P_i)\cdot(P_j) = \delta_{ij}$, viewing P_i as the ith column of P. So $PP^T = I$. □

Therefore the set of length-preserving matrices is precisely $O(n)$.

Lorentz transformations

Consider the ***Minkowski*** 1 + 1 dimension spacetime (i.e. 1 space dimension and 1 time dimension)

Definition (Minkowski inner product)**.** The ***Minkowski*** inner product of 2 vectors $\mathbf{x}$ and $\mathbf{y}$ is

$$\langle \mathbf{x} \mid \mathbf{y}\rangle = \mathbf{x}^T J\mathbf{y},$$

where

$$J = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Then $\langle \mathbf{x} \mid \mathbf{y}\rangle = x_1y_1 - x_2y_2$.

This is to be compared to the usual ***Euclidean*** inner product of $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$, given by

$$\langle \mathbf{x} \mid \mathbf{y}\rangle = \mathbf{x}^T\mathbf{y} = \mathbf{x}^T I\mathbf{y} = x_1y_1 + x_2y_2.$$

Definition (Preservation of inner product)**.** A transformation matrix M preserves the Minkowski inner product if

$$\langle \mathbf{x}|\mathbf{y}\rangle = \langle M\mathbf{x}|M\mathbf{y}\rangle$$

for all $\mathbf{x}, \mathbf{y}$.

We know that $\mathbf{x}^T J\mathbf{y} = (M\mathbf{x})^T JM\mathbf{y} = \mathbf{x}^T M^T JM\mathbf{y}$. Since this has to be true for all $\mathbf{x}$ and $\mathbf{y}$, we must have

$$J = M^T JM.$$

We can show that M takes the form of

$$H_\alpha = \begin{pmatrix} \cosh\alpha & \sinh\alpha \\ \sinh\alpha & \cosh\alpha \end{pmatrix} \text{ or } K_{\alpha/2} = \begin{pmatrix} \cosh\alpha & -\sinh\alpha \\ \sinh\alpha & -\cosh\alpha \end{pmatrix}$$

where H_α is a ***hyperbolic rotation***, and $K_{\alpha/2}$ is a ***hyperbolic reflection***.

This is technically ***all*** matrices that preserve the metric, since these only include matrices with $M_{11} > 0$. In physics, these are the matrices we want, since $M_{11} < 0$ corresponds to inverting time, which is frowned upon.

Definition (Lorentz matrix)**.** A ***Lorentz matrix*** or a ***Lorentz boost*** is a matrix in the form

$$B_v = \frac{1}{\sqrt{1-v^2}} \begin{pmatrix} 1 & v \\ v & 1 \end{pmatrix}.$$

Here $|v| < 1$, where we have chosen units in which the speed of light is equal to 1. We have $B_v = H_{\tanh^{-1} v}$

Definition (Lorentz group)**.** The ***Lorentz group*** is a group of all Lorentz matrices under matrix multiplication.

It is easy to prove that this is a group. For the closure axiom, we have $B_{v_1}B_{v_2} = B_{v_3}$, where

$$v_3 = \tanh(\tanh^{-1} v_1 + \tanh^{-1} v_2) = \frac{v_1 + v_2}{1 + v_1 v_2}$$

The set of all B_v is a group of transformations which preserve the Minkowski inner product.

Chapter 2

Dynamics

2.1 Introduction

You've been lied to. You thought you applied for mathematics. And here you have a course on physics. No, this course is not created for students taking the "Maths with Physics" option. They don't have to take this course (don't ask why).

Ever since Newton invented calculus, mathematics is becoming more and more important in physics. Physicists seek to describe the universe in a few equations, and derive everyday (physical) phenomena as mathematical consequences of these equations.

In this course, we will start with Newton's laws of motion and use it to derive a lot of physical phenomena, including planetary orbits, centrifugal forces[1] and the motion of rotating bodies.

The important thing to note is that we can "prove" all these phenomena just under the assumption that Newton's laws are correct (plus the formulas for, say, the strength of the gravitational force). We are just doing mathematics here. We don't need to do any experiments to obtain the results (of course, we need experiments to verify that Newton's laws are indeed the equations that describe this universe).

However, it turns out that Newton was wrong. While his theories were accurate for most everyday phenomena, they weren't able to adequately describe electromagnetism. This lead to Einstein discovering ***special relativity***. Special relativity is also required to describe motion that is very fast. We will have a brief introduction to special relativity at the end of the course.

[1] Yes, they exist.

2.2 Newtonian dynamics of particles

Newton's equations describe the motion of a ***(point) particle***.

Definition (Particle)**.** A ***particle*** is an object of insignificant size, hence it can be regarded as a point. It has a ***mass*** $m > 0$, and an ***electric charge*** q.

Its position at time t is described by its ***position vector***, $\mathbf{r}(t)$ or $\mathbf{x}(t)$ with respect to an origin O.

Depending on context, different things can be considered as particles. We could consider an electron to be a point particle, even though it is more accurately described by the laws of quantum mechanics than those of Newtonian mechanics. If we are studying the orbit of planets, we can consider the Sun and the Earth to be particles.

An important property of a particle is that it has no ***internal structure***. It can be completely described by its position, momentum, mass and electric charge. For example, if we model the Earth as a particle, we will have to ignore its own rotation, temperature etc.

If we want to actually describe a rotating object, we usually consider it as a collection of point particles connected together, and apply Newton's law to the individual particles.

As mentioned above, the position of a particle is described by a position ***vector***. This requires us to pick an origin of the coordinate system, as well as an orientation of the axes. Each choice is known as a frame of reference.

Definition (Frame of reference)**.** A ***frame of reference*** is a choice of coordinate axes for $\mathbf{r}$.

We don't impose many restrictions on the choice of coordinate axes. They can be fixed, moving, rotating, or even accelerating.

Using the position vector $\mathbf{r}$, we can define various interesting quantities which describe the particle.

Definition (Velocity)**.** The ***velocity*** of the particle is

$$\mathbf{v} = \dot{\mathbf{r}} = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}.$$

Definition (Acceleration)**.** The ***acceleration*** of the particle is

$$\mathbf{a} = \dot{\mathbf{v}} = \ddot{\mathbf{r}} = \frac{\mathrm{d}^2\mathbf{r}}{dt^2}.$$

Definition (Momentum)**.** The ***momentum*** of a particle is

$$\mathbf{p} = m\mathbf{v} = m\dot{\mathbf{r}}.$$

m is the ***inertial mass*** of the particle, and measures its reluctance to accelerate, as described by Newton's second law.

Newton's laws of motion

We will first state Newton's three laws of motion, and then discuss them individually.

Law (Newton's First Law of Motion)**.** A body remains at rest, or moves uniformly in a straight line, unless acted on by a force. (This is in fact Galileo's Law of Inertia)

Law (Newton's Second Law of Motion)**.** The rate of change of momentum of a body is equal to the force acting on it (in both magnitude and direction).

Law (Newton's Third Law of Motion)**.** To every action there is an equal and opposite reaction: the forces of two bodies on each other are equal and in opposite directions.

The first law might seem redundant given the second if interpreted literally. According to the second law, if there is no force, then the momentum doesn't change. Hence the body remains at rest or moves uniformly in a straight line.

So why do we have the first law? Historically, it might be there to explicitly counter Aristotle's idea that objects naturally slow down to rest. However, some (modern) physicists give it an alternative interpretation:

Note that the first law isn't always true. Take yourself as a frame of reference. When you move around your room, things will seem like they are moving around (relative to you). When you sit down, they stop moving. However, in reality, they've always been sitting there still. On second thought, this is because you, the frame of reference, is accelerating, not the objects. The first law only holds in frames that are themselves not accelerating. We call these ***inertial frames***.

Definition (Inertial frames)**.** ***Inertial frames*** are frames of references in which the frames themselves are not accelerating. Newton's Laws only hold in inertial frames.

Then we can take the first law to assert that inertial frames exists. Even though the Earth itself is rotating and orbiting the sun, for most purposes, any fixed place on the Earth counts as an inertial frame.

Galilean transformations

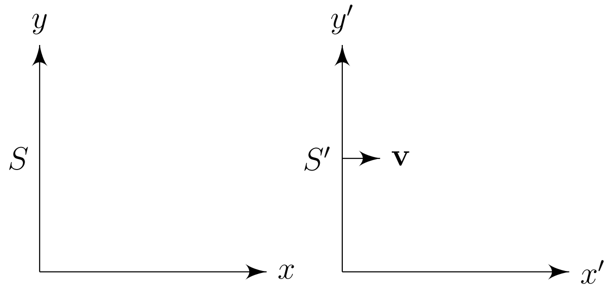
The goal of this section is to investigate inertial frames. We know that inertial frames are not unique. Given an inertial frame, what other inertial frames can we obtain?

First of all, we can rotate our axes or move our origin. In particular, we can perform the following operations:

- Translations of space:
$$\mathbf{r}' = \mathbf{r} - \mathbf{r}_0$$
- Translations of time:
$$t' = t - t_0$$
- Rotation (and reflection):
$$\mathbf{r}' = R\mathbf{r}$$
with $R \in \mathrm{O}(3)$.

These are not very interesting. They are simply symmetries of space itself.

The last possible transformation is ***uniform motion***. Suppose that S is an inertial frame. Then any other frame S' in uniform motion relative to S is also inertial:



Assuming the frames coincide at $t = 0$, then

$$\begin{aligned} x' &= x - vt \\ y' &= y \\ z' &= z \\ t' &= t \end{aligned}$$

Generally, the position vector transforms as

$$\mathbf{r}' = \mathbf{r} - \mathbf{v}t,$$

where $\mathbf{v}$ is the (constant) velocity of S' relative to S. The velocity and acceleration transform as follows:

$$\begin{aligned} \dot{\mathbf{r}}' &= \dot{\mathbf{r}} - \mathbf{v} \\ \ddot{\mathbf{r}}' &= \ddot{\mathbf{r}} \end{aligned}$$

Definition (Galilean boost)**.** A ***Galilean boost*** is a change in frame of reference by

$$\begin{aligned} \mathbf{r}' &= \mathbf{r} - \mathbf{v}t \\ t' &= t \end{aligned}$$

for a fixed, constant $\mathbf{v}$.

All these transformations together generate the ***Galilean group***, which describes the symmetry of Newtonian equations of motion.

Law (Galilean relativity)**.** The ***principle of relativity*** asserts that the laws of physics are the same in inertial frames.

This implies that physical processes work the same

- at every point of space
- at all times

- in whichever direction we face
- at whatever constant velocity we travel.

In other words, the equations of Newtonian physics must have ***Galilean invariance***.

Since the laws of physics are the same regardless of your velocity, velocity must be a ***relative concept***, and there is no such thing as an "absolute velocity" that all inertial frames agree on.

However, all inertial frames must agree on whether you are accelerating or not (even though they need not agree on the direction of acceleration since you can rotate your frame). So acceleration is an ***absolute*** concept.

Newton's Second Law

Newton's second law is often written in the form of an equation.

Law. The ***equation of motion*** for a particle subject to a force $\mathbf{F}$ is

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = \mathbf{F},$$

where $\mathbf{p} = m\mathbf{v} = m\dot{\mathbf{r}}$ is the (linear) momentum of the particle. We say m is the (inertial) mass of the particle, which is a measure of its reluctance to accelerate.

Usually, m is constant. Then

$$\mathbf{F} = m\mathbf{a} = m\ddot{\mathbf{r}}.$$

Usually, $\mathbf{F}$ is specified as a function of $\mathbf{r}, \dot{\mathbf{r}}$ and t. Then we have a second-order ordinary differential equation for $\mathbf{r}$.

To determine the solution, we need to specify the initial values of $\mathbf{r}$ and $\dot{\mathbf{r}}$, i.e. the initial position and velocity. The trajectory of the particle is then uniquely determined for all future (and past) times.

2.3 Dimensional analysis

When considering physical theories, it is important to be aware that physical quantities are not pure numbers. Each physical quantity has a ***dimension***. Roughly speaking, dimensions are what units represent, such as length, mass and time. In any equation relating physical quantities, the dimensions must be consistent, i.e. the dimensions on both sides of the equation must be equal.

For many problems in dynamics, the three basic dimensions are sufficient:

- length, L
- mass, M
- time, T

The dimensions of a physical quantity X, denoted by $[X]$ are expressible uniquely in terms of L, M and T. For example,

- $[\text{area}] = L^2$
- $[\text{density}] = L^{-3}M$
- $[\text{velocity}] = LT^{-1}$
- $[\text{acceleration}] = LT^{-2}$
- $[\text{force}] = LMT^{-2}$ since the dimensions must satisfy the equation $F = ma$.
- $[\text{energy}] = L^2MT^{-2}$, e.g. consider $E = mv^2/2$.

Physical constants also have dimensions, e.g. $[G] = L^3M^{-1}T^{-2}$ by $F = GMm/r^2$.

The only allowed operations on quantities with dimensions are sums and products (and subtraction and division), and if we sum two terms, they must have the same dimension. For example, it does not make sense to add a length with an area. More complicated functions of dimensional quantities are not allowed, e.g. e^x again makes no sense if x has a dimension, since

$$e^x = 1 + x + \frac{1}{2}x^2 + \cdots$$

and if x had a dimension, we would be summing up terms of different dimensions.

Units

People use ***units*** to represent dimensional quantities. A unit is a convenient standard physical quantity, e.g. a fixed amount of mass. In the SI system, there are base units corresponding to the basics dimensions. The three we need are

- meter (m) for length
- kilogram (kg) for mass

- second (s) for time

A physical quantity can be expressed as a pure number times a unit with the correct dimensions, e.g.

$$G = 6.67 \times 10^{-11}\,\mathrm{m}^3\,\mathrm{kg}^{-1}\,\mathrm{s}^{-2}.$$

It is important to realize that SI units are chosen arbitrarily for historical reasons only. The equation of physics must work in any consistent system of units. This is captured by the fact that physical equations must be dimensionally consistent.

Scaling

We've had so many restrictions on dimensional quantities — equations must be dimensionally consistent, and we cannot sum terms of different dimensions. However, this is ***not*** a hindrance when developing new theories. In fact, it is a very useful tool. First of all, it allows us to immediately rule out equations that do not make sense dimensionally. More importantly, it allows us to guess the form of the equation we want.

Suppose we believe that a physical quantity Y depends on 3 other physical quantities X_1, X_2, X_3, i.e. $Y = Y(X_1, X_2, X_3)$. Let their dimensions be as follows:

- $[Y] = L^a M^b T^c$
- $[X_i] = L^{a_i} M^{b_i} T^{c_i}$

Suppose further that we know that the relationship is a power law, i.e.

$$Y = C X_1^{p_1} X_2^{p_2} X_3^{p_3},$$

where C is a dimensionless constant (i.e. a pure number). Since the dimensions must work out, we know that

$$\begin{aligned} a &= p_1 a_1 + p_2 a_2 + p_3 a_3 \\ b &= p_1 b_1 + p_2 b_2 + p_3 b_3 \\ c &= p_1 c_1 + p_2 c_2 + p_3 c_3 \end{aligned}$$

for which there is a unique solution provided that the dimensions of X_1, X_2 and X_3 are independent. So just by using dimensional analysis, we can figure out the relation between the quantities up to a constant. The constant can then be found experimentally, which is much easier than finding the form of the expression experimentally.

However, in reality, the dimensions are not always independent. For example, we might have two length quantities. More importantly, the situation might involve more than 3 variables, and we do not have a unique solution.

First consider a simple case — if $X_1^2 X_2$ is dimensionless, then the relation between Y and X_i can involve more complicated terms such as $\exp(X_1^2 X_2)$, since the argument of exp is now dimensionless.

In general, suppose we have many terms, and the dimensions of X_i are not independent. We order the quantities so that the independent terms $[X_1], [X_2], [X_3]$ are at the front.

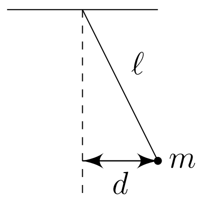
For each of the remaining variables, form a dimensionless quantity $\lambda_i = X_i X_1^{q_1} X_2^{q_2} X_3^{q_3}$. Then the relationship must be of the form

$$Y = f(\lambda_4, \lambda_5, \cdots) X_1^{p_1} X_2^{p_2} X_3^{p_3}.$$

where f is an arbitrary function of the dimensionless variables.

Formally, this results is described by the ***Buckingham's Pi theorem***, but we will not go into details.

Example (Simple pendulum)**.**



We want to find the period P. We know that P could depend on

- mass m: $[m] = M$
- length ℓ: $[\ell] = L$
- gravity g: $[g] = LT^{-2}$
- initial displacement d: $[d] = L$

and of course $[P] = T$.

We observe that m, ℓ, g have independent dimensions, and with the fourth, we can form the dimensionless group d/ℓ. So the relationship must be of the form

$$P = f\left(\frac{d}{l}\right) m^{p_1} \ell^{p_2} g^{p_3},$$

where f is a dimensionless function. For the dimensions to balance,

$$T = M^{p_1} L^{p_2} L^{p_3} T^{-2p_3}.$$

So $p_1 = 0, p_2 = -p_3 = 1/2$. Then

$$P = f\left(\frac{d}{\ell}\right) \sqrt{\frac{\ell}{g}}.$$

While we cannot find the exact formula, using dimensional analysis, we know that if both ℓ and d are quadrupled, then P will double.

2.4 Forces

Force is a central concept in Newtonian mechanics. As described by Newton's laws of motion, forces are what causes objects to accelerate, according to the formula $\mathbf{F} = m\mathbf{a}$. To completely specify the dynamics of a system, one only needs to know what forces act on what bodies.

However, unlike in Star Wars, the force is given a secondary importance in modern treatments of mechanics. Instead, the ***potential*** is what is considered to be fundamental, with force being a derived concept. In quantum mechanics, we cannot even meaningfully define forces.

Yet, there are certain systems that must be described with forces instead of potentials, the most important of which is a system that involves friction of some sort.

Force and potential energy in one dimension

To define the potential, consider a particle of mass m moving in a straight line with position $x(t)$. Suppose $F = F(x)$, i.e. it depends on position only. We define the potential energy as follows:

Definition (Potential energy)**.** Given a force field $F = F(x)$, we define the ***potential energy*** to be a function $V(x)$ such that

$$F = -\frac{\mathrm{d}V}{\mathrm{d}x}.$$

or

$$V = -\int F\ \mathrm{d}x.$$

V is defined up to a constant term. We usually pick a constant of integration such that the potential drops to 0 at infinity.

Using the potential, we can write the equation of motion as

$$m\ddot{x} = -\frac{\mathrm{d}V}{\mathrm{d}x}.$$

There is a reason why we call this the potential ***energy***. We usually consider it to be an energy of some sort. In particular, we define the total energy of a system to be

Definition (Total energy)**.** The ***total energy*** of a system is given by

$$E = T + V,$$

where V is the potential energy and $T = \frac{1}{2}m\dot{x}^2$ is the kinetic energy.

If the force of a system is derived from a potential, we can show that energy is conserved.

Proposition. Suppose the equation of a particle satisfies

$$m\ddot{x} = -\frac{\mathrm{d}V}{\mathrm{d}x}.$$

Then the total energy

$$E = T + V = \frac{1}{2}m\dot{x}^2 + V(x)$$

is conserved, i.e. $\dot{E} = 0$.

Proof.

$$\begin{aligned}\frac{\mathrm{d}E}{\mathrm{d}t} &= m\dot{x}\ddot{x} + \frac{\mathrm{d}V}{\mathrm{d}x}\dot{x}\\ &= \dot{x}\left(m\ddot{x} + \frac{\mathrm{d}V}{\mathrm{d}x}\right)\\ &= 0\end{aligned}$$

□

Example. Consider the harmonic oscillator, whose potential is given by

$$V = \frac{1}{2}kx^2.$$

Then the equation of motion is

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

This is the case of, say, Hooke's Law for a spring.
The general solution of this is

$$x(t) = A\cos(\omega t) + B\sin(\omega t)$$

with $\omega = \sqrt{k/m}$.
A and B are arbitrary constants, and are related to the initial position and velocity by $x(0) = A, \dot{x}(0) = \omega B$.

For a general potential energy $V(x)$, conservation of energy allows us to solve the problem formally:

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

Since E is a constant, from this equation, we have

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

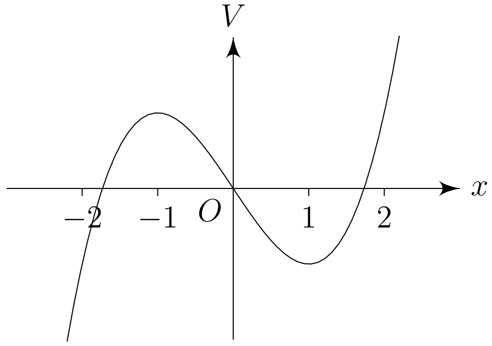
To find $x(t)$, we need to do the integral and then solve for x. This is usually not possible by analytical methods, but we can approximate the solution by numerical methods.

Motion in a potential

Given an arbitrary potential $V(x)$, it is often difficult to completely solve the equations of motion. However, just by looking at the graph of the potential, we can usually get a qualitative understanding of the dynamics.

Example. Consider $V(x) = m(x^3 - 3x)$. Note that this can be dimensionally consistent even though we add up x^3 and $-3x$, if we declare "3" to have dimension L^2.

We plot this as follows:



Suppose we release the particle from rest at $x = x_0$. Then $E = V(x_0)$. We can consider what happens to the particle for different values of x_0.

- $x_0 = \pm 1$: This is an equilibrium and the Particle stays there for all t.
- $-1 < x_0 < 2$: The particle does not have enough energy to escape the well. So it oscillates back and forth in potential well.
- $x_0 < -1$: The particle falls to $x = -\infty$.
- $x_0 > 2$: The particle has enough energy to overshoot the well and continues to $x = -\infty$.
- $x_0 = 2$: This is a special case. Obviously, the particle goes towards $x = -1$. But how long does it take, and what happens next? Here we have $E = 2m$. We noted previously
$$t - t_0 = -\int \frac{\mathrm{d}x}{\sqrt{\frac{2}{m}(E - V(x))}}.$$
Let $x = -1 + \varepsilon(t)$. Then
$$\begin{aligned}\frac{2}{m}(E - V(x)) &= 4 - 2(-1+\varepsilon)^3 + 6(-1+\varepsilon)\\ &= 6\varepsilon^2 - 2\varepsilon^3.\end{aligned}$$
So
$$t - t_0 = -\int_3^{\varepsilon} \frac{\mathrm{d}\varepsilon'}{\sqrt{6\varepsilon^2 - 2\varepsilon^3}}$$
We reach $x = -1$ when $\varepsilon \to 0$. But for small ε, the integrand is approximately $\propto 1/\varepsilon$, which integrates to $\ln \varepsilon \to -\infty$ as $\varepsilon \to 0$. So $\varepsilon = 0$ is achieved after infinite time, i.e. it takes infinite time to reach $\varepsilon = 0$, or $x = -1$.

Equilibrium points

In reality, most of the time, particles are not flying around wildly doing crazy stuff. Instead, they stay at (or near) some stable point, and only move very little in a predictable manner. We call these points ***equilibrium points***.

Definition (Equilibrium point)**.** A particle is in ***equilibrium*** if it has no tendency to move away. It will stay there for all time. Since $m\ddot{x} = -V'(x)$, the equilibrium points are the stationary points of the potential energy, i.e.

$$V'(x_0) = 0.$$

Consider motion near an equilibrium point. We assume that the motion is small and we can approximate V by a second-order Taylor expansion. Then we can write V as

$$V(x) \approx V(x_0) + \frac{1}{2}V''(x_0)(x - x_0)^2.$$

Then the equation of motion is

$$m\ddot{x} = -V''(x_0)(x - x_0).$$

If $V''(x_0) > 0$, then this is of the form of the harmonic oscillator. V has a local minimum at x_0, and we say the equilibrium point is ***stable***. The particle oscillates with angular frequency

$$\omega = \sqrt{\frac{V''(x_0)}{m}}.$$

If $V''(x_0) < 0$, then V has a local maximum at x_0. In this case, the equilibrium point is unstable, and the solution to the equation is

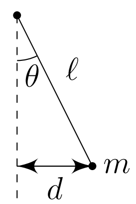
$$x - x_0 \approx Ae^{\gamma t} + Be^{-\gamma t}$$

for

$$\gamma = \sqrt{\frac{-V''(x_0)}{m}}.$$

For almost all initial conditions, $A \neq 0$ and the particle will diverge from the equilibrium point, leading to a breakdown of the approximation.
If $V''(x_0) = 0$, then further work is required to determine the outcome.

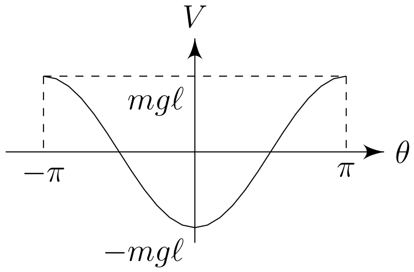
Example. Consider the simple pendulum.



Suppose that the pendulum makes an angle θ with the vertical. Then the energy is

$$E = T + V = \frac{1}{2}m\ell^2\dot{\theta}^2 - mg\ell\cos\theta.$$

Therefore $V \propto -\cos\theta$. We have a stable equilibrium at $\theta = 0$, and unstable equilibrium at $\theta = \pi$.



If $E > mg\ell$, then $\dot{\theta}$ never vanishes and the pendulum makes full circles.

If $0 < E < mg\ell$, then $\dot{\theta}$ vanishes at $\theta = \pm\theta_0$ for some $0 < \theta_0 < \pi$ i.e. $E = -mg\ell\cos\theta_0$. The pendulum oscillates back and forth. It takes a quarter of a period to reach from $\theta = 0$ to $\theta = \theta_0$. Using the previous general solution, oscillation period P is given by

$$\frac{P}{4} = \int_0^{\theta_0} = \frac{\mathrm{d}\theta}{\sqrt{\frac{2E}{m\ell^2} + \frac{2g}{\ell}\cos\theta}}.$$

Since we know that $E = -mg\ell\cos\theta_0$, we know that

$$\frac{P}{4} = \sqrt{\frac{\ell}{g}}\int_0^{\theta_0} \frac{\mathrm{d}\delta}{\sqrt{2\cos\theta - 2\cos\theta_0}}.$$

The integral is difficult to evaluate in general, but for small θ_0, we can use $\cos\theta \approx 1 - \frac{1}{2}\theta^2$. So

$$P \approx 4\sqrt{\frac{\ell}{g}}\int_0^{\theta_0} \frac{\mathrm{d}\theta}{\sqrt{\theta_0^2 - \theta^2}} = 2\pi\sqrt{\frac{\ell}{g}}$$

and is independent of the amplitude θ_0. This is of course the result for the harmonic oscillator.

Force and potential energy in three dimensions

Everything looks nice so far. However, in real life, the world has (at least) three (spatial) dimensions. To work with multiple dimensions, we will have to promote our quantities into vectors.

Consider a particle of mass m moving in 3D. The equation of motion is now a vector equation

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

We'll define the familiar quantities we've had.

Definition (Kinetic energy)**.** We define the ***kinetic energy*** of the particle to be

$$T = \frac{1}{2}m|\mathbf{v}|^2 = \frac{1}{2}m\dot{\mathbf{r}}\cdot\dot{\mathbf{r}}.$$

If we want to know how it varies with time, we obtain

$$\frac{\mathrm{d}T}{\mathrm{d}t} = m\ddot{\mathbf{r}}\cdot\dot{\mathbf{r}} = \mathbf{F}\cdot\dot{\mathbf{r}} = \mathbf{F}\cdot\mathbf{v}.$$

This is the power.

Definition (Power)**.** The ***power*** is the rate at which work is done on a particle by a force. It is given by

$$P = \mathbf{F}\cdot\mathbf{v}.$$

Definition (Work done)**.** The ***work done*** on a particle by a force is the change in kinetic energy caused by the force. The work done on a particle moving from $\mathbf{r}_1 = \mathbf{r}(t_1)$ to $\mathbf{r}_2 = \mathbf{r}(t_2)$ along a trajectory C is the line integral

$$W = \int_C \mathbf{F}\cdot\mathrm{d}\mathbf{r} = \int_{t_1}^{t_2} \mathbf{F}\cdot\dot{\mathbf{r}}\;\mathrm{d}t = \int_{t_1}^{t_2} P\;\mathrm{d}t.$$

Usually, we are interested in forces that conserve energy. These are forces which can be given a potential, and are known as ***conservative forces***.

Definition (Conservative force and potential energy)**.** A ***conservative force*** is a force field $\mathbf{F}(\mathbf{r})$ that can be written in the form

$$\mathbf{F} = -\nabla V.$$

V is the ***potential energy function***.

Proposition. If $\mathbf{F}$ is conservative, then the energy

$$\begin{aligned} E &= T + V \\ &= \frac{1}{2}m|\mathbf{v}|^2 + V(\mathbf{r}) \end{aligned}$$

is conserved. For any particle moving under this force, the work done is equal to the change in potential energy, and is independent of the path taken between the end points. In particular, if we travel around a closed loop, no work is done.

Proof.

$$\begin{aligned} \frac{\mathrm{d}E}{\mathrm{d}t} &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}m\dot{\mathbf{r}}\cdot\dot{\mathbf{r}} + V\right) \\ &= m\ddot{\mathbf{r}}\cdot\dot{\mathbf{r}} + \frac{\partial V}{\partial x_i}\frac{\mathrm{d}x_i}{\mathrm{d}t} \\ &= (m\ddot{\mathbf{r}} + \nabla V)\cdot\dot{\mathbf{r}} \\ &= (m\ddot{\mathbf{r}} - \mathbf{F})\cdot\dot{\mathbf{r}} \\ &= 0 \end{aligned}$$

So the energy is conserved. In this case, the work done is

$$W = \int_C \mathbf{F}\cdot\mathrm{d}\mathbf{r} = -\int_C (\nabla V)\cdot\mathrm{d}\mathbf{r} = V(\mathbf{r}_1) - V(\mathbf{r}_2).$$ □

Central forces

While in theory the potential can take any form it likes, most of the time, our system has ***spherical symmetry***. In this case, the potential depends only on the distance from the origin.

Definition (Central force)**.** A ***central force*** is a force with a potential $V(r)$ that depends only on the distance from the origin, $r = |\mathbf{r}|$. Note that a central force can be either attractive or repulsive.

When dealing with central forces, the following formula is often helpful:

Proposition. $\nabla r = \hat{\mathbf{r}}$.

Intuitively, this is because the direction in which r increases most rapidly is $\mathbf{r}$, and the rate of increase is clearly 1. This can also be proved algebraically:

Proof. We know that
$$r^2 = x_1^2 + x_2^2 + x_3^2.$$
Then
$$2r\frac{\partial r}{\partial x_i} = 2x_i.$$
So
$$\frac{\partial r}{\partial x_i} = \frac{x_i}{r} = (\hat{\mathbf{r}})_i.$$
□

Proposition. Let $\mathbf{F} = -\nabla V(r)$ be a central force. Then
$$\mathbf{F} = -\nabla V = -\frac{\mathrm{d}V}{\mathrm{d}r}\hat{\mathbf{r}},$$
where $\hat{\mathbf{r}} = \mathbf{r}/r$ is the unit vector in the radial direction pointing away from the origin.

Proof. Using the proof above,
$$(\nabla V)_i = \frac{\partial V}{\partial x_i} = \frac{\mathrm{d}V}{\mathrm{d}r}\frac{\partial r}{\partial x_i} = \frac{\mathrm{d}V}{\mathrm{d}r}(\hat{\mathbf{r}})_i$$
□

Since central forces have spherical symmetry, they give rise to an additional conserved quantity called ***angular momentum***.

Definition (Angular momentum)**.** The ***angular momentum*** of a particle is
$$\mathbf{L} = \mathbf{r} \times \mathbf{p} = m\mathbf{r} \times \dot{\mathbf{r}}.$$

Proposition. Angular momentum is conserved by a central force.

Proof.
$$\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t} = m\dot{\mathbf{r}} \times \dot{\mathbf{r}} + m\mathbf{r} \times \ddot{\mathbf{r}} = \mathbf{0} + \mathbf{r} \times \mathbf{F} = \mathbf{0}.$$
where the last equality comes from the fact that $\mathbf{F}$ is parallel to $\mathbf{r}$ for a central force. □

In general, for a non-central force, the rate of change of angular momentum is the ***torque***.

Definition (Torque)**.** The ***torque*** G of a particle is the rate of change of angular momentum.

$$\mathbf{G} = \frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t} = \mathbf{r} \times \mathbf{F}.$$

Note that $\mathbf{L}$ and $\mathbf{G}$ depends on the choice of origin. For a central force, only the angular momentum about the center of the force is conserved.

Gravity

We'll now study an important central force — gravity. This law was discovered by Newton and was able to explain the orbits of various planets. However, we will only study the force and potential aspects of it, and postpone the study of orbits for a later time.

Law (Newton's law of gravitation)**.** If a particle of mass M is fixed at a origin, then a second particle of mass m experiences a potential energy

$$V(r) = -\frac{GMm}{r},$$

where $G \approx 6.67 \times 10^{-11}\,\mathrm{m}^3\,\mathrm{kg}^{-1}\,\mathrm{s}^{-2}$ is the ***gravitational constant***.

The gravitational force experienced is then

$$F = -\nabla V = -\frac{GMm}{r^2}\hat{\mathbf{r}}.$$

Since the force is negative, particles are attracted to the origin.

The potential energy is a function of the masses of ***both*** the fixed mass M and the second particle m. However, it is useful what the fixed mass M does with reference to the second particle.

Definition (Gravitaional potential and field)**.** The ***gravitational potential*** is the gravitational potential energy per unit mass. It is

$$\Phi_g(r) = -\frac{GM}{r}.$$

Note that ***potential*** is confusingly different from ***potential energy***.

If we have a second particle, the potential ***energy*** is given by $V = m\Phi_g$.

The ***gravitational field*** is the force per unit mass,

$$\mathbf{g} = -\nabla\Phi_g = -\frac{GM}{r^2}\hat{\mathbf{r}}.$$

If we have many fixed masses M_i at points $\mathbf{r}_i$, we can add up their gravitational potential directly. Then the total gravitational potential is given by

$$\Phi_g(\mathbf{r}) = -\sum_i \frac{GM_i}{|\mathbf{r} - \mathbf{r}_i|}.$$

Again, $V = m\Phi_g$ for a particle of mass m.

An important (mathematical) result about gravitational fields is that we can treat spherical objects as point particles. In particular,

Proposition. The external gravitational potential of a spherically symmetric object of mass M is the same as that of a point particle with the same mass at the center of the object, i.e.

$$\Phi_g(r) = -\frac{GM}{r}.$$

The proof can be found in the Vector Calculus course.

Example. If you live on a spherical planet of mass M and radius R, and can move only a small distance $z \mathcal{L} R$ above the surface, then

$$\begin{aligned} V(r) &= V(R+z) \\ &= -\frac{GMm}{R+z} \\ &= -\frac{GMm}{R}\left(1 - \frac{z}{R} + \cdots\right) \\ &\approx \text{const.} + \frac{GMm}{R^2}z \\ &= \text{const.} + mgz, \end{aligned}$$

where $g = GM/R^2 \approx 9.8\,\mathrm{m\,s^{-2}}$ for Earth. Usually we are lazy and just say that the potential is mgz.

Example. How fast do we need to jump to escape the gravitational pull of the Earth? If we jump upwards with speed v from the surface, then

$$E = T + V = \frac{1}{2}mv^2 - \frac{GMm}{R}.$$

After escape, we must have $T \geq 0$ and $V = 0$. Since energy is conserved, we must have $E \geq 0$ from the very beginning. i.e.

$$v > v_{esc} = \sqrt{\frac{2GM}{R}}.$$

Inertial and gravitational mass

A careful reader would observe that "mass" appears in two unrelated equations:

$$\mathbf{F} = m_i\ddot{\mathbf{r}}$$

and

$$\mathbf{F} = -\frac{GM_gm_g}{r^2}\hat{\mathbf{r}},$$

and they play totally different roles. The first is the ***inertial mass***, which measures the resistance to motion, while the second is the ***gravitational mass***, which measures its response to gravitational forces.

Conceptually, these are quite different. There is no ***a priori*** reason why these two should be equal. However, experiment shows that they are indeed equivalent to each other, i.e. $m_i = m_g$, with an accuracy of 10^{-12} or better.

This (philosophical) problem was only resolved when Einstein introduced his general theory of relativity, which says that gravity is actually a ***fictitious*** force, which means that the acceleration of the particle is independent of its mass.

We can further distinct the gravitational mass by "passive" and "active", i.e. the amount of gravitational field generated by a particle (M), and the amount of gravitational force received by a particle (m), but they are still equal, and we end up calling all of them "mass".

Electromagnetism

Next we will study the laws of electromagnetism. We will only provide a very rudimentary introduction to electromagnetism. Electromagnetism will be examined more in-depth in the IB Electromagnetism and II Electrodynamics courses.

As the name suggests, electromagnetism comprises two parts — electricity and magnetism. Similar to gravity, we generally imagine electromagnetism working as follows: charges generate fields, and fields cause charges to move.

A common charged particle is the ***electron***, which is currently believed to be a fundamental particle. It has charge $q_e = -1.6 \times 10^{-19}\,\mathrm{C}$. Other particles' charges are always integer multiples of q_e (unless you are a quark).

In electromagnetism, there are two fields — the ***electric field*** $\mathbf{E}(\mathbf{r}, t)$ and the ***magnetic field*** $\mathbf{B}(\mathbf{r}, t)$. Their effects on charged particles is described by the ***Lorentz force law***.

Law (Lorentz force law)**.** The ***electromagnetic force*** experienced by a particle with electric charge q is

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}).$$

This is the first time where we see a force that depends on the ***velocity*** of the particle. In all other forces we've seen, the force depends only on the field which implicitly depends on the position only. This is weird, and seems to violate Galilean relativity, since velocity is a relative concept that depends on the frame of reference. It turns out that weird things happen to the $\mathbf{B}$ and $\mathbf{E}$ fields when you change the frame of reference. You will learn about these in the IB Electromagnetism course (if you take it).

As a result, the magnetic force is not a conservative force, and it cannot be given a (regular) potential. On the other hand, assuming that the fields are time-independent, the electric field ***is*** conservative. We write the potential as $\Phi_e(\mathbf{r})$, and $\mathbf{E} = -\nabla\Phi_e$.

Definition (Electrostatic potential)**.** The electrostatic potential is a function $\Phi_e(\mathbf{r})$ such that

$$\mathbf{E} = -\nabla\Phi_e.$$

While the magnetic force is not conservative in the traditional sense, it always acts perpendicularly to the velocity. Hence it does no work. So overall, energy is conserved under the action of the electromagnetic force.

Proposition. For time independent $\mathbf{E}(\mathbf{r})$ and $\mathbf{B}(\mathbf{r})$, the energy

$$E = T + V = \frac{1}{2}m|\mathbf{v}|^2 + q\Phi_e$$

is conserved.

Proof.

$$\begin{aligned}\frac{\mathrm{d}E}{\mathrm{d}t} &= m\ddot{\mathbf{r}}\cdot\dot{\mathbf{r}} + q(\nabla\Phi_e)\cdot\dot{\mathbf{r}}\\ &= (m\ddot{\mathbf{r}} - q\mathbf{E})\cdot\dot{\mathbf{r}}\\ &= (q\dot{\mathbf{r}}\times\mathbf{B})\cdot\dot{\mathbf{r}}\\ &= 0\end{aligned}$$

□

Motion in a uniform magnetic field

Consider the particular case where there is no electric field, i.e. $\mathbf{E} = \mathbf{0}$, and that the magnetic field is uniform throughout space. We choose our axes such that $\mathbf{B} = (0, 0, B)$ is constant.

According to the Lorentz force law, $\mathbf{F} = q(\mathbf{E} + \mathbf{v}\times\mathbf{B}) = q\mathbf{v}\times\mathbf{B}$. Since the force is always perpendicular to the velocity, we expect this to act as a centripetal force to make the particle travel in circles.

Indeed, writing out the components of the equation of motion, we obtain

$$m\ddot{x} = qB\dot{y} \tag{1}$$
$$m\ddot{y} = -qB\dot{x} \tag{2}$$
$$m\ddot{z} = 0 \tag{3}$$

From (3), we see that there is uniform motion parallel to $\mathbf{B}$, which is not interesting. We will look at the x and y components.

There are many ways to solve this system of equations. Here we solve it using complex numbers.

Let $\zeta = x + iy$. Then (1) + (2)i gives

$$m\ddot{\zeta} = -iqB\dot{\zeta}.$$

Then the solution is

$$\zeta = \alpha e^{-i\omega t} + \beta,$$

where $\omega = qB/m$ is the ***gyrofrequency***, and α and β are complex integration constants. We see that the particle goes in circles, with center β and radius α.

We can choose coordinates such that, at $t = 0$, $\mathbf{r} = 0$ and $\dot{\mathbf{r}} = (0, v, w)$, i.e. $\zeta = 0$ and $\dot{\zeta} = iv$, and $z = 0$ and $\dot{z} = w$.

The solution is then

$$\zeta = R(1 - e^{-i\omega t}).$$

with $R = v/\omega = (mv)/(qB)$ is the ***gyroradius*** or ***Larmor radius***. Alternatively,

$$\begin{aligned} x &= R(1 - \cos\omega t) \\ y &= R\sin\omega t \\ z &= wt. \end{aligned}$$

This is circular motion in the plane perpendicular to $\mathbf{B}$:

$$(x - R)^2 + y^2 = R^2,$$

combined with uniform motion parallel to $\mathbf{B}$, i.e. a helix.
Alternatively, we can solve this with vector operations. Start with

$$m\ddot{\mathbf{r}} = q\dot{\mathbf{r}} \times \mathbf{B}$$

Let $\mathbf{B} = B\mathbf{n}$ with $|\mathbf{n}| = 1$. Then

$$\ddot{\mathbf{r}} = \omega\dot{\mathbf{r}} \times \mathbf{n},$$

with our gyrofrequency $\omega = qB/m$. We integrate once, assuming $\mathbf{r}(0) = \mathbf{0}$ and $\dot{\mathbf{r}}(0) = \mathbf{v}_0$.

$$\dot{\mathbf{r}} = \omega\mathbf{r} \times \mathbf{n} + \mathbf{v}_0. \tag{$*$}$$

Now we project $(*)$ parallel to and perpendicular to $\mathbf{B}$.
First we dot $(*)$ with $\mathbf{n}$:

$$\dot{\mathbf{r}} \cdot \mathbf{n} = \mathbf{v}_0 \cdot \mathbf{n} = w = \text{const.}$$

We integrate again to obtain

$$\mathbf{r} \cdot \mathbf{n} = wt.$$

This is the part parallel to $\mathbf{B}$.
To resolve perpendicularly, write $\mathbf{r} = (\mathbf{r} \cdot \mathbf{n})\mathbf{n} + \mathbf{r}_\perp$, with $\mathbf{r}_\perp \cdot \mathbf{n} = 0$.
The perpendicular component of $(*)$ gives

$$\dot{\mathbf{r}}_\perp = w\mathbf{r}_\perp \times \mathbf{n} + \mathbf{v}_0 - (\mathbf{v}_0 \cdot \mathbf{n})\mathbf{n}.$$

We solve this by differentiating again to obtain

$$\ddot{\mathbf{r}}_\perp = \omega\dot{\mathbf{r}}_\perp \times \mathbf{n} = -\omega^2\mathbf{r}_\perp + \omega\mathbf{v}_0 \times \mathbf{n},$$

which we can solve using particular integrals and complementary functions.

Point charges

So far we've discussed the effects of the fields on particles. But how can we create fields in the first place? We'll only look at the simplest case, where a point charge generates an electric field.

Law (Columb's law)**.** A particle of charge Q, fixed at the origin, produces an electrostatic potential

$$\Phi_e = \frac{Q}{4\pi\varepsilon_0 r},$$

where $\varepsilon_0 \approx 8.85 \times 10^{-12}\,\mathrm{m^{-3}\,kg^{-1}\,s^2\,C^2}$.

The corresponding electric field is

$$\mathbf{E} = -\nabla\Phi_e = \frac{Q}{4\pi\varepsilon_0}\frac{\hat{\mathbf{r}}}{r^2}.$$

The resulting force on a particle of charge q is

$$\mathbf{F} = q\mathbf{E} = \frac{Qq}{4\pi\varepsilon_0}\frac{\hat{\mathbf{r}}}{r^2}.$$

Definition (Electric constant)**.** ε_0 is the ***electric constant*** or ***vacuum permittivity*** or ***permittivity of free space***.

The form of equations here are closely analogous to those of gravity. However, there is an important difference: charges can be positive or negative. Thus electrostatic forces can be either attractive or repulsive, whereas gravity is always attractive.

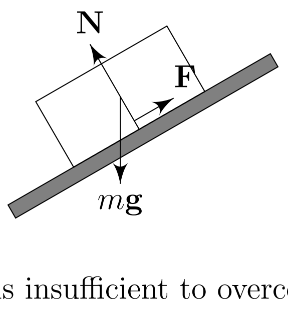
Friction

At an atomic level, energy is always conserved. However, in many everyday processes, this does not appear to be the case. This is because ***friction*** tends to take kinetic energy away from objects.

In general, we can divide friction into ***dry friction*** and ***fluid friction***.

Dry friction

When solid objects are in contact, a ***normal reaction force*** $\mathbf{N}$ (perpendicular to the contact surface) prevents them from interpenetrating, while a ***frictional force*** $\mathbf{F}$ (tangential to the surface) resists relative tangential motion (sliding or slipping).



If the tangential force is small, it is insufficient to overcome friction and no sliding occurs. We have ***static friction*** of

$$|\mathbf{F}| \le \mu_s|\mathbf{N}|,$$

where μ_s is the ***coefficient of static friction***.

When the external force on the object exceeds $\mu_s|\mathbf{N}|$, sliding starts, and we have a ***kinetic friction*** of
$$|\mathbf{F}| = \mu_k|\mathbf{N}|,$$
where μ_k is the ***coefficient of kinetic friction***.

These coefficients are measures of roughness and depend on the two surfaces involved. For example, Teflon on Teflon has coefficient of around 0.04, while rubber on asphalt has about 0.8, while a hypothetical perfectly smooth surface has coefficient 0. Usually, $\mu_s > \mu_k > 0$.

Fluid drag

When a solid object moves through a fluid (i.e. liquid or gas), it experiences a ***drag force***. There are two important regimes.

(i) Linear drag: for small things in viscous fluids moving slowly, e.g. a single cell organism in water, the friction is proportional to the velocity, i.e.

$$\mathbf{F} = -k_1\mathbf{v}.$$

where $\mathbf{v}$ is the velocity of the object relative to the fluid, and $k_1 > 0$ is a constant. This k_1 depends on the shape of the object. For example, for a sphere of radius R, Stoke's Law gives
$$k_1 = 6\pi\mu R,$$
where μ is the viscosity of the fluid.

(ii) Quadratic drag: for large objects moving rapidly in less viscous fluid, e.g. cars or tennis balls in air, the friction is proportional to the square of the velocity, i.e.

$$\mathbf{F} = -k_2|\mathbf{v}|^2\hat{\mathbf{v}}.$$

In either case, the object loses energy. The power exerted by the drag force is

$$\mathbf{F}\cdot\mathbf{v} = \begin{cases} -k_1|\mathbf{v}|^2 \\ -k_2|\mathbf{v}|^3 \end{cases}$$

Example. Consider a projectile moving in a uniform gravitational field and experiencing a linear drag force.

At $t = 0$, we throw the projectile with velocity $\mathbf{u}$ from $\mathbf{x} = \mathbf{0}$.

The equation of motion is
$$m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = m\mathbf{g} - k\mathbf{v}.$$
We first solve for $\mathbf{v}$, and then deduce $\mathbf{x}$.

We use an integrating factor $\exp(\frac{k}{m}t)$ to obtain

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{kt/m}\mathbf{v}\right) &= e^{kt/m}\mathbf{g}\\
e^{kt/m}\mathbf{v} &= \frac{m}{k}e^{kt/m}\mathbf{g} + \mathbf{c}\\
\mathbf{v} &= \frac{m}{k}\mathbf{g} + \mathbf{c}e^{-kt/m}
\end{aligned}
$$

Since $\mathbf{v} = \mathbf{u}$ at $t = 0$, we get $\mathbf{c} = \mathbf{u} - \frac{m}{k}\mathbf{g}$. So

$$
\mathbf{v} = \dot{\mathbf{x}} = \frac{m}{k}\mathbf{g} + \left(\mathbf{u} - \frac{m}{k}\mathbf{g}\right)e^{-kt/m}.
$$

Integrating once gives

$$
\mathbf{x} = \frac{m}{k}\mathbf{g}t - \frac{m}{k}\left(\mathbf{u} - \frac{m}{k}\mathbf{g}\right)e^{-kt/m} + \mathbf{d}.
$$

Since $\mathbf{x} = \mathbf{0}$ at $t = 0$. So

$$
\mathbf{d} = \frac{m}{k}\left(\mathbf{u} - \frac{m}{k}\mathbf{g}\right).
$$

So

$$
\mathbf{x} = \frac{m}{k}\mathbf{g}t + \frac{m}{k}\left(\mathbf{u} - \frac{m}{k}\mathbf{g}\right)(1 - e^{-kt/m}).
$$

In component form, let $\mathbf{x} = (x, y)$, $\mathbf{u} = (u\cos\theta, u\sin\theta)$, $\mathbf{g} = (0, -g)$. So

$$
\begin{aligned}
x &= \frac{mu}{k}\cos\theta(1 - e^{-kt/m})\\
y &= -\frac{mgt}{k} + \frac{m}{k}\left(u\sin\theta + \frac{mg}{k}\right)(1 - e^{-kt/m}).
\end{aligned}
$$

We can characterize the strength of the drag force by the dimensionless constant $ku/(mg)$, with a larger constant corresponding to a larger drag force.

Effect of damping on small oscillations

We've previously seen that particles near a potential minimum oscillate indefinitely. However, if there is friction in the system, the oscillation will damp out and energy is continually lost. Eventually, the system comes to rest at the stable equilibrium.

Example. If a linear drag force is added to a harmonic oscillator, then the equation of motion becomes

$$
m\ddot{\mathbf{x}} = -m\omega^2\mathbf{x} - k\dot{\mathbf{x}},
$$

where ω is the angular frequency of the oscillator in the absence of damping. Rewrite as

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

where $\gamma = k/2m > 0$. Solutions are $x = e^{\lambda t}$, where

$$
\lambda^2 + 2\gamma\lambda + \omega^2 = 0,
$$

or

$$\lambda = -\gamma \pm \sqrt{\gamma^2 - \omega^2}.$$

If $\gamma > \omega$, then the roots are real and negative. So we have exponential decay. We call this an overdamped oscillator.

If $0 < \gamma < \omega$, then the roots are complex with $\mathrm{Re}(\lambda) = -\gamma$. So we have decaying oscillations. We call this an underdamped oscillator.

For details, refer to Differential Equations.

2.5 Orbits

The goal of this chapter is to study the orbit of a particle in a central force,

$$m\ddot{\mathbf{r}} = -\nabla V(r).$$

While the universe is in three dimensions, the orbit is confined to a plane. This is since the angular momentum $\mathbf{L} = m\mathbf{r} \times \dot{\mathbf{r}}$ is a constant vector, as we've previously shown. Furthermore $\mathbf{L} \cdot \mathbf{r} = 0$. Therefore, the motion takes place in a plane passing through the origin, and perpendicular to $\mathbf{L}$.

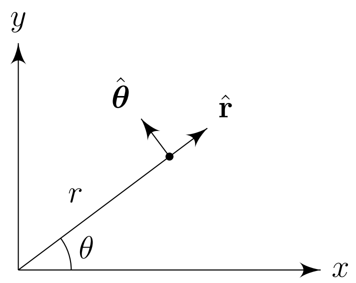
Polar coordinates in the plane

We choose our axes such that the orbital plane is $z = 0$. To describe the orbit, we introduce polar coordinates (r, θ):

$$x = r\cos\theta, \quad y = r\sin\theta.$$

Our object is to separate the motion of the particle into radial and angular components. We do so by defining unit vectors in the directions of increasing r and increasing θ:

$$\hat{\mathbf{r}} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \quad \hat{\boldsymbol{\theta}} = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}.$$



These two unit vectors form an orthonormal basis. However, they are not basis vectors in the normal sense. The directions of these basis vectors depend on time. In particular, we have

Proposition.

$$\frac{\mathrm{d}\hat{\mathbf{r}}}{\mathrm{d}\theta} = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix} = \hat{\boldsymbol{\theta}}$$

$$\frac{\mathrm{d}\hat{\boldsymbol{\theta}}}{\mathrm{d}\theta} = \begin{pmatrix} -\cos\theta \\ -\sin\theta \end{pmatrix} = -\hat{\mathbf{r}}.$$

Often, we want the derivative with respect to time, instead of θ. By the chain rule, we have

$$\frac{\mathrm{d}\hat{\mathbf{r}}}{\mathrm{d}t} = \dot{\theta}\hat{\boldsymbol{\theta}}, \quad \frac{\mathrm{d}\hat{\boldsymbol{\theta}}}{\mathrm{d}t} = -\dot{\theta}\hat{\mathbf{r}}.$$

We can now express the position, velocity and acceleration in this new polar basis. The position is given by

$$\mathbf{r} = r\hat{\mathbf{r}}.$$

Taking the derivative gives the velocity as

$$\dot{\mathbf{r}} = \dot{r}\hat{\mathbf{r}} + r\dot{\theta}\hat{\boldsymbol{\theta}}.$$

The acceleration is then

$$\begin{aligned}\ddot{\mathbf{r}} &= \ddot{r}\hat{\mathbf{r}} + \dot{r}\dot{\theta}\hat{\boldsymbol{\theta}} + \dot{r}\dot{\theta}\hat{\boldsymbol{\theta}} + r\ddot{\theta}\hat{\boldsymbol{\theta}} - r\dot{\theta}^2\hat{\mathbf{r}}\\ &= (\ddot{r} - r\dot{\theta}^2)\hat{\mathbf{r}} + (r\ddot{\theta} + 2\dot{r}\dot{\theta})\hat{\boldsymbol{\theta}}.\end{aligned}$$

Definition (Radial and angular velocity). $\dot{r}$ is the ***radial velocity***, and $\dot{\theta}$ is the ***angular velocity***.

Example (Uniform motion in a circle). If we are moving in a circle, then $\dot{r} = 0$ and $\dot{\theta} = \omega = \text{constant}$. So

$$\dot{\mathbf{r}} = r\omega\hat{\boldsymbol{\theta}}.$$

The speed is given by

$$v = |\dot{\mathbf{r}}| = r|\omega| = \text{const}$$

and the acceleration is

$$\ddot{\mathbf{r}} = -r\omega^2\hat{\mathbf{r}}.$$

Hence in order to make a particle of mass m move uniformly in a circle, we must supply a ***centripetal force*** mv^2/r towards the center.

Motion in a central force field

Now let's put in our central force. Since $V = V(r)$, we have

$$\mathbf{F} = -\nabla V = \frac{\mathrm{d}V}{\mathrm{d}r}\hat{\mathbf{r}}.$$

So Newton's 2nd law in polar coordinates is

$$m(\ddot{r} - r\dot{\theta}^2)\hat{\mathbf{r}} + m(r\ddot{\theta} + 2\dot{r}\dot{\theta})\hat{\boldsymbol{\theta}} = -\frac{\mathrm{d}V}{\mathrm{d}r}\hat{\mathbf{r}}.$$

The θ component of this equation is

$$m(r\ddot{\theta} - 2\dot{r}\dot{\theta}) = 0.$$

We can rewrite it as

$$\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}t}(mr^2\dot{\theta}) = 0.$$

Let $L = mr^2\dot{\theta}$. This is the z component (and the only component) of the conserved angular momentum $\mathbf{L}$:

$$\begin{aligned}
\mathbf{L} &= m\mathbf{r} \times \dot{\mathbf{r}} \\
&= mr\hat{\mathbf{r}} \times (\dot{r}\hat{\mathbf{r}} + r\dot{\theta}\hat{\boldsymbol{\theta}}) \\
&= mr^2\dot{\theta}\, \hat{\mathbf{r}} \times \hat{\boldsymbol{\theta}} \\
&= mr^2\dot{\theta}\, \hat{\mathbf{z}}.
\end{aligned}$$

So the angular component tells us that L is constant, which is the conservation of angular momentum.

However, a more convenient quantity is the angular momentum ***per unit mass***:

Notation (Angular momentum per unit mass)**.** The ***angular momentum per unit mass*** is

$$h = \frac{L}{m} = r^2\dot{\theta} = \text{const.}$$

Now the radial (r) component of the equation of motion is

$$m(\ddot{r} - r\dot{\theta}^2) = -\frac{\mathrm{d}V}{\mathrm{d}r}.$$

We eliminate $\dot{\theta}$ using $r^2\dot{\theta} = h$ to obtain

$$m\ddot{r} = -\frac{\mathrm{d}V}{\mathrm{d}r} + \frac{mh^2}{r^3} = -\frac{\mathrm{d}V_{\text{eff}}}{\mathrm{d}r},$$

where

$$V_{\text{eff}}(r) = V(r) + \frac{mh^2}{2r^2}.$$

We have now reduced the problem to 1D motion in an (effective) potential — as studied previously.

The total energy of the particle is

$$\begin{aligned}
E &= \frac{1}{2}m|\dot{\mathbf{r}}|^2 + V(r) \\
&= \frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) + V(r)
\end{aligned}$$

(since $\dot{\mathbf{r}} = \dot{r}\hat{\mathbf{r}} + r\dot{\theta}\hat{\boldsymbol{\theta}}$, and $\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\theta}}$ are orthogonal)

$$\begin{aligned}
&= \frac{1}{2}m\dot{r}^2 + \frac{mh^2}{2r^2} + V(r) \\
&= \frac{1}{2}m\dot{r}^2 + V_{\text{eff}}(r).
\end{aligned}$$

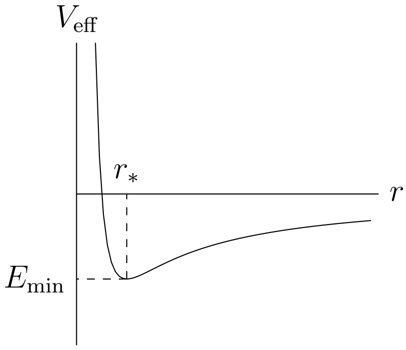
Example. Consider an attractive force following the inverse-square law (e.g. gravity). Here

$$V = -\frac{mk}{r},$$

for some constant k. So

$$V_{\text{eff}} = -\frac{mk}{r} + \frac{mh^2}{2r^2}.$$

We have two terms of opposite signs and different dependencies on r. For small r, the second term dominates and V_{eff} is large. For large r, the first term dominates. Then V_{eff} asymptotically approaches 0 from below.



The minimum of V_{eff} is at

$$r_* = \frac{h^2}{k}, \quad E_{\min} = -\frac{mk^2}{2h^2}.$$

We have a few possible types of motion:

- If $E = E_{\min}$, then r remains at r_* and $\dot{\theta}h/r^2$ is constant. So we have a uniform motion in a circle.
- If $E_{\min} < E < 0$, then r oscillates and $\dot{r} = h/r^2$ does also. This is a non-circular, bounded orbit.

  We'll now introduce a lot of (pointless) definitions:

  Definition (Periapsis, apoapsis and apsides)**.** The points of minimum and maximum r in such an orbit are called the ***periapsis*** and ***apoapsis***. They are collectively known as the ***apsides***.

  Definition (Perihelion and aphelion)**.** For an orbit around the Sun, the periapsis and apoapsis are known as the ***perihelion*** and ***aphelion***.

  In particular

  Definition (Perigee and apogee)**.** The perihelion and aphelion of the Earth are known as the ***perigee*** and ***apogee***.

- If $E \geq 0$, then r comes in from ∞, reaches a minimum, and returns to infinity. This is an unbounded orbit.

We will later show that in the case of motion in an inverse square force, the trajectories are conic sections (circles, ellipses, parabolae and hyperbolae).

Stability of circular orbits

We'll now look at circular orbits, since circles are nice. Consider a general potential energy $V(r)$. We have to answer two questions:

- Do circular orbits exist?
- If they do, are they stable?

The conditions for existence and stability are rather straightforward. For a circular orbit, $r = r_* = \text{const}$ for some value of $h \neq 0$ (if $h = 0$, then the object is just standing still!). Since $\ddot{r} = 0$ for constant r, we require

$$V'_{\text{eff}}(r_*) = 0.$$

The orbit is stable if r_* is a minimum of V_{eff}, i.e.

$$V''_{\text{eff}}(r_*) > 0.$$

In terms of $V(r)$, circular orbit requires

$$V'(r_*) = \frac{mh^2}{r_*^3}$$

and stability further requires

$$V''(r_*) + \frac{3mh^2}{r_*^4} = V''(r_*) + \frac{3}{r_*}V'(r_*) > 0.$$

In terms of the radial force $F(r) = -V'(r)$, the orbit is stable if

$$F'(r_*) + \frac{3}{r}F(r_*) < 0.$$

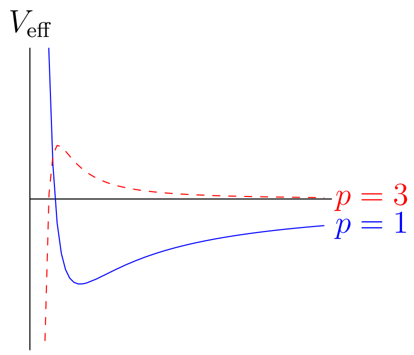
Example. Consider a central force with

$$V(r) = -\frac{mk}{r^p}$$

for some $k, p > 0$. Then

$$V''(r) + \frac{3}{r}V'(r) = \big(-p(p+1) + 3p\big)\frac{mk}{r^{p+2}} = p(2-p)\frac{mk}{r^{p+2}}.$$

So circular orbits are stable for $p < 2$. This is illustrated by the graphs of $V_{\text{eff}}(r)$ for $p = 1$ and $p = 3$.



Equation of the shape of the orbit

In general, we could determine $r(t)$ by integrating the energy equation

$$E = \frac{1}{2}m\dot{r}^2 + V_{\text{eff}}(r)$$
$$t = \pm\sqrt{\frac{m}{2}}\int \frac{\mathrm{d}r}{\sqrt{E - V_{\text{eff}}(r)}}$$

However, this is usually not practical, because we can't do the integral. Instead, it is usually much easier to find the shape $r(\theta)$ of the orbit.

Still, solving for $r(\theta)$ is also not easy. We will need a magic trick — we introduce the new variable

Notation.

$$u = \frac{1}{r}.$$

Then

$$\dot{r} = \frac{\mathrm{d}r}{\mathrm{d}\theta}\dot{\theta} = \frac{\mathrm{d}r}{\mathrm{d}\theta}\frac{h}{r^2} = -h\frac{\mathrm{d}u}{\mathrm{d}\theta},$$

and

$$\ddot{r} = \frac{\mathrm{d}}{\mathrm{d}t}\left(-h\frac{\mathrm{d}u}{\mathrm{d}\theta}\right) = -h\frac{\mathrm{d}^2u}{\mathrm{d}\theta^2}\dot{\theta} = -h\frac{\mathrm{d}^2u}{\mathrm{d}\theta^2}\frac{h}{r^2} = -h^2u^2\frac{\mathrm{d}^2u}{\mathrm{d}\theta^2}.$$

This doesn't look very linear with u^2, but it will help linearizing the equation when we put in other factors.

The radial equation of motion

$$m\ddot{r} - \frac{mh^2}{r^3} = F(r)$$

then becomes

Proposition (Binet's equation)**.**

$$-mh^2u^2\left(\frac{\mathrm{d}^2u}{\mathrm{d}\theta^2} + u\right) = F\left(\frac{1}{u}\right).$$

This still looks rather complicated, but observe that for an inverse square force, $F(1/u)$ is proportional to u^2, and then the equation is linear!

In general, given an arbitrary $F(r)$, we aim to solve this second order ODE for $u(\theta)$. If needed, we can then work out the time-dependence via

$$\dot{\theta} = hu^2.$$

The Kepler problem

The Kepler problem is the study of the orbits of two objects interacting via a central force that obeys the inverse square law. The goal is to classify the possible orbits and study their properties. One of the most important examples of the Kepler problem is the orbit of celestial objects, as studied by Kepler himself.

Shapes of orbits

For a planet orbiting the sun, the potential and force are given by

$$V(r) = \frac{mk}{r}, \quad F(r) = -\frac{mk}{r^2}$$

with $k = GM$ (for the Coulomb attraction of opposite charges, we have the same equation with $k = -\dfrac{Qq}{4\pi\varepsilon_0 m}$).
Binet's equation then becomes linear, and

$$\frac{\mathrm{d}^2 u}{\mathrm{d}\theta^2} + u = \frac{k}{h^2}.$$

We write the general solution as

$$u = \frac{k}{h^2} + A\cos(\theta - \theta_0),$$

where $A \geq 0$ and θ_0 are arbitrary constants.
If $A = 0$, then u is constant, and the orbit is circular. Otherwise, u reaches a maximum at $\theta = \theta_0$. This is the periapsis. We now re-define polar coordinates such that the periapsis is at $\theta = 0$. Then

Proposition. The orbit of a planet around the sun is given by

$$r = \frac{\ell}{1 + e\cos\theta}, \tag{$*$}$$

with $\ell = h^2/k$ and $e = Ah^2/k$. This is the polar equation of a conic, with a focus (the sun) at the origin.

Definition (Eccentricity)**.** The dimensionless parameter $e \geq 0$ in the equation of orbit is the ***eccentricity*** and determines the shape of the orbit.

We can rewrite $(*)$ in Cartesian coordinates with $x = r\cos\theta$ and $y = r\sin\theta$. Then we obtain

$$(1 - e^2)x^2 + 2e\ell x + y^2 = \ell^2. \tag{$\dagger$}$$

There are three different possibilities:

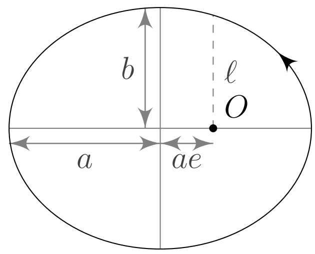
- Ellipse: $(0 \leq e < 1)$. r is bounded by

  $$\frac{\ell}{1+e} \leq r \leq \frac{\ell}{1-e}.$$

  $(\dagger)$ can be put into the equation of an ellipse centered on $(-ea, 0)$,

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

  where $a = \dfrac{\ell}{1 - e^2}$ and $b = \dfrac{\ell}{\sqrt{1 - e^2}} \leq a$.

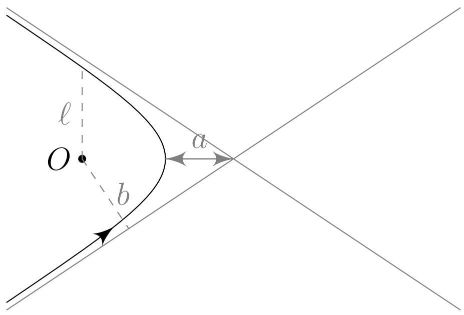


a and b are the semi-major and semi-minor axis. ℓ is the ***semi-latus rectum***. One focus of the ellipse is at the origin. If $e = 0$, then $a = b = \ell$ and the ellipse is a circle.

– Hyperbola: ($e > 1$). For $e > 1$, $r \to \infty$ as $\theta \to \pm\alpha$, where $\alpha = \cos^{-1}(1/e) \in (\pi/2, \pi)$. Then (†) can be put into the equation of a hyperbola centered on $(ea, 0)$,

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

with $a = \dfrac{\ell}{e^2 - 1}$, $b = \dfrac{\ell}{\sqrt{e^2 - 1}}$.



This corresponds to an unbound orbit that is deflected (scattered) by an attractive force.

b is both the semi-minor axis and the ***impact parameter***. It is the distance by which the planet would miss the object if there were no attractive force.

The asymptote is $y = \frac{b}{a}(x - ea)$, or

$$x\sqrt{e^2 - 1} - y = eb.$$

Alternatively, we can write the equation of the asymptote as

$$(x, y) \cdot \left(\frac{\sqrt{e^2 - 1}}{e}, -\frac{1}{e}\right) = b$$

or $\mathbf{r} \cdot \mathbf{n} = b$, the equation of a line at a distance b from the origin.

– Parabola: ($e = 1$). Then ($*$) becomes

$$r = \frac{\ell}{1 + \cos\theta}.$$

We see that $r \to \infty$ as $\theta \to \pm\pi$. (†) becomes the equation of a parabola, $y^2 = \ell(\ell - 2x)$. The trajectory is similar to that of a hyperbola.

Energy and eccentricity

We can figure out which path a planet follows by considering its energy.

$$\begin{aligned}E &= \frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) - \frac{mk}{r}\\ &= \frac{1}{2}mh^2\left(\left(\frac{\mathrm{d}u}{\mathrm{d}\theta}\right)^2 + u^2\right) - mku\end{aligned}$$

Substitute $u = \frac{1}{\ell}(1 + e\cos\theta)$ and $\ell = \frac{h^2}{k}$, and it simplifies to

$$E = \frac{mk}{2\ell}(e^2 - 1),$$

which is independent of θ, as it must be.

Orbits are bounded for $e < 1$. This corresponds to the case $E < 0$. Unbounded orbits have $e > 1$ and thus $E > 0$. A parabolic orbit has $e = 1$, $E = 0$, and is "marginally bound".

Note that the condition $E > 0$ is equivalent to $|\dot{\mathbf{r}}| > \sqrt{\frac{2GM}{r}} = v_{\text{esc}}$, which means you have enough kinetic energy to escape orbit.

Kepler's laws of planetary motion

When Kepler first studied the laws of planetary motion, he took a telescope, observed actual planets, and came up with his famous three laws of motion. We are now going to derive the laws with pen and paper instead.

Law (Kepler's first law)**.** The orbit of each planet is an ellipse with the Sun at one focus.

Law (Kepler's second law)**.** The line between the planet and the sun sweeps out equal areas in equal times.

Law (Kepler's third law)**.** The square of the orbital period is proportional to the cube of the semi-major axis, or

$$P^2 \propto a^3.$$

We have already shown that Law 1 follows from Newtonian dynamics and the inverse-square law of gravity. In the solar system, planets generally have very low eccentricity (i.e. very close to circular motion), but asteroids and comets can have very eccentric orbits. In other solar systems, even planets have have highly eccentric orbits. As we've previously shown, it is also possible for the object to have a parabolic or hyperbolic orbit. However, we tend not to call these "planets".

Law 2 follows simply from the conservation of angular momentum: The area swept out by moving $\mathrm{d}\theta$ is $\mathrm{d}A = \frac{1}{2}r^2\,\mathrm{d}\theta$ (area of sector of circle). So

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \frac{1}{2}r^2\dot{\theta} = \frac{h}{2} = \text{const.}$$

and is true for ***any*** central force.

Law 3 follows from this: the total area of the ellipse is $A = \pi ab = \frac{h}{2}P$ (by the second law). But $b^2 = a^2(1-e^2)$ and $h^2 = k\ell = ka(1-e^2)$. So

$$P^2 = \frac{(2\pi)^2 a^4(1-e^2)}{ka(1-e^2)} = \frac{(2\pi)^2 a^3}{k}.$$

Note that the third law is very easy to prove directly for circular orbits. Since the radius is constant, $\ddot{r} = 0$. So the equations of motion give

$$-r\dot{\theta}^2 = -\frac{k}{r^2}$$

So

$$r^3\dot{\theta}^2 = k$$

Since $\dot{\theta} \propto P^{-1}$, the result follows.

Rutherford scattering

Finally, we will consider the case where the force is ***repulsive*** instead of attractive. An important example is the Rutherford gold foil experiment, where Rutherford bombarded atoms with alpha particles, and the alpha particles are repelled by the nucleus of the atom.

Under a repulsive force, the potential and force are given by

$$V(r) = +\frac{mk}{r}, \quad F(r) = +\frac{mk}{r^2}.$$

For Coulomb repulsion of like charges,

$$k = \frac{Qq}{4\pi\varepsilon_0 m} > 0.$$

The solution is now

$$u = -\frac{k}{h^2} + A\cos(\theta - \theta_0).$$

wlog, we can take $A \geq 0, \theta_0 = 0$. Then

$$r = \frac{\ell}{e\cos\theta - 1}$$

with

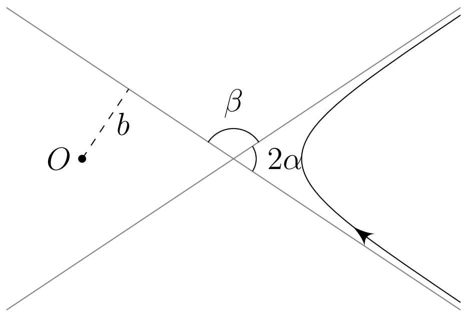
$$\ell = \frac{h^2}{k}, \quad e = \frac{Ah^2}{k}.$$

We know that r and ℓ are positive. So we must have $e \geq 1$. Then $r \to \infty$ as $\theta \to \pm\alpha$, where $\alpha = \cos^{-1}(1/e)$.

The orbit is a hyperbola, again given by

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

with $a = \frac{\ell}{e^2-1}$ and $b = \frac{\ell}{\sqrt{e^2-1}}$. However, this time, the trajectory is the other branch of the hyperbola.



It seems as if the particle is deflected by O.

We can characterize the path of the particle by the impact parameter b and the incident speed v (i.e. the speed when far away from the origin). We know that the angular momentum per unit mass is $h = bv$ (velocity $\times$ perpendicular distance to O).

How does the scattering angle $\beta = \pi - 2\alpha$ depend on the impact parameter b and the incident speed v?

Recall that the angle α is given by $\alpha = \cos^{-1}(1/e)$. So we obtain

$$\frac{1}{e} = \cos\alpha = \cos\left(\frac{\pi}{2} - \frac{\beta}{2}\right) = \sin\left(\frac{\beta}{2}\right),$$

So

$$b = \frac{\ell}{\sqrt{e^2 - 1}} = \frac{(bv)^2}{k}\tan\frac{\beta}{2}.$$

So

$$\beta = 2\tan^{-1}\left(\frac{k}{bv^2}\right).$$

We see that if we have a small impact parameter, i.e. $b\mathcal{L}k/v^2$, then we can have a scattering angle approaching π.

2.6 Rotating frames

Recall that Newton's laws hold only in inertial frames. However, sometimes, our frames are not inertial. In this chapter, we are going to study a particular kind of non-inertial frame — a rotating frame. An important rotating frame is the Earth itself, but there are also other examples such as merry-go-rounds.

Motion in rotating frames

Now suppose that S is an inertial frame, and S' is rotating about the z axis with angular velocity $\omega = \dot{\theta}$ with respect to S.

Definition (Angular velocity vector)**.** The ***angular velocity vector*** of a rotating frame is $\boldsymbol{\omega} = \omega\hat{\mathbf{z}}$, where $\hat{\mathbf{z}}$ is the axis of rotation and ω is the angular speed.

First we wish to relate the basis vectors $\{\mathbf{e}_i\}$ and $\{\mathbf{e}'_i\}$ of S and S' respectively.

Consider a particle at rest in $\mathbf{S}'$. From the perspective of S, its velocity is

$$\left(\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right)_S = \boldsymbol{\omega} \times \mathbf{r},$$

where $\boldsymbol{\omega} = \omega\hat{\mathbf{z}}$ is the ***angular velocity vector*** (aligned with the rotation axis). This formula also applies to the basis vectors of S'.

$$\left(\frac{\mathrm{d}\mathbf{e}'_i}{\mathrm{d}t}\right)_S = \boldsymbol{\omega} \times \mathbf{e}'_i.$$

Now given a general time-dependent vector $\mathbf{a}$, we can express it in the $\{\mathbf{e}'_i\}$ basis as follows:

$$\mathbf{a} = \sum a'_i(t)\mathbf{e}'_i.$$

From the perspective of S', $\mathbf{e}'_i$ is constant and the time derivative of $\mathbf{a}$ is given by

$$\left(\frac{\mathrm{d}\mathbf{a}}{\mathrm{d}t}\right)_{S'} = \sum \frac{\mathrm{d}a'_i}{\mathrm{d}t}\mathbf{e}'_i.$$

In S, however, $\mathbf{e}'_i$ is not constant. So we apply the product rule to obtain the time derivative of $\mathbf{a}$:

$$\left(\frac{\mathrm{d}\mathbf{a}}{\mathrm{d}t}\right)_S = \sum \frac{\mathrm{d}a_i}{\mathrm{d}t}\mathbf{e}'_i + \sum a'_i\boldsymbol{\omega} \times \mathbf{e}'_i = \left(\frac{\mathrm{d}\mathbf{a}}{\mathrm{d}t}\right)_{S'} + \boldsymbol{\omega} \times \mathbf{a}.$$

This key identity applies to all vectors and can be written as an operator equation:

Proposition. If S is an inertial frame, and S' is rotating relative to S with angular velocity $\boldsymbol{\omega}$, then

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_S = \left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_{S'} + \boldsymbol{\omega} \times .$$

Applied to the position vector $\mathbf{r}(t)$ of a particle, it gives

$$\left(\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right)_S = \left(\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right)_{S'} + \boldsymbol{\omega} \times \mathbf{r}.$$

We can interpret this as saying that the difference in velocity measured in the two frames is the relative velocity of the frames.

We apply this formula a second time, and allow $\boldsymbol{\omega}$ to depend on time. Then we have

$$\begin{aligned}\left(\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\right)_S &= \left(\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_{S'} + \boldsymbol{\omega}\times\right)\left(\left(\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right)_{S'} + \boldsymbol{\omega}\times\mathbf{r}\right).\\ &= \left(\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\right)_{S'} + 2\boldsymbol{\omega}\times\left(\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right)_{S'} + \dot{\boldsymbol{\omega}}\times\mathbf{r} + \boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r})\end{aligned}$$

Since S is inertial, Newton's Second Law is

$$m\left(\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\right)_S = \mathbf{F}.$$

So

Proposition.

$$m\left(\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\right)_{S'} = \mathbf{F} - 2m\boldsymbol{\omega}\times\left(\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right)_{S'} - m\dot{\boldsymbol{\omega}}\times\mathbf{r} - m\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r}).$$

Definition (Fictious forces)**.** The additional terms on the RHS of the equation of motion in rotating frames are ***fictitious forces***, and are needed to explain the motion observed in S'. They are

- ***Coriolis force***: $-2m\boldsymbol{\omega}\times\left(\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right)_{S'}$.
- ***Euler force***: $-m\dot{\boldsymbol{\omega}}\times\mathbf{r}$
- ***Centrifugal force***: $-m\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r})$.

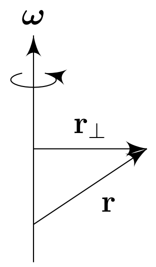
In most cases, $\boldsymbol{\omega}$ is constant and can neglect the Euler force.

The centrifugal force

What exactly does the centrifugal force do? Let $\boldsymbol{\omega} = \omega\hat{\boldsymbol{\omega}}$, where $|\hat{\boldsymbol{\omega}}| = 1$. Then

$$-m\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r}) = -m\big((\boldsymbol{\omega}\cdot\mathbf{r})\boldsymbol{\omega} - (\boldsymbol{\omega}\cdot\boldsymbol{\omega})\mathbf{r}\big) = m\omega^2\mathbf{r}_\perp,$$

where $\mathbf{r}_\perp = \mathbf{r} - (\mathbf{r}\cdot\hat{\boldsymbol{\omega}})\hat{\boldsymbol{\omega}}$ is the projection of the position on the plane perpendicular to $\boldsymbol{\omega}$. So the centrifugal force is directed away from the axis of rotation, and its magnitude is $m\omega^2$ times the distance form the axis.



Note that

$$\begin{aligned}\mathbf{r}_\perp \cdot \mathbf{r}_\perp &= \mathbf{r}\cdot\mathbf{r} - (\mathbf{r}\cdot\hat{\boldsymbol{\omega}})^2\\ \nabla(|\mathbf{r}_\perp|^2) &= 2\mathbf{r} - 2(\mathbf{r}\cdot\hat{\boldsymbol{\omega}})\hat{\boldsymbol{\omega}} = 2\mathbf{r}_\perp.\end{aligned}$$

So

$$-m\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r}) = -\nabla\left(-\frac{1}{2}m\omega^2|\mathbf{r}_\perp|^2\right) = -\nabla\left(-\frac{1}{2}m|\boldsymbol{\omega}\times\mathbf{r}|^2\right).$$

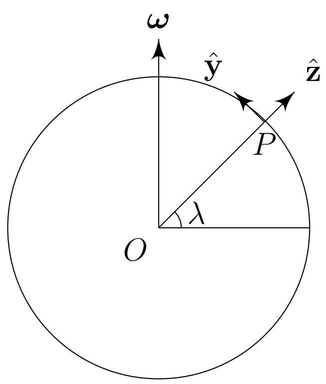
Thus the centrifugal force is a conservative (fictitious) force.

On a rotating planet, the gravitational and centrifugal forces per unit mass combine to make the ***effective gravity***,

$$\mathbf{g}_{\text{eff}} = \mathbf{g} + \omega^2\mathbf{r}_\perp.$$

This gravity will not be vertically downwards. Consider a point P at latitude λ on the surface of a spherical planet of radius R.

We construct orthogonal axes:



with $\hat{\mathbf{x}}$ into the page. So $\hat{\mathbf{z}}$ is up, $\hat{\mathbf{y}}$ is North, and $\hat{\mathbf{x}}$ is East.

At P, we have

$$\begin{aligned}\mathbf{r} &= R\hat{\mathbf{z}}\\ \mathbf{g} &= -g\hat{\mathbf{z}}\\ \boldsymbol{\omega} &= \omega(\cos\lambda\hat{\mathbf{y}} + \sin\lambda\hat{\mathbf{z}})\end{aligned}$$

So

$$\begin{aligned}\mathbf{g}_{\text{eff}} &= \mathbf{g} + \omega^2\mathbf{r}_\perp\\ &= -g\hat{\mathbf{z}} + \omega^2 R\cos\lambda(\cos\lambda\hat{\mathbf{z}} - \sin\lambda\hat{\mathbf{y}})\\ &= -\omega^2 R\cos\lambda\sin\lambda\hat{\mathbf{y}} - (g - \omega^2 R\cos^2\lambda)\hat{\mathbf{z}}.\end{aligned}$$

So the angle α between $\mathbf{g}$ and $\mathbf{g}_{\text{eff}}$ is given by

$$\tan\alpha = \frac{\omega^2 R\cos\lambda\sin\lambda}{g - \omega^2 R\cos^2\lambda}.$$

This is 0 at the equator and the poles, and greatest when you are halfway between. However, this is still tiny on Earth, and does not affect our daily life.

The Coriolis force

The Coriolis force is a more subtle force. Writing $\mathbf{v} = \left(\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right)_{S'}$, we can write the force as

$$\mathbf{F} = -2m\boldsymbol{\omega} \times \mathbf{v}.$$

Note that this has the same form as the Lorentz force caused by a magnetic field, and is velocity-dependent. However, unlike the effects of a magnetic field, particles do not go around in circles in a rotating frame, since we also have the centrifugal force in play.

Since this force is always perpendicular to the velocity, it does no work.

Consider motion parallel to the Earth's surface. We only care about the effect of the Coriolis force on the horizontal trajectory, and ignore the vertical component that is tiny compared to gravity.

So we only take the vertical component of $\boldsymbol{\omega}$, $\omega \sin\lambda\hat{\mathbf{z}}$. The horizontal velocity $\mathbf{v} = v_x\hat{\mathbf{x}}+v_y\hat{\mathbf{y}}$ generates a horizontal Coriolis force:

$$-2m\omega\sin\lambda\hat{\mathbf{z}} \times \mathbf{v} = -2m\omega\sin\lambda(v_y\hat{\mathbf{x}} - v_x\hat{\mathbf{y}}).$$

In the Northern hemisphere ($0 < \lambda < \pi/2$), this causes a deflection towards the right. In the Southern Hemisphere, the deflection is to the left. The effect vanishes at the equator.

Note that only the horizontal effect of horizontal motion vanishes at the equator. The vertical effects or those caused by vertical motion still exist.

Example. Suppose a ball is dropped from a tower of height h at the equator. Where does it land?

In the rotating frame,

$$\ddot{\mathbf{r}} = \mathbf{g} - 2\boldsymbol{\omega} \times \dot{\mathbf{r}} - \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}).$$

We work to first order in ω. Then

$$\ddot{\mathbf{r}} = \mathbf{g} - 2\boldsymbol{\omega} \times \dot{\mathbf{r}} + O(\omega^2).$$

Integrate wrt t to obtain

$$\dot{\mathbf{r}} = \mathbf{g}t - 2\boldsymbol{\omega} \times (\mathbf{r} - \mathbf{r}_0) + O(\omega^2),$$

where $\mathbf{r}_0$ is the initial position. We substitute into the original equation to obtain

$$\ddot{\mathbf{r}} = \mathbf{g} - 2\boldsymbol{\omega} \times \mathbf{g}t + O(\omega^2).$$

(where some new ω^2 terms are thrown into $O(\omega^2)$). We integrate twice to obtain

$$\mathbf{r} = \mathbf{r}_0 + \frac{1}{2}\mathbf{g}t^2 - \frac{1}{3}\boldsymbol{\omega} \times \mathbf{g}t^3 + O(\omega^2).$$

In components, we have $\mathbf{g} = (0, 0, -g)$, $\boldsymbol{\omega} = (0, \omega, 0)$ and $\mathbf{r}_0 = (0, 0, R + h)$. So

$$\mathbf{r} = \left(\frac{1}{3}\omega g t^3, 0, R + h - \frac{1}{2}gt^2\right) + O(\omega^2).$$

So the particle hits the ground at $t = \sqrt{2h/g}$, and its eastward displacement is $\frac{1}{3}wg\left(\frac{2h}{g}\right)^{3/2}$.

This can be understood in terms of angular momentum conservation in the non-rotating frame. At the beginning, the particle has the same angular velocity with the Earth. As it falls towards the Earth, to maintain the same angular momentum, the angular velocity has to increase to compensate for the decreased radius. So it spins faster than the Earth and drifts towards the East, relative to the Earth.

Example. Consider a pendulum that is free to swing in any plane, e.g. a weight on a string. At the North pole, it will swing in a plane that is fixed in an inertial frame, while the Earth rotates beneath it. From the perspective of the rotating frame, the plane of the pendulum rotates backwards. This can be explained as a result of the Coriolis force.

In general, at latitude λ, the plane rotates rightwards with period $\frac{1 \text{ day}}{\sin \lambda}$.

2.7 Systems of particles

Now suppose we have N interacting particles. We adopt the following notation: particle i has mass m_i, position $\mathbf{r}_i$, and momentum $\mathbf{p}_i = m_i\dot{\mathbf{r}}_i$. Note that the subscript denotes which particle it is referring to, not vector components.

Newton's Second Law for particle i is

$$m_i\ddot{\mathbf{r}}_i = \dot{\mathbf{p}}_i = \mathbf{F}_i,$$

where $\mathbf{F}_i$ is the total force acting on particle i. We can write $\mathbf{F}_i$ as

$$\mathbf{F}_i = \mathbf{F}_i^{\text{ext}} + \sum_{j=1}^{N} \mathbf{F}_{ij},$$

where $\mathbf{F}_{ij}$ is the force on particle i by particle j, and $\mathbf{F}_i^{\text{ext}}$ is the external force on i, which comes from particles outside the system.

Since a particle cannot exert a force on itself, we have $\mathbf{F}_{ii} = \mathbf{0}$. Also, Newton's third law requires that

$$\mathbf{F}_{ij} = -\mathbf{F}_{ji}.$$

For example, if the particles interact only via gravity, then we have

$$\mathbf{F}_{ij} = -\frac{Gm_im_j(\mathbf{r}_i - \mathbf{r}_j)}{|\mathbf{r}_i - \mathbf{r}_j|^3} = -\mathbf{F}_{ji}.$$

Motion of the center of mass

Sometimes, we are interested in the collection of particles as a whole. For example, if we treat a cat as a collection of particles, we are more interested in how the cat as a whole falls, instead of tracking the individual particles of the cat.

Hence, we define some aggregate quantities of the system such as the total mass and investigate how these quantities relate.

Definition (Total mass)**.** The ***total mass*** of the system is $M = \sum m_i$.

Definition (Center of mass)**.** The ***center of mass*** is located at

$$\mathbf{R} = \frac{1}{M}\sum_{i=1}^{N} m_i\mathbf{r}_i.$$

This is the mass-weighted average position.

Definition (Total linear momentum)**.** The ***total linear momentum*** is

$$\mathbf{P} = \sum_{i=1}^{N} \mathbf{p}_i = \sum_{i=1}^{N} m_i\dot{\mathbf{r}}_i = M\dot{\mathbf{R}}.$$

Note that this is equivalent to the momentum of a single particle of mass M at the center of mass.

Definition (Total external force)**.** The ***total external force*** is

$$\mathbf{F} = \sum_{i=1}^{N} \mathbf{F}_i^{\text{ext}}.$$

We can now obtain the equation of motion of the center of mass:

Proposition.

$$M\ddot{\mathbf{R}} = \mathbf{F}.$$

Proof.

$$\begin{aligned}
M\ddot{\mathbf{R}} &= \dot{\mathbf{P}} \\
&= \sum_{i=1}^{N} \dot{\mathbf{p}}_i \\
&= \sum_{i=1}^{N} \mathbf{F}_i^{\text{ext}} + \sum_{i=1}^{N}\sum_{j=1}^{N} \mathbf{F}_{ij} \\
&= \mathbf{F} + \frac{1}{2}\sum_i \sum_j (\mathbf{F}_{ij} + \mathbf{F}_{ji}) \\
&= \mathbf{F}
\end{aligned}$$

□

This means that if we don't care about the internal structure, we can treat the system as a point particle of mass M at the center of mass $\mathbf{R}$. This is why Newton's Laws apply to macroscopic objects even though they are not individual particles.

Law (Conservation of momentum)**.** If there is no external force, i.e. $\mathbf{F} = \mathbf{0}$, then $\dot{\mathbf{P}} = \mathbf{0}$. So the total momentum is conserved.

If there is no external force, then the center of mass moves uniformly in a straight line. In this case, we can pick a particularly nice frame of reference, known as the ***center of mass frame***.

Definition (Center of mass frame)**.** The ***center of mass frame*** is an inertial frame in which $\mathbf{R} = 0$ for all time.

Doing calculations in the center of mass frame is usually much more convenient than using other frames,

After doing linear motion, we can now look at angular motion.

Definition (Total angular momentum)**.** The ***total angular momentum*** of the system about the origin is

$$\mathbf{L} = \sum_i \mathbf{r}_i \times \mathbf{p}_i.$$

How does the total angular momentum change with time? Here we have to assume a stronger version of Newton's Third law, saying that

$$\mathbf{F}_{ij} = -\mathbf{F}_{ji} \text{ and is parallel to } (\mathbf{r}_i - \mathbf{r}_j).$$

This is true, at least, for gravitational and electrostatic forces.

Then we have

$$\begin{aligned}
\dot{\mathbf{L}} &= \sum_i \mathbf{r}_i \times \dot{\mathbf{p}}_i + \dot{\mathbf{r}}_i \times \mathbf{p}_i \\
&= \sum_i \mathbf{r}_i \times \left(\mathbf{F}_i^{\text{ext}} + \sum_j F_{ij}\right) + m(\dot{\mathbf{r}}_i \times \dot{\mathbf{r}}_i) \\
&= \sum_i \mathbf{r}_i \times \mathbf{F}_i^{\text{ext}} + \sum_i \sum_j \mathbf{r}_i \times \mathbf{F}_{ij} \\
&= \sum_i \mathbf{G}_i^{\text{ext}} + \frac{1}{2} \sum_i \sum_j (\mathbf{r}_i \times \mathbf{F}_{ij} + \mathbf{r}_j \times \mathbf{F}_{ji}) \\
&= \mathbf{G} + \frac{1}{2} \sum_i \sum_j (\mathbf{r}_i - \mathbf{r}_j) \times \mathbf{F}_{ij} \\
&= \mathbf{G},
\end{aligned}$$

where

Definition (Total external torque)**.** The ***total external torque*** is

$$\mathbf{G} = \sum_i \mathbf{r}_i \times \mathbf{F}_i^{\text{ext}}.$$

So the total angular momentum is conserved if $\mathbf{G} = \mathbf{0}$, ie the total external torque vanishes.

Motion relative to the center of mass

So far, we have shown that externally, a multi-particle system behaves as if it were a point particle at the center of mass. But internally, what happens to the individual particles themselves?

We write $\mathbf{r}_i = \mathbf{R} + \mathbf{r}_i^c$, where $\mathbf{r}_i^c$ is the position of particle i relative to the center of mass.

We first obtain two useful equalities:

$$\sum_i m_i \mathbf{r}_i^c = \sum m_i \mathbf{r}_i - \sum m_i \mathbf{R} = M\mathbf{R} - M\mathbf{R} = \mathbf{0}.$$

Differentiating gives

$$\sum_i m_i \dot{\mathbf{r}}_i^c = \mathbf{0}.$$

Using these equalities, we can express the angular momentum and kinetic energy in terms of $\mathbf{R}$ and $\mathbf{r}_i^c$ only:

$$\begin{aligned}
\mathbf{L} &= \sum_i m_i(\mathbf{R}+\mathbf{r}_i^c)\times(\dot{\mathbf{R}}+\dot{\mathbf{r}}_i^c) \\
&= \sum_i m_i\mathbf{R}\times\dot{\mathbf{R}} + \mathbf{R}\times\sum_i m_i\dot{\mathbf{r}}_i^c + \sum_i m_i\mathbf{r}_i^c\times\dot{\mathbf{R}} + \sum_i m_i\mathbf{r}_i^c\times\dot{\mathbf{r}}_i^c \\
&= M\mathbf{R}\times\dot{\mathbf{R}} + \sum_i m_i\mathbf{r}_i^c\times\dot{\mathbf{r}}_i^c \\
T &= \frac{1}{2}\sum_i m_i|\dot{\mathbf{r}}_i|^2 \\
&= \frac{1}{2}\sum_i m_I(\dot{\mathbf{R}}+\dot{\mathbf{r}_i}^c)\cdot(\dot{\mathbf{R}}+\dot{\mathbf{r}}_i^c) \\
&= \frac{1}{2}\sum_i m_i\dot{\mathbf{R}}\cdot\dot{\mathbf{R}} + \dot{\mathbf{R}}\cdot\sum_i m_i\dot{\mathbf{r}}_i^c + \frac{1}{2}\sum_i m_i\dot{\mathbf{r}}_i^c\cdot\dot{\mathbf{r}}_i^c \\
&= \frac{1}{2}M|\dot{\mathbf{R}}|^2 + \frac{1}{2}\sum_i m_i|\dot{\mathbf{r}}_i^c|^2
\end{aligned}$$

We see that each item is composed of two parts — that of the center of mass and that of motion relative to center of mass.

If the forces are conservative in the sense that

$$\mathbf{F}_i^{\text{ext}} = -\nabla_i V_i(\mathbf{r}_i),$$

and

$$\mathbf{F}_{ij} = -\nabla_i V_{ij}(\mathbf{r}_i-\mathbf{r}_j),$$

where ∇_i is the gradient with respect to $\mathbf{r}_i$, then energy is conserved in the from

$$E = T + \sum_i V_i(\mathbf{r}_i) + \frac{1}{2}\sum_i\sum_j V_{ij}(\mathbf{r}_i-\mathbf{r}_j) = \text{const}.$$

The two-body problem

The ***two-body problem*** is to determine the motion of two bodies interacting only via gravitational forces.

The center of mass is at

$$\mathbf{R} = \frac{1}{M}(m_1\mathbf{r}_1 + m_2\mathbf{r}_2),$$

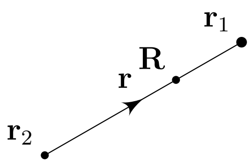
where $M = m_1 + m_2$.

The magic trick to solving the two-body problem is to define the separation vector (or relative position vector)

$$\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2.$$

Then we write everything in terms of $\mathbf{R}$ and $\mathbf{r}$.

$$\mathbf{r}_1 = \mathbf{R} + \frac{m_2}{M}\mathbf{r}, \quad \mathbf{r}_2 = \mathbf{R} - \frac{m_1}{M}\mathbf{r}.$$



Since the external force $\mathbf{F} = \mathbf{0}$, we have $\ddot{\mathbf{R}} = \mathbf{0}$, i.e. the center of mass moves uniformly. Meanwhile,

$$\begin{aligned}\ddot{\mathbf{r}} &= \ddot{\mathbf{r}}_1 - \ddot{\mathbf{r}}_2\\ &= \frac{1}{m_1}\mathbf{F}_{12} - \frac{1}{m_2}\mathbf{F}_{21}\\ &= \left(\frac{1}{m_1} + \frac{1}{m_2}\right)\mathbf{F}_{12}\end{aligned}$$

We can write this as

$$\mu\ddot{\mathbf{r}} = \mathbf{F}_{12}(\mathbf{r}),$$

where

$$\mu = \frac{m_1 m_2}{m_1 + m_2}$$

is the ***reduced mass***. This is the same as the equation of motion for ***one particle*** of mass μ with position vector $\mathbf{r}$ relative to a fixed origin — as studied previously.

For example, with gravity,

$$\mu\ddot{\mathbf{r}} = -\frac{G m_1 m_2 \hat{\mathbf{r}}}{|\mathbf{r}|^2}.$$

So

$$\ddot{\mathbf{r}} = -\frac{GM\hat{\mathbf{r}}}{|\mathbf{r}|^2}.$$

For example, give a planet orbiting the Sun, both the planet and the sun moves in ellipses about their center of mass. The orbital period depends on the total mass.

It can be shown that

$$\begin{aligned}\mathbf{L} &= M\mathbf{R}\times\dot{\mathbf{R}} + \mu\mathbf{r}\times\dot{\mathbf{r}}\\ T &= \frac{1}{2}M|\dot{\mathbf{R}}|^2 + \frac{1}{2}\mu|\dot{\mathbf{r}}|^2\end{aligned}$$

by expressing $\mathbf{r}_1$ and $\mathbf{r}_2$ in terms of $\mathbf{r}$ and $\mathbf{R}$.

Variable-mass problem

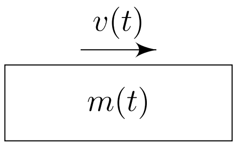
All systems we've studied so far have fixed mass. However, in real life, many objects have changing mass, such as rockets, fireworks, falling raindrops and rolling snowballs.

Again, we will use Newton's second law, which states that

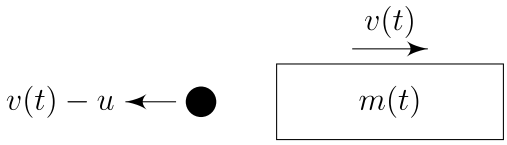
$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = \mathbf{F}, \quad \text{with } \mathbf{p} = m\dot{\mathbf{r}}.$$

We will consider a rocket moving in one dimension with mass $m(t)$ and velocity $v(t)$. The rocket propels itself forwards by burning fuel and ejecting the exhaust at velocity $-u$ relative to the rocket.

At time t, the rocket looks like this:



At time $t + \delta t$, it ejects exhaust of mass $m(t) - m(t + \delta t)$ with velocity $v(t) - u + O(\delta t)$.



The change in total momentum of the system (rocket + exhaust) is

$$\begin{aligned}
\delta p &= m(t+\delta t)v(t+\delta t) + [m(t) - m(t+\delta t)][v(t) - u(t) + O(\delta t)] - m(t)v(t)\\
&= (m + \dot{m}\delta t + O(\delta t^2))(v + \dot{v}\delta t + O(\delta t^2)) - \dot{m}\delta t(v - u) + O(\delta t^2) - mv\\
&= (\dot{m}v + m\dot{v} - \dot{m}v + \dot{m}u)\delta t + O(\delta t^2)\\
&= (m\dot{v} + \dot{m}u)\delta t + O(\delta t^2).
\end{aligned}$$

Newton's second law gives

$$\lim_{\delta \to 0} \frac{\delta p}{\delta t} = F$$

where F is the external force on the rocket. So we obtain

Proposition (Rocket equation)**.**

$$m\frac{\mathrm{d}v}{\mathrm{d}t} + u\frac{\mathrm{d}m}{\mathrm{d}t} = F.$$

Example. Suppose that we travel in space with $F = 0$. Assume also that u is constant. Then we have

$$m\frac{\mathrm{d}v}{\mathrm{d}t} = -u\frac{\mathrm{d}m}{dt}.$$

So

$$v = v_0 + u\log\left(\frac{m_0}{m(t)}\right),$$

Note that we are expressing things in terms of the mass remaining m, not time t.

Note also that the velocity does not depend on the rate at which mass is ejected, only the velocity at which it is ejected. Of course, if we expressed v as a function of time, then the velocity at a specific time ***does*** depend on the rate at which mass is ejected.

Example. Consider a falling raindrop of mass $m(t)$, gathering mass from a stationary cloud. In this case, $u = v$. So

$$m\frac{\mathrm{d}v}{\mathrm{d}t} + v\frac{\mathrm{d}m}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}(mv) = mg,$$

with v measured downwards. To obtain a solution of this, we will need a model to determine the rate at which the raindrop gathers mass.

2.8 Rigid bodies

This chapter is somewhat similar to the previous chapter. We again have a lot of particles and we study their motion. However, instead of having forces between the individual particles, this time the particles are constrained such that their relative positions are fixed. This corresponds to a solid object that cannot deform. We call these ***rigid bodies***.

Definition (Rigid body)**.** A ***rigid body*** is an extended object, consisting of N particles that are constrained such that the distance between any pair of particles, $|\mathbf{r}_i - \mathbf{r}_j|$, is fixed.

The possible motions of a rigid body are the continuous isometries of Euclidean space, i.e. translations and rotations. However, as we have previously shown, pure translations of rigid bodies are uninteresting — they simply correspond to the center of mass moving under an external force. Hence we will first study rotations.

Later, we will combine rotational and translational effects and see what happens.

Angular velocity

We'll first consider the cases where there is just one particle, moving in a circle of radius s about the z axis. Its position and velocity vectors are

$$\begin{aligned}\mathbf{r} &= (s\cos\theta, s\sin\theta, z)\\ \dot{\mathbf{r}} &= (-s\dot{\theta}\sin\theta, s\dot{\theta}\cos\theta, 0).\end{aligned}$$

We can write

$$\dot{\mathbf{r}} = \boldsymbol{\omega}\times\mathbf{r},$$

where

$$\boldsymbol{\omega} = \dot{\theta}\hat{\mathbf{z}}$$

is the angular velocity vector.

In general, we write

$$\boldsymbol{\omega} = \dot{\theta}\hat{\mathbf{n}} = \omega\hat{\mathbf{n}},$$

where $\hat{\mathbf{n}}$ is a unit vector parallel to the rotation axis.

The kinetic energy of this particle is thus

$$\begin{aligned}T &= \frac{1}{2}m|\dot{\mathbf{r}}|^2\\ &= \frac{1}{2}ms^2\dot{\theta}^2\\ &= \frac{1}{2}I\omega^2\end{aligned}$$

where $I = ms^2$ is the ***moment of inertia***. This is the counterpart of "mass" in rotational motion.

Definition (Moment of inertia)**.** The ***moment of inertia*** of a particle is

$$I = ms^2 = m|\hat{\mathbf{n}}\times\mathbf{r}|^2,$$

where s is the distance of the particle from the axis of rotation.

Moment of inertia

In general, consider a rigid body in which all N particles rotate about the same axis with the same angular velocity:

$$\dot{\mathbf{r}}_i = \boldsymbol{\omega} \times \mathbf{r}_i.$$

This ensures that

$$\frac{\mathrm{d}}{\mathrm{d}t}|\mathbf{r}_i - \mathbf{r}_j|^2 = 2(\dot{\mathbf{r}}_i - \dot{\mathbf{r}}_j) \cdot (\mathbf{r}_i - \mathbf{r}_j) = 2\big(\boldsymbol{\omega} \times (\mathbf{r}_i - \mathbf{r}_j)\big) \cdot (\mathbf{r}_i - \mathbf{r}_j) = 0,$$

as required for a rigid body.

Similar to what we had above, the rotational kinetic energy is

$$T = \frac{1}{2}\sum_{i=1}^{N} m_i|\dot{\mathbf{r}}_i|^2 = \frac{1}{2}I\omega^2,$$

where

Definition (Moment of inertia)**.** The ***moment of inertia*** of a rigid body about the rotation axis $\hat{\mathbf{n}}$ is

$$I = \sum_{i=1}^{N} m_i s_i^2 = \sum_{i=1}^{N} m_i|\hat{\mathbf{n}} \times \mathbf{r}_i|^2.$$

Again, we define the angular momentum of the system:

Definition. The ***angular momentum*** is

$$\mathbf{L} = \sum_i m_i \mathbf{r}_i \times \dot{\mathbf{r}}_i = \sum_i m_i \mathbf{r}_i \times (\boldsymbol{\omega} \times \mathbf{r}_i).$$

Note that our definitions for angular motion are analogous to those for linear motion. The moment of inertia I is defined such that $T = \frac{1}{2}I\omega^2$. Ideally, we would want the momentum to be $\mathbf{L} = I\boldsymbol{\omega}$. However, this not true. In fact, $\mathbf{L}$ need not be parallel to $\boldsymbol{\omega}$.

What ***is*** true, is that the component of $\mathbf{L}$ parallel to $\boldsymbol{\omega}$ is equal to $I\omega$. Write $\boldsymbol{\omega} = \omega\hat{\mathbf{n}}$. Then we have

$$\begin{aligned}
\mathbf{L} \cdot \hat{\mathbf{n}} &= \omega \sum_i m_i \hat{\mathbf{n}} \cdot (\mathbf{r}_i \times (\hat{\mathbf{n}} \times \mathbf{r}_i)) \\
&= \omega \sum_i m(\hat{\mathbf{n}} \times \mathbf{r}_i) \cdot (\hat{\mathbf{n}} \times \mathbf{r}_i) \\
&= I\omega.
\end{aligned}$$

What does $\mathbf{L}$ itself look like? Using vector identities, we have

$$\mathbf{L} = \sum_i m_i\big((\mathbf{r}_i \cdot \mathbf{r}_i)\boldsymbol{\omega} - (\mathbf{r}_i \cdot \boldsymbol{\omega})\mathbf{r}_i\big)$$

Note that this is a linear function of $\boldsymbol{\omega}$. So we can write

$$\mathbf{L} = I\boldsymbol{\omega},$$

where we abuse notation to use I for the ***inertia tensor***. This is represented by a symmetric matrix with components

$$I_{jk} = \sum_i m_i(|\mathbf{r}_i|^2\delta_{jk} - (\mathbf{r}_i)_j(\mathbf{r}_i)_k),$$

where i refers to the index of the particle, and j, k are dummy suffixes.

If the body rotates about a ***principal axis***, i.e. one of the three orthogonal eigenvectors of I, then $\mathbf{L}$ will be parallel to $\boldsymbol{\omega}$. Usually, the principal axes lie on the axes of rotational symmetry of the body.

Calculating the moment of inertia

For a solid body, we usually want to think of it as a continuous substance with a mass density, instead of individual point particles. So we replace the sum of particles by a volume integral weighted by the mass density $\rho(\mathbf{r})$.

Definition (Mass, center of mass and moment of inertia)**.** The ***mass*** is

$$M = \int \rho \, \mathrm{d}V.$$

The ***center of mass*** is

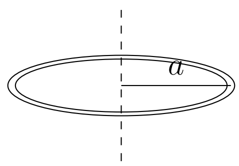
$$\mathbf{R} = \frac{1}{M}\int \rho\mathbf{r} \, \mathrm{d}V$$

The ***moment of inertia*** is

$$I = \int \rho s^2 \, \mathrm{d}V = \int \rho|\hat{\mathbf{n}} \times \mathbf{r}|^2 \, \mathrm{d}V.$$

In theory, we can study inhomogeneous bodies with varying ρ, but usually we mainly consider homogeneous ones with constant ρ throughout.

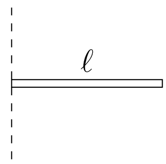
Example (Thin circular ring)**.** Suppose the ring has mass M and radius a, and a rotation axis through the center, perpendicular to the plane of the ring.



Then the moment of inertia is

$$I = Ma^2.$$

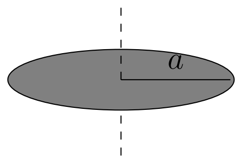
Example (Thin rod)**.** Suppose a rod has mass M and length ℓ. It rotates through one end, perpendicular to the rod.



The mass per unit length is M/ℓ. So the moment of inertia is

$$I = \int_0^\ell \frac{M}{\ell} x^2 \, \mathrm{d}x = \frac{1}{3} M\ell^2.$$

Example (Thin disc)**.** Consider a disc of mass M and radius a, with a rotation axis through the center, perpendicular to the plane of the disc.



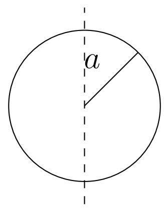
Then

$$\begin{aligned} I &= \int_0^{2\pi} \int_0^a \underbrace{\frac{M}{\pi a^2}}_{\text{mass per unit length}} \underbrace{r^2}_{s^2} \underbrace{r \, \mathrm{d}r \, \mathrm{d}\theta}_{\text{area element}} \\ &= \frac{M}{\pi a^2} \int_0^a r^3 \, \mathrm{d}r \int_0^{2\pi} \mathrm{d}\theta \\ &= \frac{M}{\pi a^2} \frac{1}{4} a^4 (2\pi) \\ &= \frac{1}{2} M a^2. \end{aligned}$$

Now suppose that the rotation axis is in the plane of the disc instead (also rotating through the center). Then

$$\begin{aligned} I &= \int_0^{2\pi} \int_0^a \underbrace{\frac{M}{\pi a^2}}_{\text{mass per unit length}} \underbrace{(r \sin\theta)^2}_{s^2} \underbrace{r \, \mathrm{d}r \, \mathrm{d}\theta}_{\text{area element}} \\ &= \frac{M}{\pi a^2} \int_0^a r^3 \, \mathrm{d}r \int_0^{2\pi} \sin^2\theta \, \mathrm{d}\theta \\ &= \frac{M}{\pi a^2} \frac{1}{4} a^4 \pi \\ &= \frac{1}{4} M a^2. \end{aligned}$$

Example. Consider a solid sphere with mass M, radius a, with a rotation axis though the center.



Using spherical polar coordinates (r, θ, ϕ) based on the rotation axis,

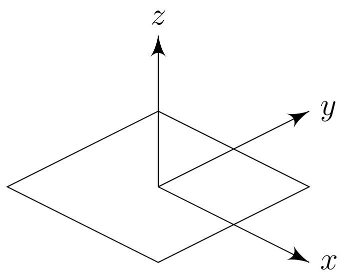
$$\begin{aligned}
I &= \int_0^{2\pi}\int_0^{\pi}\int_0^{a} \underbrace{\frac{M}{\frac{4}{3}\pi a^3}}_{\rho} \underbrace{(r\sin\theta)^2}_{s^2} \underbrace{r^2 \sin\theta \;\mathrm{d}r\;\mathrm{d}\theta\;\mathrm{d}\phi}_{\text{volume element}}\\
&= \frac{M}{\frac{4}{3}\pi a^3}\int_0^a r^4\;\mathrm{d}r \int_0^{\pi}(1-\cos^2)\sin\theta\;\mathrm{d}\theta\int_0^{2\pi}\;\mathrm{d}\phi\\
&= \frac{M}{\frac{4}{3}\pi a^3}\cdot\frac{1}{5}a^5\cdot\frac{4}{3}\cdot 2\pi\\
&= \frac{2}{5}Ma^2.
\end{aligned}$$

Usually, finding the moment of inertia involves doing complicated integrals. We will now come up with two theorems that help us find moments of inertia.

Theorem (Perpendicular axis theorem)**.** For a two-dimensional object (a lamina), and three perpendicular axes x, y, z through the same spot, with z normal to the plane,

$$I_z = I_x + I_y,$$

where I_z is the moment of inertia about the z axis.



Note that this does ***not*** apply to 3D objects! For example, in a sphere, $I_x = I_y = I_z$.

Proof. Let ρ be the mass per unit volume. Then

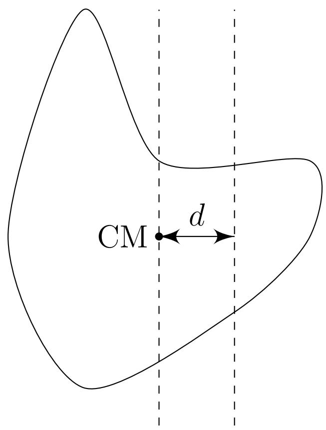
$$\begin{aligned}
I_x &= \int \rho y^2\;\mathrm{d}A\\
I_y &= \int \rho x^2\;\mathrm{d}A\\
I_z &= \int \rho(x^2+y^2)\;\mathrm{d}A = I_x + I_y.
\end{aligned}$$

□

Example. For a disc, $I_x = I_y$ by symmetry. So $I_z = 2I_x$.

Theorem (Parallel axis theorem)**.** If a rigid body of mass M has moment of inertia I^C about an axis passing through the center of mass, then its moment of inertia about a parallel axis a distance d away is

$$I = I^C + Md^2.$$



Proof. With a convenient choice of Cartesian coordinates such that the center of mass is at the origin and the two rotation axes are $x = y = 0$ and $x = d, y = 0$,

$$I^C = \int \rho(x^2 + y^2)\;\mathrm{d}V,$$

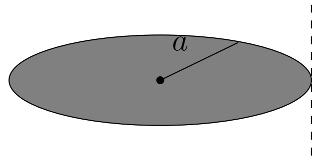
and

$$\int \rho\mathbf{r}\;\mathrm{d}V = \mathbf{0}.$$

So

$$\begin{aligned} I &= \int \rho((x - d)^2 + y^2)\;\mathrm{d}V\\ &= \int \rho(x^2 + y^2)\;\mathrm{d}V - 2d\int \rho x\;\mathrm{d}V + \int d^2\rho\;\mathrm{d}V\\ &= I^c + 0 + Md^2\\ &= I^c + Md^2. \end{aligned}$$ □

Example. Take a disc of mass M and radius a, and rotation axis through a point on the circumference, perpendicular to the plane of the disc. Then



$$I = I^c + Ma^2 = \frac{1}{2}Ma^2 + Ma^2 = \frac{3}{2}Ma^2.$$

Motion of a rigid body

The general motion of a rigid body can be described as a translation of its centre of mass, following a trajectory $\mathbf{R}(t)$, together with a rotation about an axis through the center of mass. As before, we write

$$\mathbf{r}_i = \mathbf{R} + \mathbf{r}_i^c.$$

Then

$$\dot{\mathbf{r}}_i = \dot{\mathbf{R}} + \dot{\mathbf{r}}_i^c.$$

Using this, we can break down the velocity and kinetic energy into translational and rotational parts.

If the body rotates with angular velocity $\boldsymbol{\omega}$ about the center of mass, then

$$\dot{\mathbf{r}}_i^c = \boldsymbol{\omega} \times \mathbf{r}_i^c.$$

Since $\mathbf{r}_i^c = \mathbf{r}_i - \mathbf{R}$, we have

$$\dot{\mathbf{r}}_i = \dot{\mathbf{R}} + \boldsymbol{\omega} \times \mathbf{r}_i^c = \dot{\mathbf{R}} + \boldsymbol{\omega} \times (\mathbf{r}_i - \mathbf{R}).$$

On the other hand, the kinetic energy, as calculated in previous lectures, is

$$\begin{aligned} T &= \frac{1}{2}M|\dot{\mathbf{R}}|^2 + \frac{1}{2}\sum_i m_i|\dot{\mathbf{r}}_i^c|^2 \\ &= \underbrace{\frac{1}{2}M|\dot{\mathbf{R}}|^2}_{\text{translational KE}} + \underbrace{\frac{1}{2}I^c\omega^2}_{\text{rotational KE}}. \end{aligned}$$

Sometimes we do not want to use the center of mass as the center. For example, if an item is held at the edge and spun around, we'd like to study the motion about the point at which the item is held, and not the center of mass.

So consider any point Q, with position vector $\mathbf{Q}(t)$ that is not the center of mass but moves with the rigid body, i.e.

$$\dot{\mathbf{Q}} = \dot{\mathbf{R}} + \boldsymbol{\omega} \times (\mathbf{Q} - \mathbf{R}).$$

Usually this is a point inside the object itself, but we do not assume that in our calculation. Then we can write

$$\begin{aligned} \dot{\mathbf{r}}_i &= \dot{\mathbf{R}} + \boldsymbol{\omega} \times (\mathbf{r}_i - \mathbf{R}) \\ &= \dot{\mathbf{Q}} - \boldsymbol{\omega} \times (\mathbf{Q} - \mathbf{R}) + \boldsymbol{\omega} \times (\mathbf{r}_i - \mathbf{R}) \\ &= \dot{\mathbf{Q}} + \boldsymbol{\omega} \times (\mathbf{r}_i - \mathbf{Q}). \end{aligned}$$

Therefore the motion can be considered as a translation of Q (with ***different*** velocity than the center of mass), together with rotation about Q (with the ***same*** angular velocity $\boldsymbol{\omega}$).

Equations of motion

As shown previously, the linear and angular momenta evolve according to

$$\dot{\mathbf{P}} = \mathbf{F} \quad \text{(total external force)}$$
$$\dot{\mathbf{L}} = \mathbf{G} \quad \text{(total external torque)}$$

These two equations determine the translational and rotational motion of a rigid body.

$\mathbf{L}$ and $\mathbf{G}$ depend on the choice of origin, which could be any point that is fixed in an inertial frame. More surprisingly, it can also be applied to the center of mass, even if this is accelerated: If

$$m_i \ddot{\mathbf{r}}_i = \mathbf{F}_i,$$

then

$$m_i \ddot{\mathbf{r}}_i^c = \mathbf{F}_i + m_i \ddot{\mathbf{R}}.$$

So there is a fictitious force $m_i \ddot{\mathbf{R}}$ in the center-of-mass frame. But the total torque of the fictitious forces about the center of mass is

$$\sum_i \mathbf{r}_i^c \times \left(-m_i \ddot{\mathbf{R}}\right) = -\left(\sum m_i \mathbf{r}_i^c\right) \times \ddot{\mathbf{R}} = \mathbf{0} \times \dot{\mathbf{R}} = 0.$$

So we can still apply the above two equations.

In summary, the laws of motion apply in any inertial frame, or the center of mass (possibly non-inertial) frame.

Motion in a uniform gravitational field

In a uniform gravitational field $\mathbf{g}$, the total gravitational force and torque are the same as those that would act on a single particle of mass M located at the center of mass (which is also the ***center of gravity***):

$$\mathbf{F} = \sum_i \mathbf{F}_i^{\text{ext}} = \sum_i m_i \mathbf{g} = M\mathbf{g},$$

and

$$\mathbf{G} = \sum_i \mathbf{G}_i^{\text{ext}} = \sum_i \mathbf{r}_i \times (m_i \mathbf{g}) = \sum m_i \mathbf{r}_i \times \mathbf{g} = M\mathbf{R} \times \mathbf{g}.$$

In particular, the gravitational torque about the center of mass vanishes: $\mathbf{G}^c = \mathbf{0}$.

We obtain a similar result for gravitational potential energy.

The gravitational potential in a uniform $\mathbf{g}$ is

$$\Phi_g = -\mathbf{r} \cdot \mathbf{g}.$$

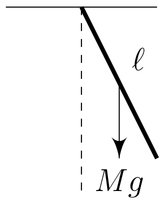
(since $\mathbf{g} = -\nabla \Phi_g$ by definition)

So

$$\begin{aligned} V^{\text{ext}} &= \sum_i V_i^{\text{ext}} \\ &= \sum_i m_i(-\mathbf{r}_i \cdot \mathbf{g}) \\ &= M(-\mathbf{R} \cdot \mathbf{g}). \end{aligned}$$

Example (Thrown stick)**.** Suppose we throw a symmetrical stick. So the center of mass is the actual center. Then the center of the stick moves in a parabola. Meanwhile, the stick rotates with constant angular velocity about its center due to the absence of torque.

Example. Swinging bar.



This is an example of a ***compound pendulum***.

Consider the bar to be rotating about the pivot (and not translating). Its angular momentum is $L = I\dot{\theta}$ with $I = \frac{1}{3}M\ell^2$. The gravitational torque about the pivot is

$$G = -Mg\frac{\ell}{2}\sin\theta.$$

The equation of motion is

$$\dot{L} = G.$$

So

$$I\ddot{\theta} = -Mg\frac{\ell}{2}\sin\theta,$$

or

$$\ddot{\theta} = -\frac{3g}{2\ell}\sin\theta.$$

which is exactly equivalent to a simple pendulum of length $2\ell/3$, with angular frequency $\sqrt{\frac{3g}{2\ell}}$.

This can also be obtained from an energy argument:

$$E = T + V = \frac{1}{2}I\dot{\theta}^2 - Mg\frac{\ell}{2}\cos\theta.$$

We differentiate to obtain

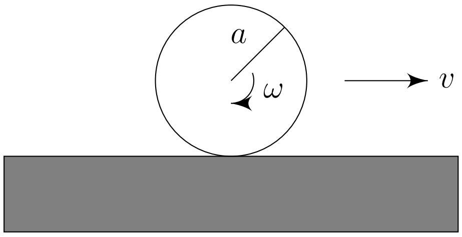
$$\frac{\mathrm{d}E}{\mathrm{d}t} = \dot{\theta}(I\ddot{\theta} + Mg\frac{\ell}{2}\sin\theta) = 0.$$

So

$$I\ddot{\theta} = -Mg\frac{\ell}{2}\sin\theta.$$

Sliding versus rolling

Consider a cylinder or sphere of radius a, moving along a stationary horizontal surface.



In general, the motion consists of a translation of the center of mass (with velocity v) plus a rotation about the center of mass (with angular velocity ω).

The horizontal velocity at the point of contact is

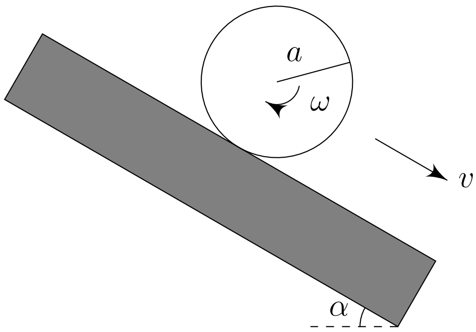
$$v_{\text{slip}} = v - a\omega.$$

For a pure sliding motion, $v \neq 0$ and $\omega = 0$, in which case $v - a\omega \neq 0$: the point of contact moves relative to the surface and kinetic friction may occur.

For a pure rolling motion, $v \neq 0$ and $\omega \neq 0$ such that $v - a\omega = 0$: the point of contact is stationary. This is the no-slip condition.

The rolling body can alternatively be considered to be rotating instantaneously about the point of contact (with angular velocity ω) and not translating.

Example (Rolling down hill)**.**



Consider a cylinder or sphere of mass M and radius a rolling down a rough plane inclined at angle α. The no-slip (rolling) condition is

$$v - a\omega = 0.$$

The kinetic energy is

$$T = \frac{1}{2}Mv^2 + \frac{1}{2}I\omega^2 = \frac{1}{2}\left(M + \frac{I}{a^2}\right)v^2.$$

The total energy is

$$E = \frac{1}{2}\left(M + \frac{I}{a^2}\right)\dot{x}^2 - Mgx\sin\alpha,$$

where x is the distance down slope. While there is a frictional force, the instantaneous velocity is 0, and no work is done. So energy is conserved, and we have

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \dot{x}\left(\left(M + \frac{I}{a^2}\right)\ddot{x} - Mg\sin\alpha\right) = 0.$$

So

$$\left(M + \frac{I}{a^2}\right)\ddot{x} = Mg\sin\alpha.$$

For example, if we have a uniform solid cylinder,

$$I = \frac{1}{2}Ma^2 \quad \text{(as for a disc)}$$

and so

$$\ddot{x} = \frac{2}{3}g\sin\alpha.$$

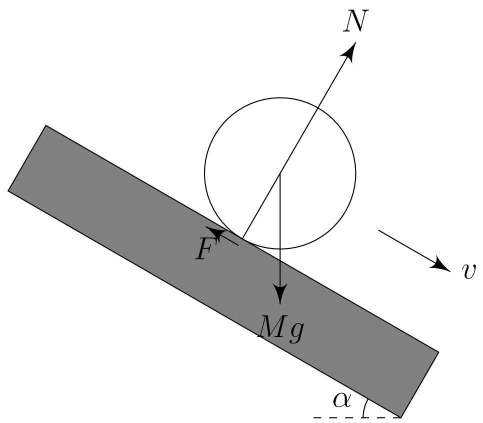
For a thin cylindrical shell,

$$I = Ma^2.$$

So

$$\ddot{x} = \frac{1}{2}g\sin\alpha.$$

Alternatively, we may do it in terms of forces and torques,



The equations of motion are

$$M\dot{v} = Mg\sin\alpha - F$$

and

$$I\dot{\omega} = aF.$$

While rolling,

$$\dot{v} - a\dot{\omega} = 0.$$

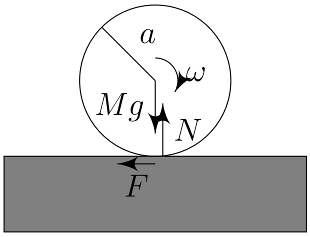
So

$$M\dot{v} = Mg\sin\alpha - \frac{I}{a^2}\dot{v},$$

leading to the same result.

Note that even though there is a frictional force, it does no work, since $v_{\text{slip}} = 0$. So energy is still conserved.

Example (Snooker ball)**.**



It is struck centrally so as to initiate translation, but not rotation. Sliding occurs initially. Intuitively, we think it will start to roll, and we'll see that's the case.

The constant frictional force is

$$F = \mu_k N = \mu_k M g,$$

which applies while $v - a\omega > 0$.

The moment of inertia about the center of mass is

$$I = \frac{2}{5} M a^2.$$

The equations of motion are

$$M\dot{v} = -F$$
$$I\dot{\omega} = aF$$

Initially, $v = v_0$ and $\omega = 0$. Then the solution is

$$v = v_0 - \mu_k g t$$
$$\omega = \frac{5}{2}\frac{\mu_k g}{a} t$$

as long as $v - a\omega > 0$. The slip velocity is

$$v_{\text{slip}} = v - a\omega = v_0 - \frac{7}{2}\mu_k g t = v_0 \left(1 - \frac{t}{t_{\text{roll}}}\right),$$

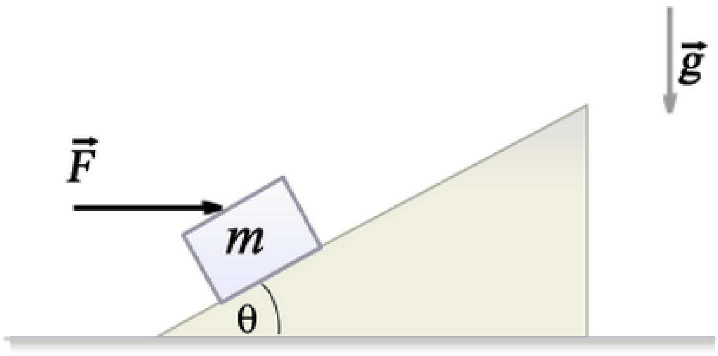
where

$$t_{\text{roll}} = \frac{2v_0}{7\mu_k g}.$$

This is valid up till $t = t_{\text{roll}}$. Then the slip velocity is 0, rolling begins and friction ceases. At this point, $v = a\omega = \frac{5}{7}v_0$. The energy is then $\frac{5}{14}Mv_0^2 < \frac{1}{2}Mv_0^2$. So energy is lost to friction.

2.9 PROBLEMS

1. A block on a frictionless ramp



A block of mass $m = 4kg$ is pressed with a horizontal force F against a frictionless ramp of angle $\theta = 38°$.

Assuming the block is at rest on the ramp, answer the following:

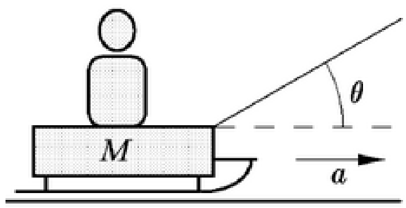
a) What is the magnitude of the normal force exerted by the incline surface on the block?

b) What is the magnitude of the force F exerted on the block?

†198

2. Towing a sled

A mother tows her daughter on a sled on level ice. The friction between the sled and the ice is negligible, and the tow rope makes an angle of θ to the horizontal. The combined mass of the sled and the child is M. The sled has an acceleration in the horizontal direction of magnitude a.



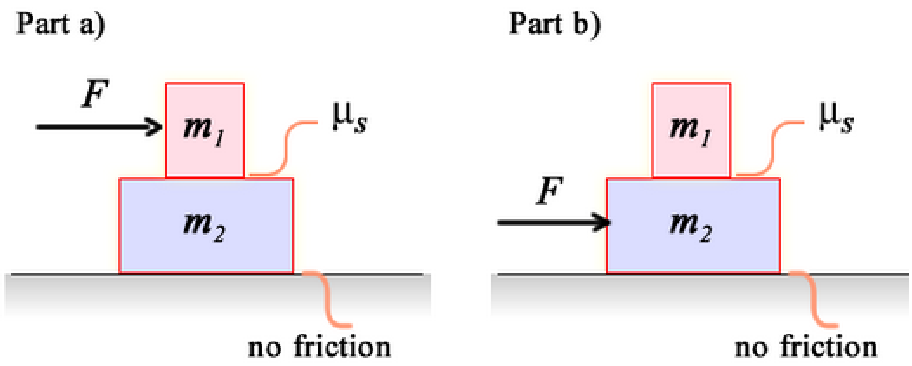
Express the following in terms of M, a, g, and θ.

a) The tension, T, in the rope.

b) The magnitude of the normal force, N, exerted by the ice on the sled.

[†199](#)

3. Stacked blocks



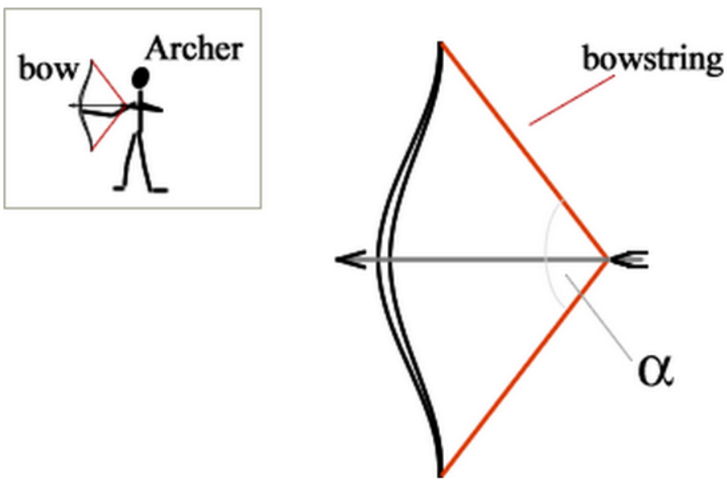
Consider two blocks that are resting one on top of the other. The lower block has mass $m_2 = 4.3\,\text{kg}$ and is resting on a frictionless table. The upper block has mass $m_1 = 1.2\,\text{kg}$. Suppose the coefficient of static friction between the two blocks is given by $\mu_s = 0.6$.

a) A force of magnitude F is applied as shown in the left figure above. What is the maximum force for which the upper block can be pushed horizontally so that the two blocks move together without slipping?

b) A force of magnitude F as shown in the right figure above. What is the maximum force for which the lower block can be pushed horizontally so that the two blocks move together without slipping?'

†199

4. Tension in string



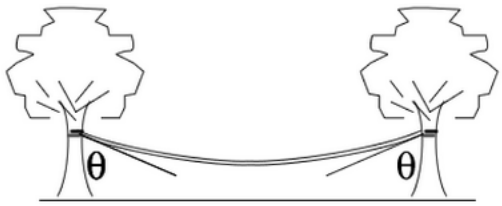
An archer is preparing to shoot an arrow. He grabs the center of the bowstring and pulls straight back with a force of magnitude $F = 118$ N. The upper and lower halves of the string form an angle $\alpha = 124°$ with respect to each other. Assume that the bowstring is massless.

a) What is the magnitude of the tension in the upper half of the bowstring?

†201

5. Rope between trees

Suppose a rope of mass m hangs between two trees. The ends of the rope are at the same height and they make an angle θ with the trees.



Express in terms of m, g, and θ

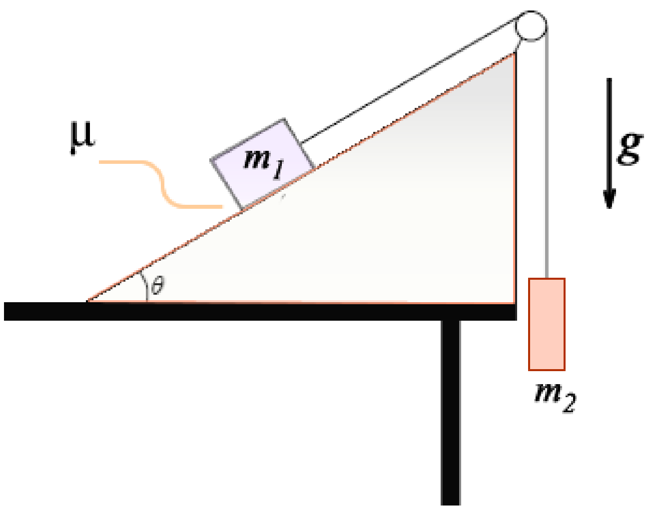
a) The tension at the ends of the rope where it is connected to the trees?

b) The tension in the rope at a point midway between the trees?

†201

6. Blocks and ramp with friction

A block of mass $m_1 = 28\,\mathrm{kg}$ rests on a wedge of angle $\theta = 47°$ which is itself attached to a table (the wedge does not move in this problem). An inextensible string is attached to m_1, passes over a frictionless pulley at the top of the wedge, and is then attached to another block of mass $m_2 = 3\,\mathrm{kg}$. The coefficient of kinetic friction between block 1 and the plane is $\mu = 0.8$. The string and wedge are long enough to ensure neither block hits the pulley or the table in this problem, and you may assume that block 1 never reaches the table. Take g to be $9.81\,\mathrm{m/s^2}$.



The system is released from rest as shown above, at $t = 0$.

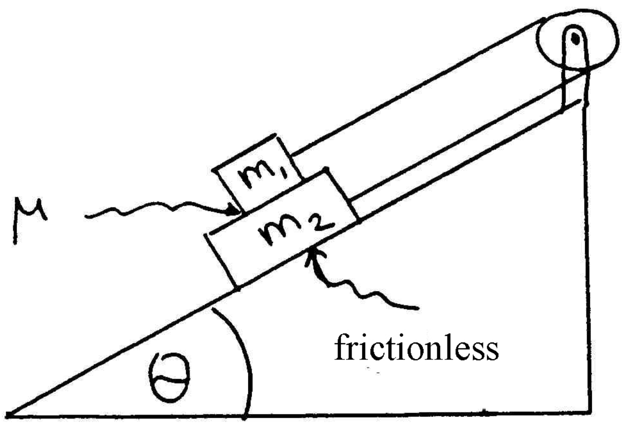
a) Find the magnitude of the acceleration of block 1 when it is released (in m/s^2).

b) How many cm down the plane will block 1 have traveled when 0.47 s have elapsed?'

†201

7. Friction between blocks on a ramp

Two blocks with masses m_1 and m_2 such that $m_1 \mathcal{L} m_2$ are connected by a massless inextensible string and a massless pulley as shown in the figure below. The pulley is rigidly connected to the top of a wedge with angle θ. The coefficient of friction between the blocks is μ. The surface between the lower block and the wedge is frictionless. The goal of this problem is to find the magnitude of the acceleration of each block.

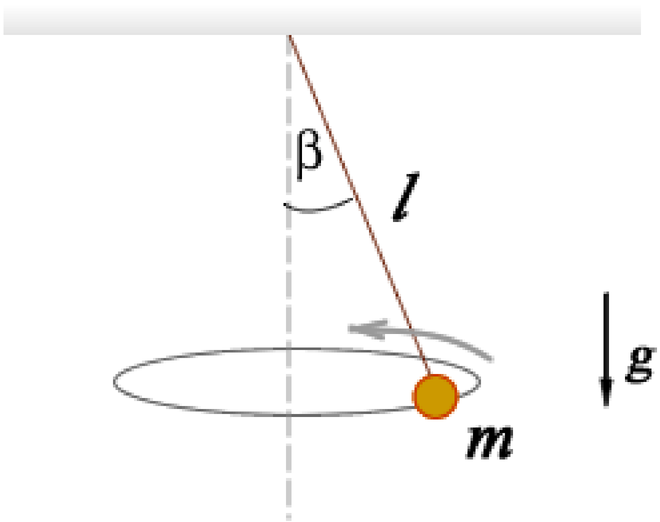


What are the magnitudes of the acceleration of the two blocks? Express your answer in terms of g, μ, m_1, m_2, and θ.

†202

8. Conical pendulum

A conical pendulum is constructed from a rope of length ℓ and negligible mass, which is suspended from a fixed pivot attached to the ceiling. A small ball of mass m is attached to the lower end of the rope. The ball moves in a circle with constant speed in the horizontal plane, while the rope makes an angle β with respect to the vertical, as shown in the diagram.



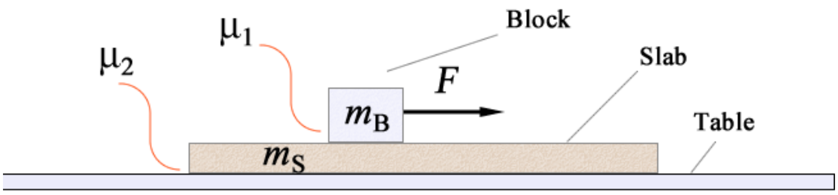
a) Find the tension F_T in the rope. Express your answer in terms of g, m, ℓ, and β.

b) Find the period of the motion (how long does it take the ball to make one circle in the horizontal plane). Express your answer in terms of g, m, ℓ, and β.

†203

9. Stacked blocks 2

A block of mass $m_B = 15\,\text{kg}$ is on top of a long slab of mass $m_S = 9\,\text{kg}$, and the slab is on top of a horizontal table as shown. A horizontal force of magnitude $F = 294\,\text{N}$ is applied on the block. As a result the block moves relative to the slab and the slab moves relative to the table. There is friction between all surfaces. The coefficient of kinetic friction between the block and the slab is $\mu_1 = 0.7$, and the coefficient of kinetic friction between the slab and the table is $\mu_2 = 0.1$. Take g to be $9.81\,\text{m/s}^2$, and enter your answer to 3 significant figures.



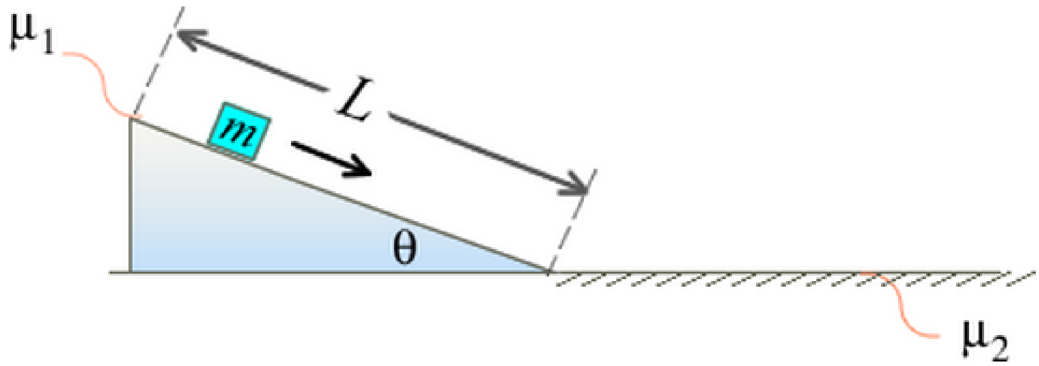
a) What is the magnitude of the block's acceleration?

b) What is the magnitude of the slab's acceleration?

†205

10. Rough surfaces

A block of mass m, starting from rest, slides down an inclined plane of length L and angle θ with respect to the horizontal. The coefficient of kinetic friction between the block and the inclined surface is μ_1. At the bottom of the incline, the block slides along a horizontal and rough surface with a coefficient of kinetic friction μ_2. The goal of this problem is to find out how far the block slides along the rough surface.



a) What is the work done by the friction force on the block while it is sliding down the inclined plane?

b) What is the work done by the gravitational force on the block while it is sliding down the inclined plane?

c) What is the kinetic energy of the block just at the bottom of the inclined plane?

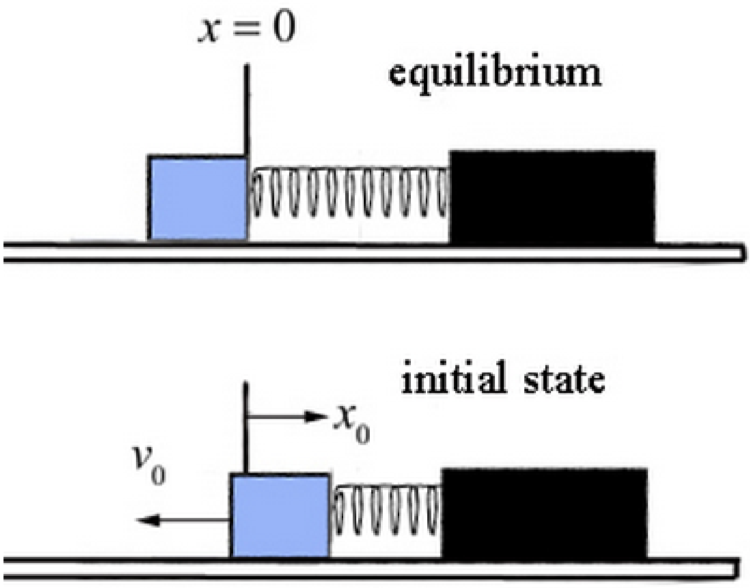
d) After leaving the incline, the block slides along the rough surface until it comes to rest. How far has it traveled?

Express your answers in terms of g, m, L, θ, μ_1 and μ_2.

†206

11. Oscillating block

Consider an ideal spring that has an unstretched length $\ell_0 = 3.1\,\text{m}$. Assume the spring has a constant $k = 36\,\text{N/m}$. Suppose the spring is attached to a mass $m = 7\,\text{kg}$ that lies on a horizontal frictionless surface. The spring-mass system is compressed a distance of $x_0 = 1.8\,\text{m}$ from equilibrium and then released with an initial speed $v_0 = 3\,\text{m/s}$ toward the equilibrium position.



a) What is the period of oscillation T for this system?

b) What is the position of the block as a function of time. Express your answer in terms of t.

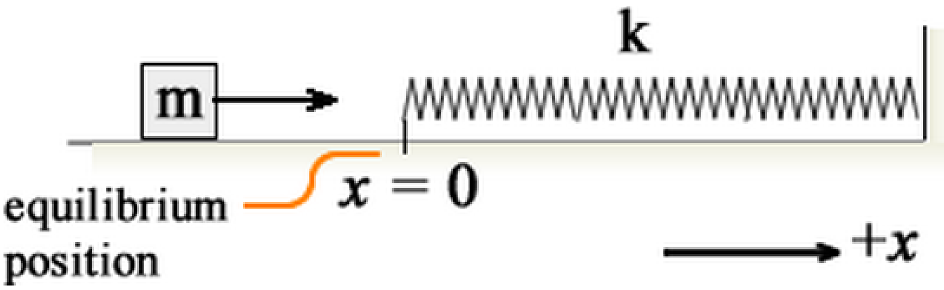
c) How long will it take for the mass to first return to the equilibrium position?

d) How long will it take for the spring to first become completely extended?

†207

12. Spring block with friction

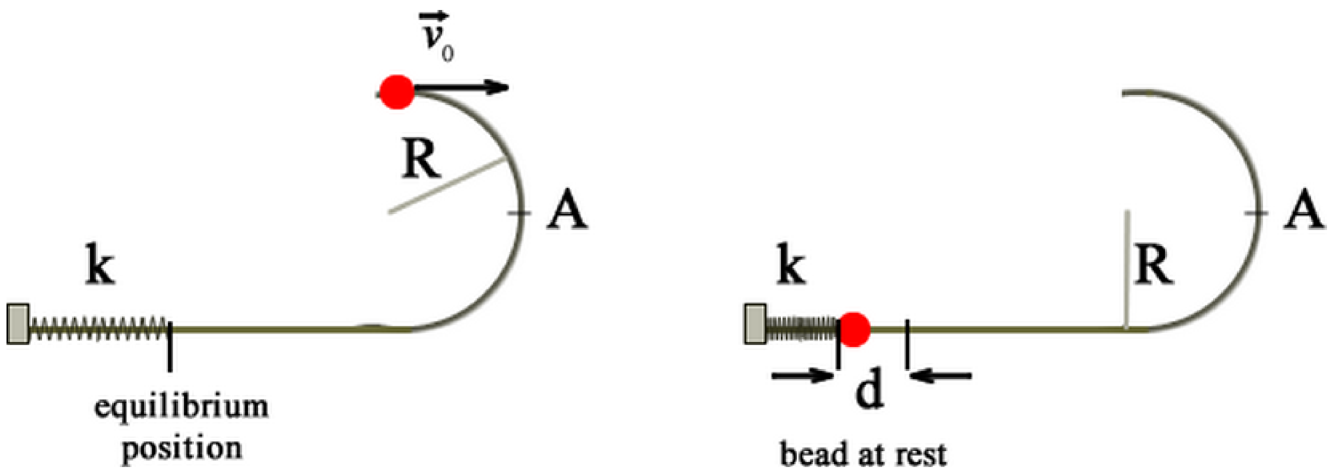
A block of mass $m = 4\,\mathrm{kg}$ slides along a horizontal table when it encounters the free end of a horizontal spring of spring constant $k = 16\,\mathrm{N/m}$. The spring is initially on its equilibrium state, defined when its free end is at $x = 0$ in the figure. Right before the collision, the block is moving with a speed $v_i = 4\,\mathrm{m/s}$. There is friction between the block and the surface. The coefficient of friction is given by $\mu = 0.83$. How far did the spring compress when the block first momentarily comes to rest? Take $g = 10\,\mathrm{m/s^2}$.



†209

13. Half loop

A small bead of mass m is constrained to move along a frictionless track as shown. The track consists of a semicircular portion of radius R followed by a straight part. At the end of the straight portion there is a horizontal spring of spring constant k attached to a fixed support. At the top of the circular portion of the track, the bead is pushed with an unknown speed v_0. The bead comes momentarily to rest after compressing the spring a distance d. The magnitude of the acceleration due to the gravitational force is g.

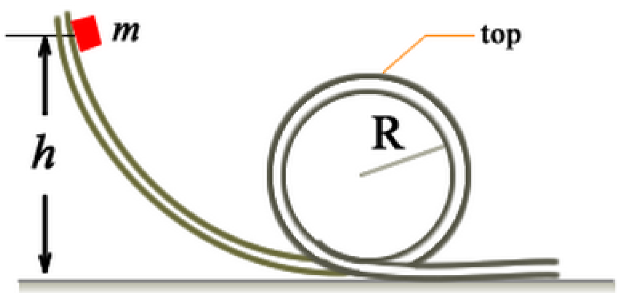


What is the magnitude of the normal force exerted by the track on the bead at the point A, a height R above the base of the track? Express your answer in terms of m, k, R, d, and g but ***not*** in terms of v_0.

†210

14. Full loop

An object of mass m is released from rest at a height h above the surface of a table. The object slides along the inside of the loop-the-loop track consisting of a ramp and a circular loop of radius R shown in the figure. Assume that the track is frictionless.

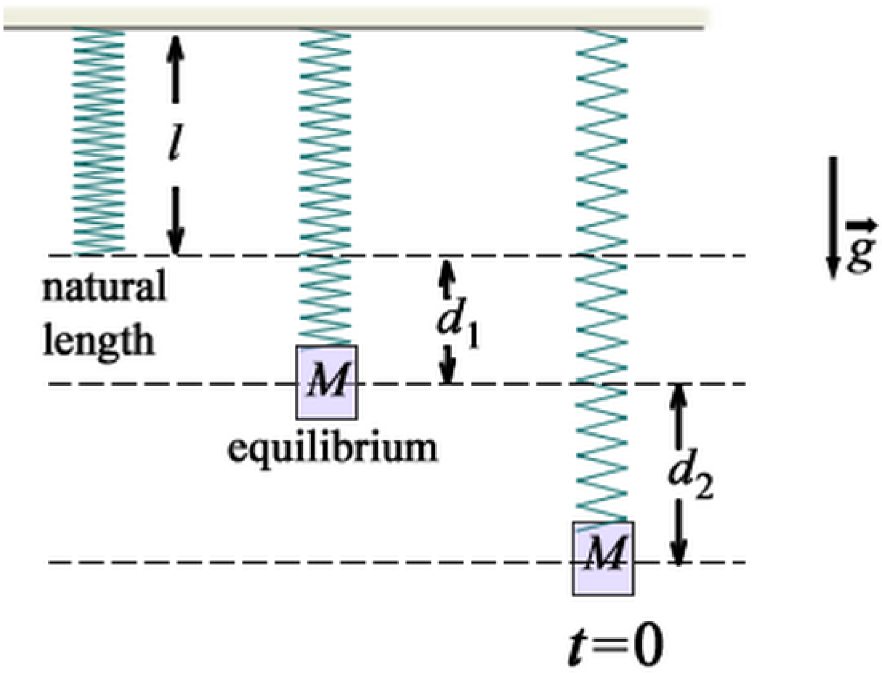


When the object is at the top of the loop it barely loses contact with the track. What height h was the object released from? Express your answer in terms of some or all of the given variables m, g, and R.

†211

15. Vertical spring

A spring of negligible mass, spring constant $k = 99\,\mathrm{N/m}$, and natural length $\ell = 1.3\,\mathrm{m}$ is hanging vertically. This is shown in the left figure below where the spring is neither stretched nor compressed. In the central figure, a block of mass $M = 2\,\mathrm{kg}$ is attached to the free end. When equilibrium is reached (the block is at rest), the length of the spring has increased by d_1 with respect to ℓ. We now lower the block by an additional $d_2 = 0.4\,\mathrm{m}$ as shown in the right figure below. At $t = 0$ we release it (zero speed) and the block starts to oscillate. Take $g = 9.81\,\mathrm{m/s^2}$.



a) Find d_1.

b) What is the frequency (Hz) of the oscillations?

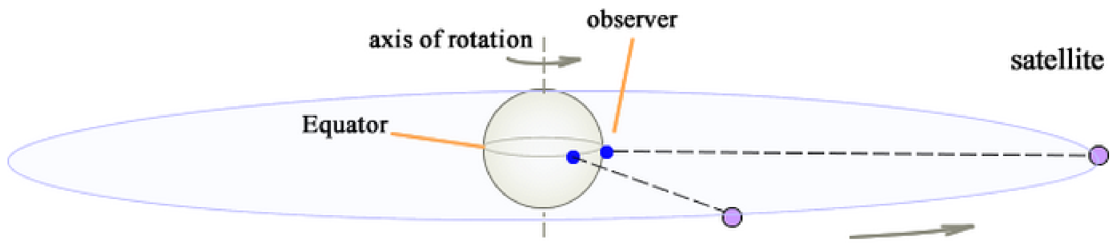
c) What is the length of the spring when the block reaches its highest point during the oscillations?

d) What is maximum speed of the block?

†212

16. Geosynchronous orbit

A satellite with a mass of $m_s = 3 \times 10^3\,\mathrm{kg}$ is in a planet's equatorial plane in a circular "synchronous" orbit. This means that an observer at the equator will see the satellite being stationary overhead (see figure below). The planet has mass $m_p = 5.16 \times 10^{25}\,\mathrm{kg}$ and a day of length $T = 0.7$ earth days (1 earth day = 24 hours).



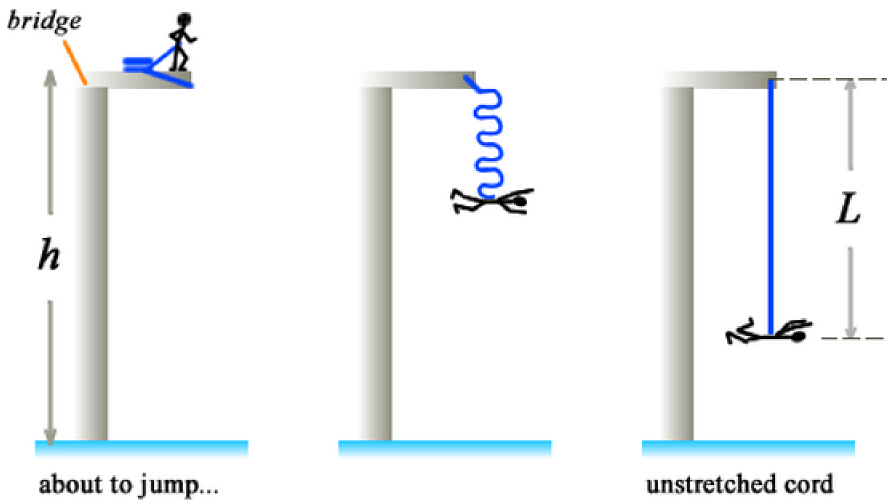
a) How far from the center (in m) of the planet is the satellite?

b) What is the escape velocity (in km/sec) of any object that is at the same distance from the center of the planet that you calculated in (a)?

†213

17. Bungee jumper

A bungee jumper jumps (with no initial speed) from a tall bridge attached to a light elastic cord (bungee cord) of unstretched length L. The cord first straightens and then extends as the jumper falls. This prevents her from hitting the water! Suppose that the bungee cord behaves like a spring with spring constant $k = 90\,\mathrm{N/m}$. The bridge is $h = 100\,\mathrm{m}$ high and the jumper's mass is $m = 65\,\mathrm{kg}$. Use $g = 10\,\mathrm{m/s^2}$.



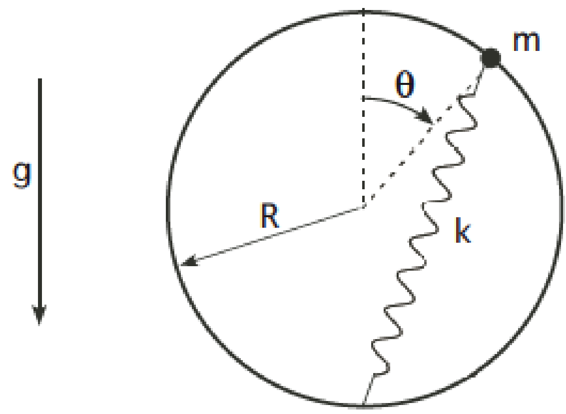
a) What is the maximum allowed length L of the unstretched bungee cord (in m) to keep the jumper alive? (Assume that the spring constant doesn't depend on L).

b) Before jumping, our jumper verified the spring constant of the cord. She lowered herself very slowly from the bridge to the full extent of the cord and when she is at rest she measured the distance to the water surface. What was the measured distance (in m)?

†214

18. Loop, spring and bead

A bead of mass m slides without friction on a vertical hoop of radius R. The bead moves under the combined action of gravity and a spring, with spring constant k, attached to the bottom of the hoop. Assume that the equilibrium (relaxed) length of the spring is R. The bead is released from rest at $\theta = 0$ with a non-zero but negligible speed to the right.



a) What is the speed v of the bead when $\theta = 90°$?

b) What is the magnitude of the force the hoop exerts on the bead when $\theta = 90°$?

Express your answers in terms of m, R, k, and g.

†215

19. Moon

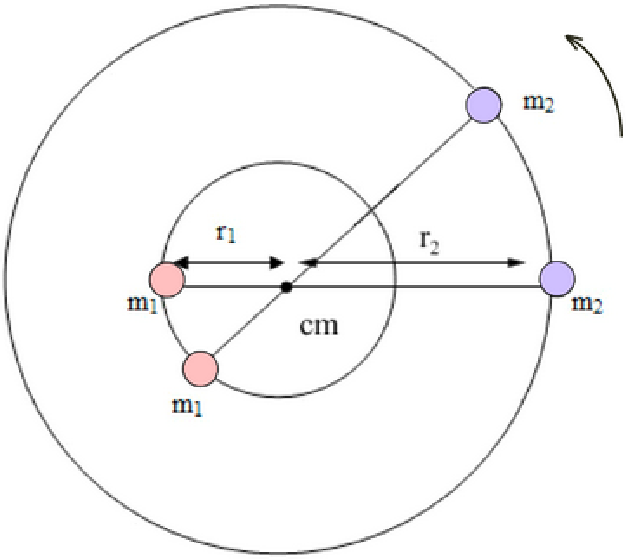
A planet has a single moon that is solely influenced by the gravitational interaction between the two bodies. We will assume that the moon is moving in a circular orbit around the planet and that the moon travels with a constant speed in that orbit. The mass of the planet is $m_p = 3.03 \times 10^{25}$ kg. The mass of the moon is $m_m = 9.65 \times 10^{22}$ kg. The radius of the orbit is $R = 2.75 \times 10^8$ m.

What is the period of the moon's orbit around the planet in earth days (1 earth day = 24 hours).

†217

20. Double star system

Consider a double star system under the influence of the gravitational force between the stars. Star 1 has mass $m_1 = 2.22 \times 10^{31}$ kg and Star 2 has mass $m_2 = 1.64 \times 10^{31}$ kg. Assume that each star undergoes uniform circular motion about the center of mass of the system (cm). In the figure below r_1 is the distance between Star 1 and cm, and r_2 is the distance between Star 2 and cm.



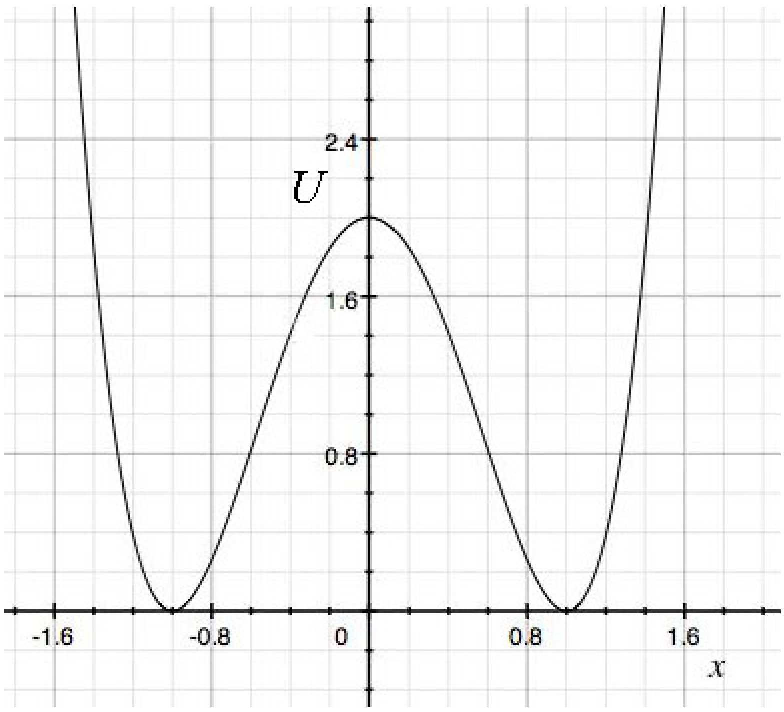
†217

21. Potential energy diagram

A body of mass $m = 1$ kg is moving along the x-axis. Its potential energy is given by the function

$$U(x) = 2(x^2 - 1)^2$$

Note: The units were dropped for the numbers in the equation above. You should note that 2 would carry units of $\text{J} \cdot \text{m}^{-4}$ and 1 would carry units of m^2.



a) What is the x component of the force associated with the potential energy given by $U(x)$? Give an expression in terms of x.

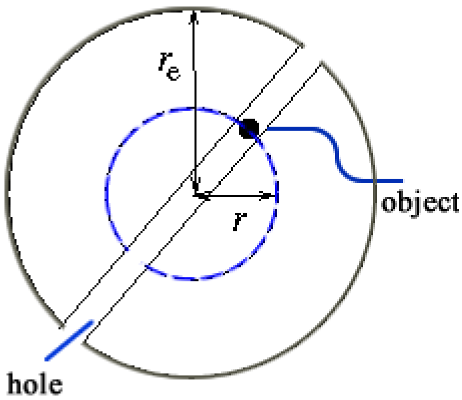
b) At what positive value of x ($x > 0$) in m, does the potential have a stable equilibrium point?

c) Suppose the body starts with zero speed at $x = 1.5$ m. What is its speed (in m/s) at $x = 0$ m and at $x = -1$ m?

†218

22. Earth drilling

A hole is drilled with smooth sides straight through the center of the earth to the other side of the earth. The air is removed from this tube (and the tube doesn't fill up with water, liquid rock or iron from the core). An object is dropped into one end of the tube and just reaches the opposite end. You can assume the earth is of uniform mass density. You can neglect the amount of mass drilled out and the rotation of the earth.



a) The gravitational force on an object of mass m located inside the earth a distance $r < r_e$ from the center (r_e is the radius of the earth) is due only to the mass of the earth that lies within a solid sphere of radius r. What is the magnitude of the gravitational force as a function of the distance r from the center of the earth?

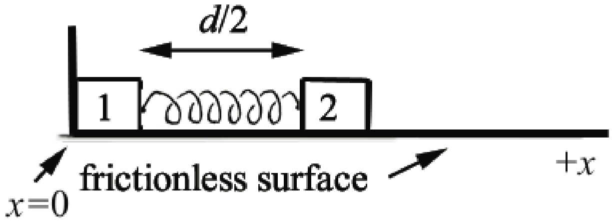
b) How long would it take for this object to reach the other side of the earth?

Express your answers in terms of the gravitational constant at the surface of earth g, m, and r_e as needed.

Note: you do not need the mass of the earth m_e or the universal gravitation constant G to answer this question but you will need to find an expression relating m_e and G to g and r_e.

†220

23. Two blocks and a spring



A system is composed of two non-identical blocks connected by a spring. The blocks slide on a frictionless plane. The unstretched length of the spring is d. Initially block 2 is held so that the spring is compressed to $d/2$ and block 1 is forced against a stop as shown in the figure above. Block 2 is released.

Which of the following statements is true? (Note: more than one statement may be true.)

(a) When the position of block 2 is $x_2 > d$, the center of mass of the system is accelerating to the right.

(b) When the position of block 2 is $x_2 > d$, the center of mass of the system is moving at a constant speed to the right.

(c) When the position of block 2 is $x_2 > d$, the center of mass of the system is at rest.

(d) When the position of block 2 is $x_2 < d$, the center of mass of the system is accelerating to the right.

(e) When the position of block 2 is $x_2 < d$, the center of mass of the system is moving at a constant speed to the right.

(f) When the position of block 2 is $x_2 < d$, the center of mass is at rest.

†221

24. Pushing a baseball bat

The greatest acceleration of the center of mass of a baseball bat will be produced by pushing with a force F at

(a) Position 1 (at the handle)

(b) Position 2 (at the center of mass, around the middle of the bat)

(c) Position 3 (at to the very edge)

(d) Any point. The acceleration is the same.

(e) Not enough information is given to decide.”

†222

25. Jumping off the ground

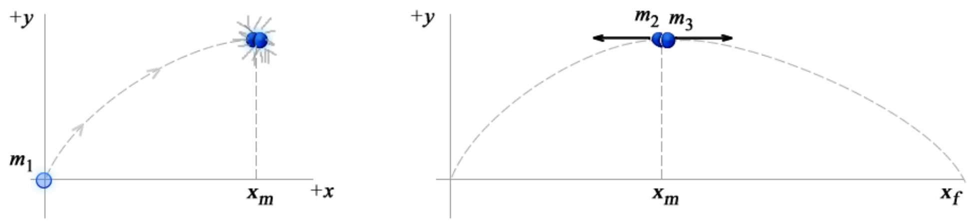
A person of mass m jumps off the ground. Suppose the person pushes off the ground with a constant force of magnitude F for T seconds.

What was the magnitude of the displacement of the center of mass of the person while they were in contact with the ground? Express your answer in terms of m, F, T, and g as needed.

†222

26. Exploding projectile

An instrument-carrying projectile of mass m_1 accidentally explodes at the top of its trajectory. The horizontal distance between launch point and the explosion is x_m. The projectile breaks into two pieces which fly apart horizontally. The larger piece, m_3, has three times the mass of the smaller piece, m_2. To the surprise of the scientist in charge, the smaller piece returns to earth at the launching station. Neglect air resistance and effects due to the earth's curvature.



How far away, x_f, from the original launching point does the larger piece land? Express your answer in terms of some or all of the given variables m_1, x_m, and g.

†222

27. Center of mass of the Earth-Moon system

The mean distance from the center of the earth to the center of the moon is $r_{em} = 3.84 \times 10^8$ m. The mass of the earth is $m_e = 5.98 \times 10^{24}$ kg and the mass of the moon is $m_m = 7.34 \times 10^{22}$ kg. The mean radius of the earth is $r_e = 6.37 \times 10^6$ m. The mean radius of the moon is $r_m = 1.74 \times 10^6$ m.

How far from the center of the earth is the center of mass of the earth-moon system located?"

†223

28. Bouncing ball

A superball of mass m, starting at rest, is dropped from a height h_i above the ground and bounces back up to a height of h_f. The collision with the ground occurs over a total time t_c. You may ignore air resistance.

a) What is the magnitude of the momentum of the ball immediately before the collision? Express your answer in terms of m, h_i, and g as needed.

b) What is the magnitude of the momentum of the ball immediately after the collision? Express your answer in terms of m, h_f, and g as needed.

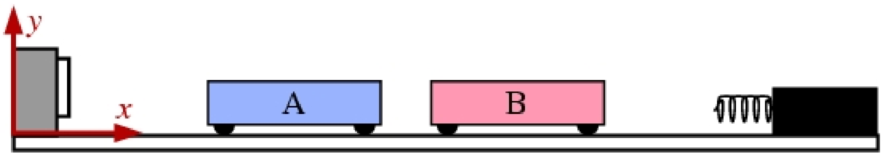
c) What is the magnitude of the impulse imparted to the ball? Express your answer in terms of m, h_i, h_f, t_c, and g as needed.

d) What is the magnitude of the average force of the ground on the ball? Express your answer in terms of m, h_i, h_f, t_c, and g as needed.

†223

29. Colliding carts

The figure below shows the experimental setup to study the collision between two carts.



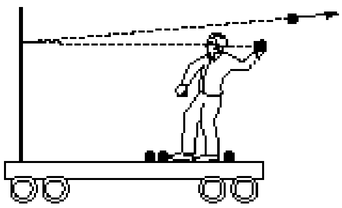
In the experiment cart A rolls to the right on the level track, away from the motion sensor at the left end of the track. Cart B is initially at rest. The mass of cart A is equal to the mass of cart B. Suppose the two carts stick together after the collision. Assume the carts move frictionlessly.

The kinetic energy of the two carts after the collision:

(a) is equal to one half the kinetic energy of cart A before the collision.
(b) is equal to one quarter the kinetic energy of cart A before the collision.
(c) is equal to the kinetic energy of cart A before the collision.
(d) is equal to twice the kinetic energy of cart A before the collision.
(e) is equal to four times the kinetic energy of cart A before the collision.
(f) None of the above.

†224

30. Man on cart throwing balls



Suppose you are on a cart, initially at rest on a track with very little friction. You throw balls at a partition that is rigidly mounted on the cart. If the balls bounce straight back as shown in the figure, is the cart put in motion?

(a) Yes, it moves to the right.

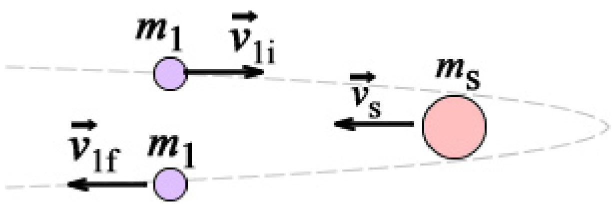
(b) Yes, it moves to the left.

(c) No, it remains in place.

(d) Not enough information is given to decide.

†225

31. Gravitational slingshot

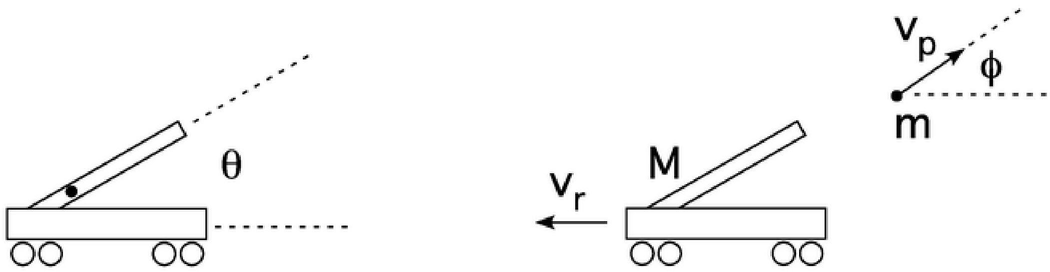


A spacecraft of mass $m_1 = 4757$ kg with a speed $v_{1i} = 3 \times 10^3$ m/s approaches Saturn which is moving in the opposite direction with a speed $v_s = 9.6 \times 10^3$ m/s. After interacting gravitationally with Saturn, the spacecraft swings around Saturn and heads off in the opposite direction it approached. The mass of Saturn is $m_s = 5.69 \times 10^{26}$ kg. Find the final speed v_{1f} (in m/s) of the spacecraft after it is far enough away from Saturn to be nearly free of Saturn's gravitational pull.

†225

32. Railroad gun

A railroad gun of mass $M = 2.0$ kg fires a shell of mass $m = 1.0$ kg at an angle of $\theta = 45°$ with respect to the horizontal as measured relative to the gun. After the firing is complete, the final speed of the projectile relative to the gun (muzzle velocity) is $v_0 = 130.0$ m/s. The gun recoils with speed v_r and the instant the projectile leaves the gun, it makes an angle ϕ with respect to the ground.



a) What is v_p, the speed of the projectile with respect to the ground (in m/s)?

b) What is ϕ, the angle that the projectile makes with the horizontal with respect to the ground (in degrees)?

†226

33. turntables

A turntable is a uniform disc of mass m and radius R. The turntable is initially spinning clockwise when looked down on from above at a constant frequency f_0. The motor is turned off at $t = 0$ and the turntable slows to a stop in time t with constant angular deceleration.

a) What is the magnitude of the initial angular velocity ω_0 of the turntable?

b) What is the magnitude of the angular acceleration α of the turntable?

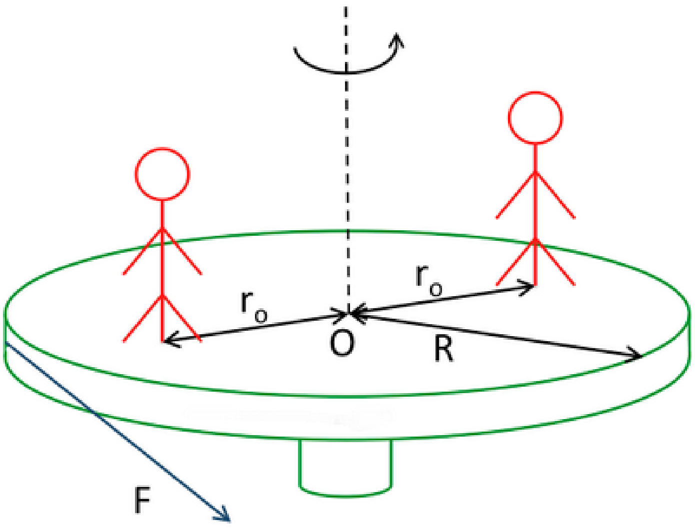
c) What is the magnitude of the total angle $\Delta\theta$ in radians that the turntable spins while slowing down?

Express your answer in terms of f_0 and t."

†229

34. Angular dynamics

A playground merry-go-round has a radius of $R = 2$ m and has a moment of inertia $I_{cm} = 5 \times 10^3\,\mathrm{kgm}^2$ about a vertical axis passing through the center of mass. There is negligible friction about this axis. Two children each of mass $m = 25$ kg are standing on opposite sides at a distance $r_o = 1.4$ m from the central axis. The merry-go-round is initially at rest. A person on the ground applies a constant tangential force of $F = 2 \times 10^2$ N at the rim of the merry-go-round for a time $\Delta t = 10$ s . For your calculations, assume the children to be point masses.



a) What is the angular acceleration α of the merry-go-round (in rad/s^2)?

b) What is the angular velocity ω_{final} of the merry-go-round when the person stopped applying the force (in rad/s)?

c) What average power P_{avg} does the person put out while pushing the merry-go-round (in Watts)?

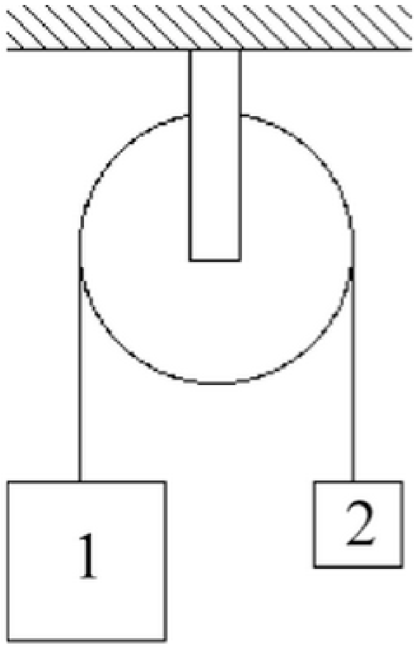
d) What is the rotational kinetic energy R.K.Efinal of the merry-go-round when the person stopped applying the force (in kg m^2/s^2)?

†229

35. Atwood machine

A pulley of mass m_p, radius R, and moment of inertia about the center of mass $I_c = \frac{1}{2} m_p R^2$, is suspended from a ceiling. The pulley rotates about a frictionless axle. An inextensible string of negligible mass is wrapped around the pulley and it does not slip on the pulley. The string is attached on one end to an object of mass m_1 and on the other end to an object of mass with $m_2 < m_1$.

At time $t = 0$, the objects are released from rest.



a) Find the magnitude of the acceleration of the two objects.

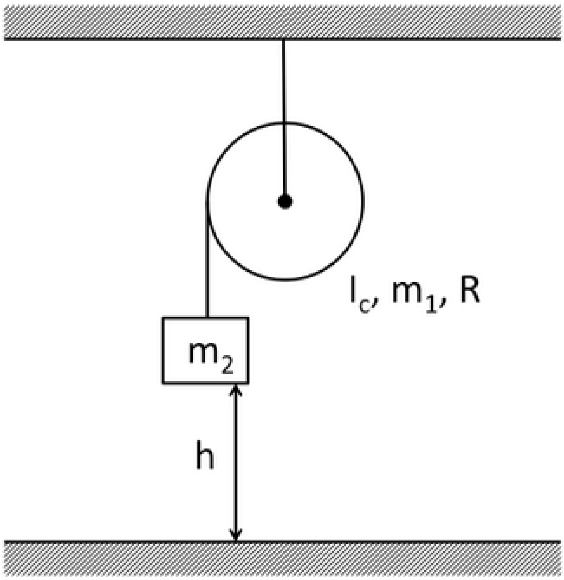
b) How long does it take the objects to move a distance d?

Express your answer in terms of d, m_1, m_2, m_p, R and acceleration due to gravity g.

†231

36. Pulley-object rotational dynamics

A light inflexible cable is wrapped around a cylinder of mass m_1, radius R, and moment of inertia about the center of mass I_c. The cylinder rotates about its axis without friction. The cable does not slip on the cylinder when set in motion. The free end of the cable is attached to an object of mass m_2. The object is released from rest at a height h above the floor. You may assume that the cable has negligible mass. Let g be the acceleration due to gravity.



a) Find the acceleration a of the falling object.

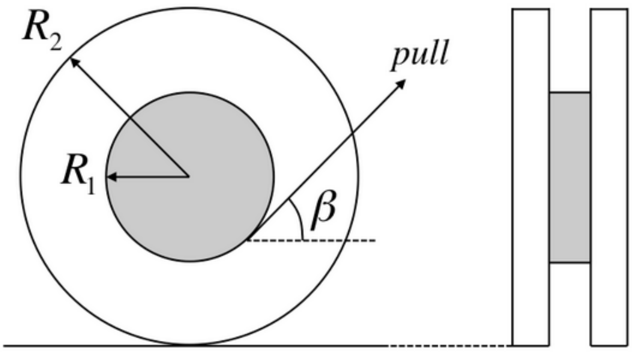
b) Find the tension T in the cable.

c) Find the speed v of the falling object just before it hits the floor.

Express your answer in terms of m_2, R, I_{cm}, h and g as needed.

†232

37. Yo-yo



A yo-yo of mass m rests on the floor (the static friction coefficient with the floor is μ). The inner (shaded) portion of the yo-yo has a radius R_1, the two outer disks have radii R_2. A string is wrapped around the inner part. Someone pulls on the string at an angle β (see sketch). The "pull" is very gentle, and is carefully increased until the yo-yo starts to roll without slipping. Try it at Home; it's Fun!

For what angles of β will the yo-yo roll to the left and for what angles to the right?

(i) Yo-Yo rolls to the left if $\sin\beta < \dfrac{R_1}{R_2}$, and to the right if $\sin\beta > \dfrac{R_1}{R_2}$.

(ii) Yo-Yo rolls to the left if $\sin\beta > \dfrac{R_1}{R_2}$, and to the right if $\sin\beta < \dfrac{R_1}{R_2}$.

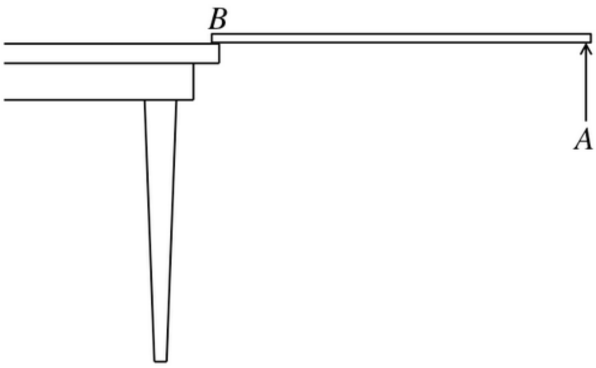
(iii) Yo-Yo rolls to the left if $\cos\beta < \dfrac{R_1}{R_2}$, and to the right if $\cos\beta > \dfrac{R_1}{R_2}$.

(iv) Yo-Yo rolls to the left if $\cos\beta > \dfrac{R_1}{R_2}$, and to the right if $\cos\beta < \dfrac{R_1}{R_2}$.

†234

38. Stick on table

A uniform stick of mass m and length ℓ is suspended horizontally with end B at the edge of a table as shown in the diagram, and the other end A is originally held by hand. The hand at A is suddenly released.



At the instant immediately after the release:

a) What is the magnitude of the torque (τ_B) about the end B at the edge of the table?

b) What is the magnitude of the angular acceleration α about the end B at the edge of the table?

c) What is the magnitude of the vertical acceleration a of the center of mass?

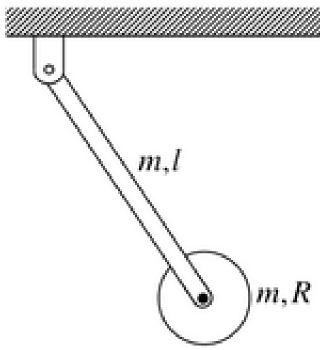
d) What is the magnitude of the vertical component of the hinge force (N) at B?

Express your answer in terms of m, ℓ and acceleration due to gravity g as needed.

†235

39. Physical pendulum

A physical pendulum consists of a disc of radius R and mass m fixed at the end of a rod of mass m and length ℓ.



a) Find the period of the pendulum for small angles of oscillation.

b) For small angles of oscillation, what is the new period of oscillation if the disk is mounted to the rod by a frictionless bearing so that it is perfectly free to spin?

Express your answer in terms of m, R, ℓ and acceleration due to gravity g as needed.

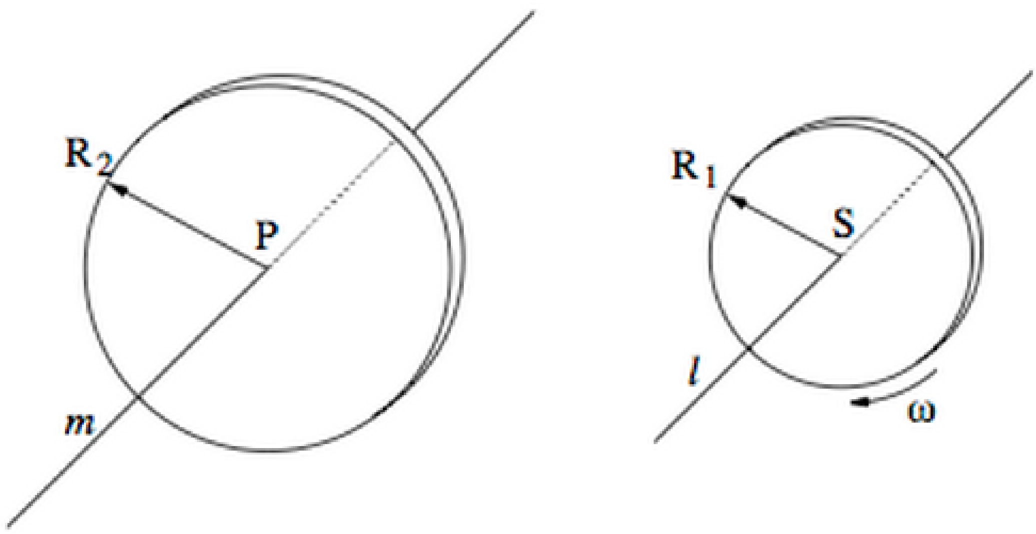
†237

40. Two rotating disks

A solid disk 1 with radius R_1 is spinning freely about a frictionless horizontal axle ℓ at an angular speed ω initially. The axle ℓ is perpendicular to disk 1, and goes through the center S of disk 1.

The circumference of disk 1 is pushed against the circumference of another disk (disk 2). Disk 2 has the same thickness and density as disk 1, but has a radius R_2, and it is initially at rest. Disk 2 can rotate freely about a horizontal axle m through its center P. Axles m and ℓ are parallel. The friction coefficient between the two touching surfaces (disk circumferences) is μ.

We wait until an equilibrium situation is reached (i.e. the circumferences of the two disks are no longer slipping against each other). At that time, disk 1 is spinning with angular velocity ω_1, and disk 2 is spinning with angular velocity ω_2.



Calculate the magnitude of the angular velocities $|\omega_1|$ and $|\omega_2|$ in terms of R_1, R_2 and ω.

†239

41. Translation and rotation

A rod is lying at rest on a perfectly smooth horizontal surface (no friction). We give rod a short impulse (a hit) perpendicular to the length direction of the rod at X. The mass of the rod is 3 kg, and its length is 50 cm. The impulse is 4 kgm/s. The distance from the center C of the rod to X is 15 cm.

a) What is the translational speed $|\mathbf{v_{cm}}|$ of C after the rod is hit (in m/s)?

b) What is the magnitude of the angular velocity ω of the rod about C (in rad/s)?

c) How far (distance D in meters) has the center C of the rod moved from its initial position 8 seconds after it was hit? And what is the angle θ (in radians) between the direction of the rod at 8 seconds after it was hit, and its initial direction (before it was hit)? Give the smaller angle.

d) What is the total kinetic energy K of the rod after it was hit? (in joules)

†243

42. Going to the Sun

A spacecraft of mass m is first brought into an orbit around the earth. The earth (together with the spacecraft) orbits the sun in a near circular orbit with radius R (R is the mean distance between the earth and the sun; it is about 150 million km).

a) What is the speed v_0 (in m/s) of the earth in its orbit of radius $R = 1.5 \times 10^{11}$ m around the sun with a mass $M = 1.99 \times 10^{30}$ kg? Take the gravitational constant $G = 6.674 \times 10^{-11}\,\mathrm{m^3kg^{-1}s^{-2}}$.

b) What is the total impulse I_0 that would have to be given by the rocket to the spacecraft to accomplish this? You may ignore the effect of the earth's gravitation as well as the orbital speed of the spacecraft around the earth as the latter is much smaller than the speed of the earth around the sun. Thus, you may assume that the spacecraft, before the rocket is fired, has the same speed in its orbit around the sun as the earth. Express your answer in terms of m and v_0.

c) Calculate the impulse I_1 required at the first rocket burn (the boost). Express your answer in terms of I_0, R and r."

d) What is the speed v_2 of the spacecraft at aphelion? Express your answer in terms of v_0, R and r.

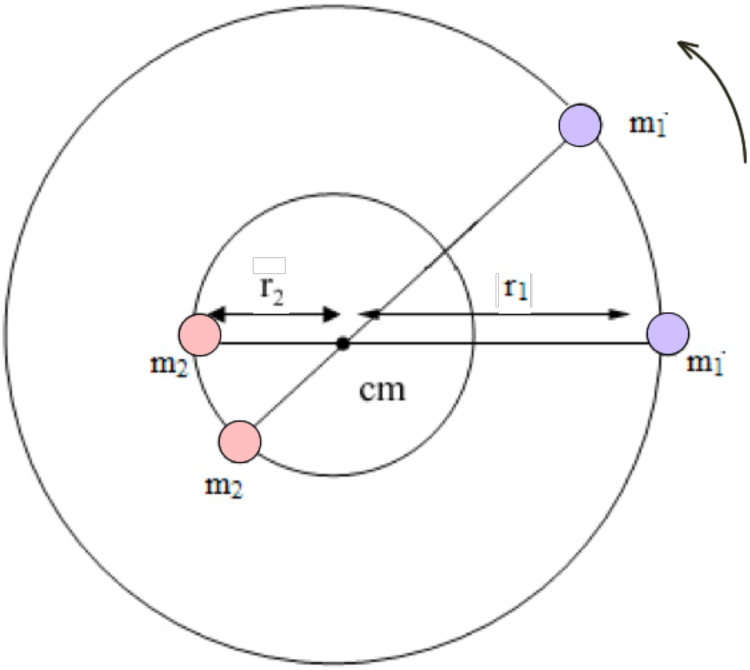
e) Calculate the impulse I_2 required at the second rocket burn (at aphelion). Express your answer in terms of I_0, R and r.

f) Compare the impulse under b) with the sum of the impulses under c) and e) (i.e find $I_0 - (I_1 + I_2)$), and convince yourself that the latter procedure is more economical. Express your answer in terms of I_0, R and r.

†244

43. Black hole in X-ray binary

An X-ray binary consists of 2 stars with masses m_1 (the accreting compact object) and m_2 (the donor). The orbits are circular with radii r_1 and r_2 centered on the center of mass.



a) Find the orbital period T of the binary following the guidelines given in lectures. Express your answer in terms of $(m_1 + m_2)$, $(r_1 + r_2)$ and G.

b) In the case of Cyg X-1 (as discussed in lectures), the orbital period is 5.6 days. The donor star is a "supergiant" with a mass 30 times that of the sun. Doppler shift measurements indicate that the donor star has an orbital speed v_2 of about 148 km/sec. Calculate r_2 (in meters).

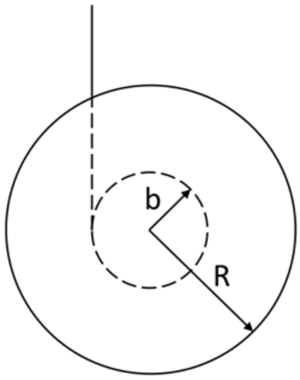
c) Calculate r_1 (in meters).

d) Now calculate the mass m1 of the accreting compact object (express that as ratio to the mass of the sun m_1/M_{Sun})."

†248

44. Torque, rotation and translation

A Yo-Yo of mass m has an axle of radius b and a spool of radius R. Its moment of inertia about the center can be taken to be $I = \frac{1}{2}mR^2$ and the thickness of the string can be neglected. The Yo-Yo is released from rest. You will need to assume that the center of mass of the Yo-Yo descends vertically, and that the string is vertical as it unwinds.



a) What is magnitude of the tension in the cord as the Yo-Yo descends?

b) Find the angular speed of the Yo-Yo when it reaches the bottom of the string, when a length ℓ of the string has unwound.

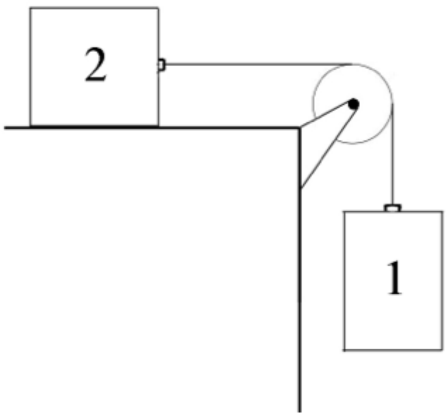
c) Find the magnitude of the average tension in the string over the course of the Yo-Yo reversing its direction at the bottom of its descent (see figure below).

Express your answer in terms of m, b, R, ℓ and acceleration due to gravity g.

†249

45. Double block pulley

A pulley of mass m_p, radius R, and moment of inertia about its center of mass I_c, is attached to the edge of a table. An inextensible string of negligible mass is wrapped around the pulley and attached on one end to block 1 that hangs over the edge of the table. The other end of the string is attached to block 2 which slides along a table. The coefficient of sliding friction between the table and the block 2 is μ_k . Block 1 has mass m_1 and block 2 has mass m_2, with $m_1 > \mu_k m_2$. At time $t = 0$, the blocks are released from rest. At time $t = t_1$, block 1 hits the ground. Let g denote the gravitational acceleration near the surface of the earth.



a) Find the magnitude of the linear acceleration of the blocks.

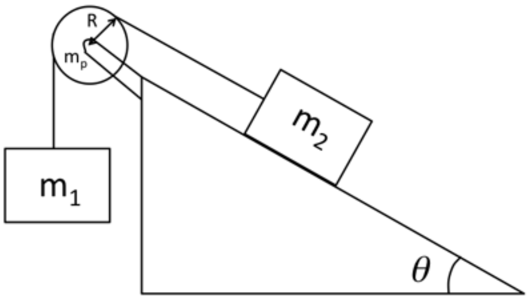
b) How far did the block 1 fall before hitting the ground?

Express your answer in terms of m_1, m_2, I_c, R, μ_k, t_1 and g as needed.

†252

46. Wheel, inclined plane, two masses and a rope

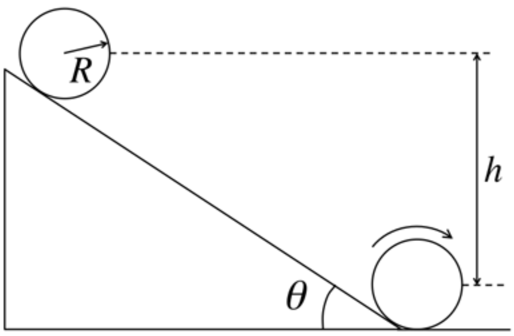
A wheel in the shape of a uniform disk of radius R and mass m_p is mounted on a frictionless horizontal axis. The wheel has moment of inertia about the center of mass $I_{cm} = \frac{1}{2} m_p R^2$. A massless cord is wrapped around the wheel and one end of the cord is attached to an object of mass m_2 that can slide up or down a frictionless inclined plane. The other end of the cord is attached to a second object of mass m_1 that hangs over the edge of the inclined plane. The plane is inclined from the horizontal by an angle θ. Once the objects are released from rest, the cord moves without slipping around the disk. Find the magnitude of accelerations of each object, and the magnitude of tensions in the string on either side of the pulley. Assume that the cord doesn't stretch ($a_1 = a_2 = a$). Express your answers in terms of the masses m_1, m_2, m_p, angle θ and the gravitational acceleration due to gravity near earth's surface g.'



†253

47. Rolling object on an incline

A hollow cylinder of outer radius R and mass M with moment of inertia about the center of mass $I_{cm} = MR^2$ starts from rest and moves down an incline tilted at an angle θ from the horizontal. The center of mass of the cylinder has dropped a vertical distance h when it reaches the bottom of the incline. Let g denote the acceleration due to gravity. The coefficient of static friction between the cylinder and the surface is μ_s. The cylinder rolls without slipping down the incline. The goal of this problem is to find an expression for the smallest possible value of μ_s such that the cylinder rolls without slipping down the incline plane and the velocity of the center of mass of the cylinder when it reaches the bottom of the incline.



a) What is the magnitude of the acceleration a of the center of mass of the cylinder on the incline?

b) What is the minimum value for the coefficient of static friction μ_s such that the cylinder rolls without slipping down the incline plane?

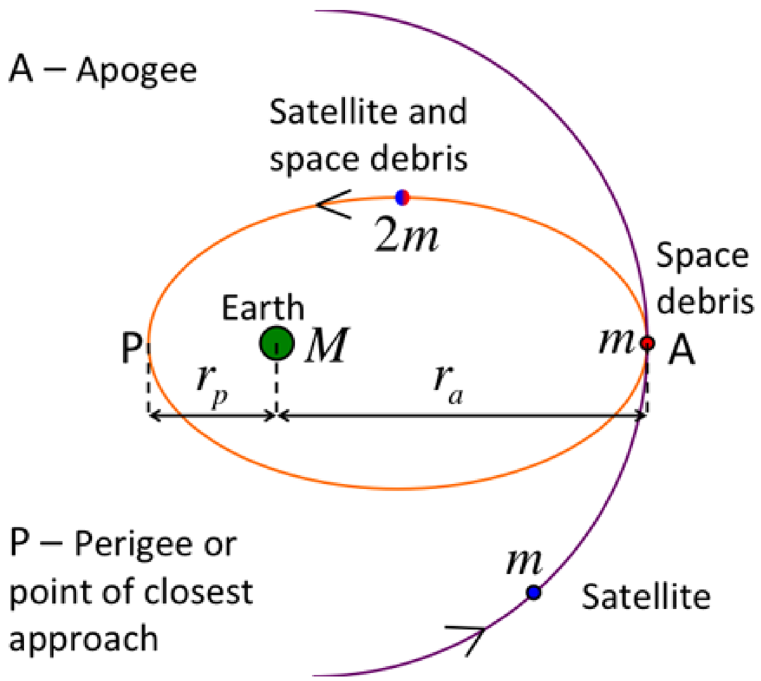
c) What is the magnitude of the velocity of the center of mass of the cylinder when it reaches the bottom of the incline?

Express your answer in terms of θ, h and g as needed.

†255

48. Space debris collision

A satellite of mass m is orbiting the earth, mass M, in a circular orbit of radius r_a. Unfortunately a piece of space debris left by a passing rocket lies directly in the satellite's path. The piece of debris has the same mass m as the satellite. The debris collides with the satellite and sticks to the satellite. Assume that the debris has negligible speed just before the collision. After the collision, the satellite and debris enter an elliptical orbit around the earth. The distance of closest approach to the earth of the satellite and the debris is r_p. Let G be the universal constant of gravity. You may assume that $M \gg m$.



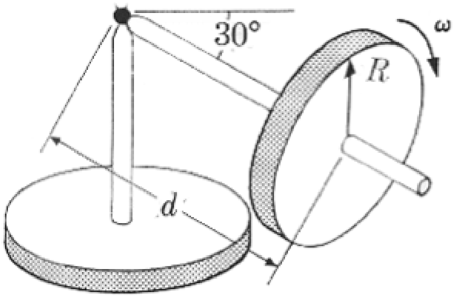
a) Find an expression for the speed v_0 of the satellite before the collision.

b) Calculate the ratio r_a/r_p.

You may express your answer in terms of M, r_a and G as needed.

†257

49. Turntable solutions

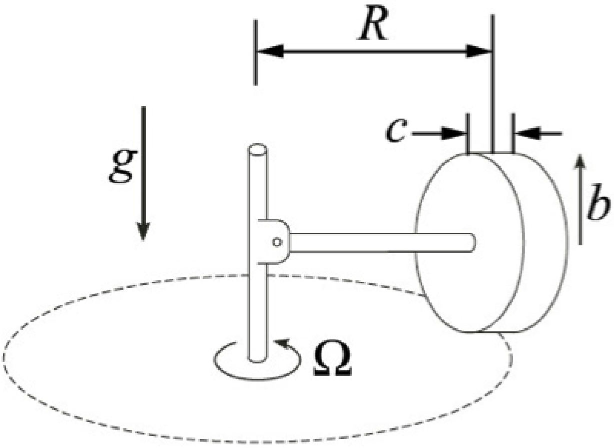


A gyroscope consists of a uniform disc of mass radius $M = 2$ kg and radius $R = 0.3$ m. The disc spins with an angular speed $\omega = 400$ rad/s as shown in the figure [above]. The gyroscope precesses, with its axle at an angle 30° below the horizontal (see figure). The gyroscope is pivoted about a point $d = 0.6$ m from the center of the disc. What is the magnitude of the precessional angular velocity Ω (in radians/sec)?

†258

50. Grain mill

In a grain mill, grain is ground by a massive wheel which rolls without slipping in a circle on a flat horizontal surface driven by a vertical shaft. The rolling wheel has radius b and is constrained to roll in a horizontal circle of radius R at angular speed Ω. Because of the stone's angular momentum, the contact force with the surface can be considerably greater than the weight of the wheel. In this problem, the angular speed Ω about the shaft is such that the contact force between the ground and the wheel is equal to twice the weight. The goal of the problem is to find Ω. Assume that the wheel is closely fitted to the axle so that it cannot tip, and that the width of the wheel $c\mathcal{L}R$. Neglect friction and the mass of the axle of the wheel. Let g denote the acceleration due to gravity.



a) How is the angular speed ω of the wheel about its axis related to the angular speed Ω about the shaft?

b) What is the horizontal component of the angular momentum vector about the point P in the figure above? Although we have not shown this, for this situation it is correct to compute the horizontal component of the angular momentum by completely ignoring the rotation of the mill wheel about the vertical axis, taking into account only the rotation of the mill wheel about its own axle.

c) What is the magnitude of the torque about the joint (about the point P in the figure above) due to the forces acting on the axle-wheel combination?

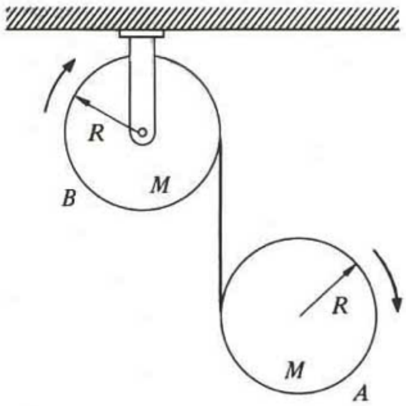
d) What is the value of Ω if the contact force between the stone and the ground $N = 2Mg$?

Express your answers in terms of Ω, M, N, b, R and g as necessary.

†259

51. Double drums rotating

A drum A of mass M and radius R is suspended from a drum B also of mass M and radius R, which is free to rotate about its axis. The suspension is in the form of a massless metal tape wound around the outside of each drum, and free to unwind. Gravity is directed downwards. Both drums are initially at rest. Consider the drums to be uniform disks.

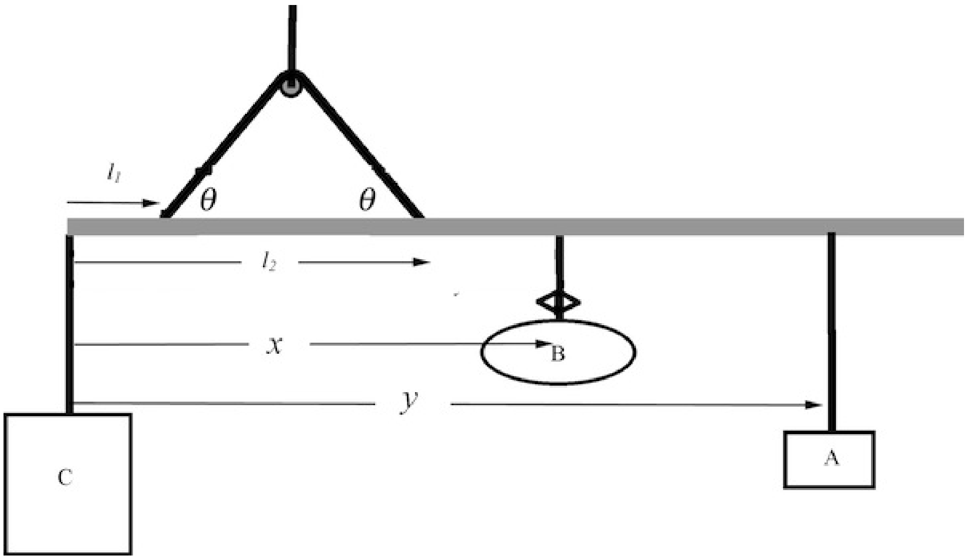


Find the initial acceleration of drum A, assuming that it moves straight down. Express your answer in terms of M, R and acceleration due to gravity g as needed.

†261

52. Crane

A crane is configured as below, with the beam suspended at two points ℓ_1 and ℓ_2 by each end of a cable passing over a frictionless pulley. The two ends of the cable each make an angle θ with the beam. A counterbalance object C with mass m_C is fixed at one end of the beam. A balance object B of mass m_B is attached to the beam and can move horizontally in order to maintain static equilibrium. The crane lifts an object A with mass m_A at a distance y from the counterbalance. For simplicity, assume the pulley, beam and cable to be massless.



a) What is the tension in the cable that runs over the pulley?

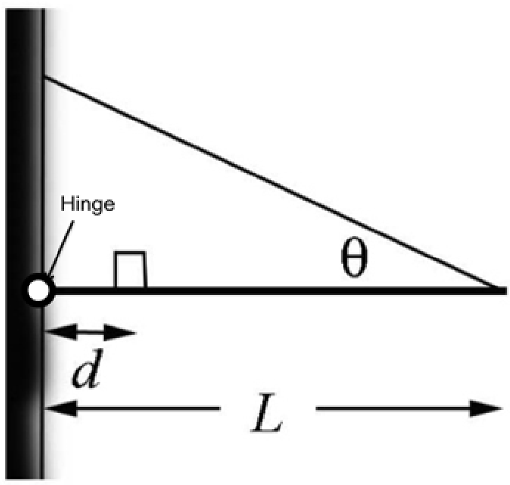
b) At what horizontal position, x, should one put the balance object B such that the crane doesn't tilt?

Express your answer in terms of g, m_A, m_B, m_C, ℓ_1, ℓ_2 θ and y as needed.

†263

53. Steel beam and cable

A uniform steel beam of mass $m_1 = 150.0\,\text{kg}$ is held up by a steel cable that is connected to the beam a distance $L = 5.0$ m from the wall, at an angle $\theta = 35.0°$ as shown in the sketch. The beam is bolted to the wall with an unknown force $\mathbf{F}$ exerted by the wall on the beam. An object of mass $m_2 = 60.0\,\text{kg}$ resting on top of the beam, is placed a distance $d = 2.0$ m from the wall. For simplicity, assume the steel cable to be massless. Use $g = 9.8\,\text{m/s}^2$ for the gravitational acceleration.



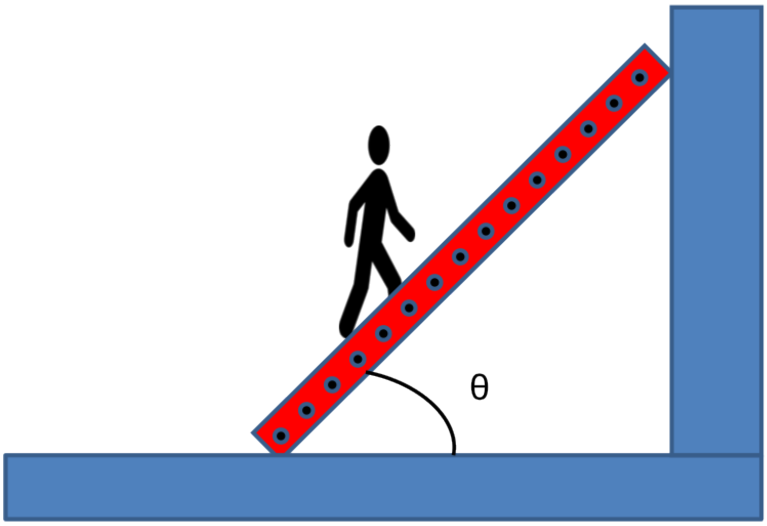
a) Find the tension (in Newton) in the cable. Start by drawing a free-body diagram for the beam, then find equations for static equilibrium for the beam (this will involve force equations and torque relations).

b) Find the horizontal and vertical components of the force (in Newton) that the wall exerts on the beam."

†264

54. Person on ladder

A person of mass $m_2 = 85.0$ kg is standing on a rung, one third of the way up a ladder of length $d = 4.0$ m. The mass of the ladder is $m_1 = 15.0$ kg, uniformly distributed. The ladder is initially inclined at an angle $\theta = 40.0°$ with respect to the horizontal. Assume that there is no friction between the ladder and the wall but that there is friction between the base of the ladder and the floor with a coefficient of static friction μ_s.



Start this problem by drawing a free-body force diagrams showing all the forces acting on the person and the ladder. Indicating a choice of unit vectors on your free-body diagrams may be helpful.

a) Using the equations of static equilibrium for both forces and torque, find expressions for the normal and horizontal components of the contact force between the ladder and the floor, and the normal force between the ladder and the wall. Consider carefully which point to use for computing the torques. Determine the magnitude of the frictional force (in N) between the base of the ladder and the floor below.

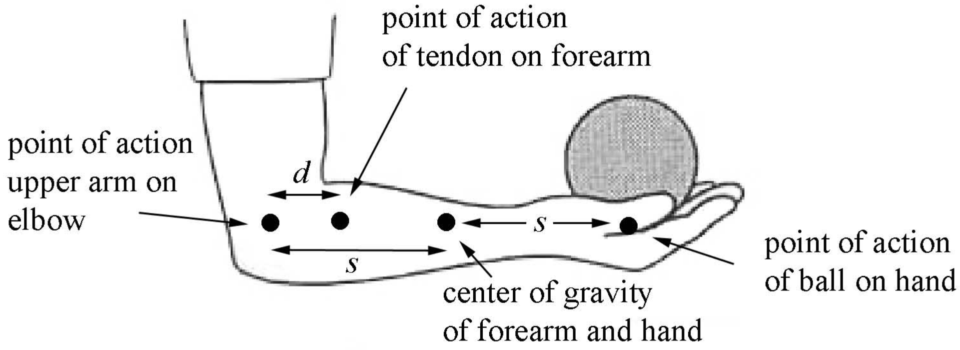
b) Find the magnitude for the minimum coefficient of friction between the ladder and the floor so that the person and ladder does not slip.

c) Find the magnitude $C_{ladder,ground}$ (in N) of the contact force that the floor exerts on the ladder. Remember, the contact force is the vector sum of the normal force and friction. Find the direction of the contact force that the floor exerts on the ladder. i.e. determine the angle α (in radians) that the contact force makes with the horizontal to indicate the direction.

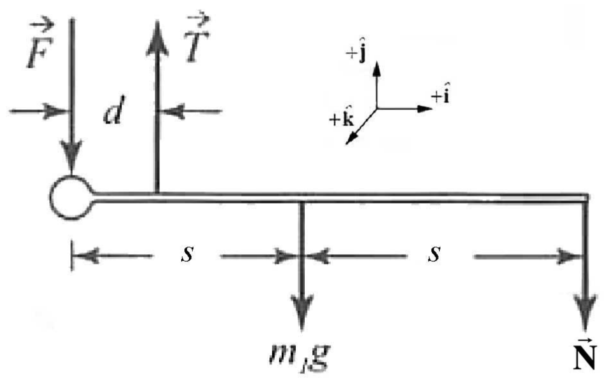
†265

55. Static equilibrium arm

You are holding a ball of mass m_2 in your hand. In this problem you will solve for the upward force $\mathbf{T}$ that the tendon of your biceps muscle exerts to keep the forearm horizontal and the downward force $\mathbf{F}$ that the upper arm exerts on the forearm at the elbow joint. Assume the outstretched arm has a mass of m_1, the center of mass of the outstretched arm is a distance s from the elbow, the tendon attaches to the bone a distance d from the elbow, and the ball is a distance $2s$ from the elbow. (Taking $\mathbf{T}$ to be upward and $\mathbf{F}$ to be downward, with no horizontal components, indicates that this is a simplified model.)
A schematic representation of this situation is shown below:



Hint: The forces can be modeled as shown in the following Free Body Diagram:



a) What is the magnitude of the tension $T \equiv |\mathbf{T}|$ in the tendon?

b) What is the magnitude of the force that the upper arm exerts on the forearm at the elbow joint?

Express your answer in terms of s, m_1, m_2, d and g as needed.

†266

56. Specific strength

A metal meter stick made of steel rotates about its midpoint. The angular speed is slowly increased. At what value of the angular speed will the stick break apart at the center? Give your answer in rad/s.

Hint: find a relationship between the maximum angular frequency and the breaking (ultimate tensile strength) of steel. Use the values that are given in this table in the handout of lecture 26 [link not copied].'

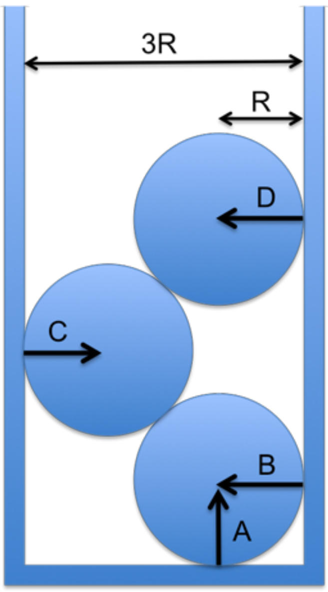
†267

57. Static friction of stick leaning against a wall

A stick of length $\ell = 60.0$ cm rests against a wall. The coefficient of static friction between stick and the wall and between the stick and the floor are equal. The stick will slip off the wall if placed at an angle greater than $\theta = 40.0$ degrees. What is the coefficient of static friction, μ_s, between the stick and the wall and floor?

†268

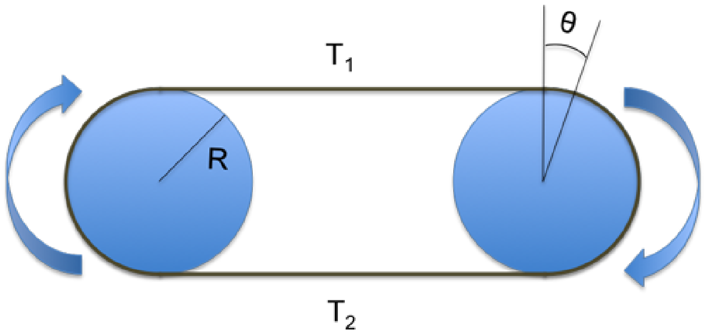
58. Three balls in a tube



Three smooth balls of iron of mass m and radius R are placed inside a tube of diameter $3R$ (see Figure). Find the magnitude of the forces (A, B, C and D) exerted by the sides of the container on each ball. Write your answers in terms of m, g and R.

†270

59. Two flywheels and a drive belt



The flywheel of a motor is connected to the flywheel of an electric generator by a drive belt. The flywheels are of equal size each of radius R. While the flywheels are rotating the tension in the upper and lower portions of the drive belt are T_1 and T_2 respectively. The drive belt exerts a torque $\tau = (T_2 - T_1)R$ on the generator (around its center). The coefficient of static friction between the drive belt and each flywheel is μ_s. Assume the tension is as high as possible with no slipping between the belt and the flywheel, and that the drive belt is massless.

a) Derive a differential expression representing the change of tension along the portion of the belt in contact with one of the flywheels. That is find the value of dT/T for one of the two flywheels. $dT/T =$

(i) $\frac{1}{\mu_s} d\theta$

(ii) $\frac{1}{\mu_s R} d\theta$

(iii) $\mu_s d\theta$

(iv) $R\mu_s d\theta$

What is T_1?

(i) $\frac{\tau}{R}\frac{1}{e^{\mu_s \pi} - 1}$

(ii) $\frac{\tau}{R}\frac{1}{1 - e^{-\mu_s \pi}}$

(iii) $\frac{\tau}{R} e^{\mu_s \pi}$

(iv) $\frac{\tau}{R} e^{-\mu_s \pi}$

(v) $\frac{\tau}{R}(1 - e^{\mu_s \pi})$

What is T_2?

(i) $\frac{\tau}{R}\frac{1}{e^{\mu_s \pi} - 1}$

(ii) $\dfrac{\tau}{R}\dfrac{1}{1-e^{-\mu_s\pi}}$

(iii) $\dfrac{\tau}{R}e^{\mu_s\pi}$

(iv) $\dfrac{\tau}{R}e^{-\mu_s\pi}$

(v) $\dfrac{\tau}{R}(1-e^{\mu_s\pi})$

†272

60. Hanging rod length

A long rod hangs straight down from one end. How long (in meters) can the rod be before its weight causes it to break off at the end if it is made of iron? Titanium? Give your answer in meters.

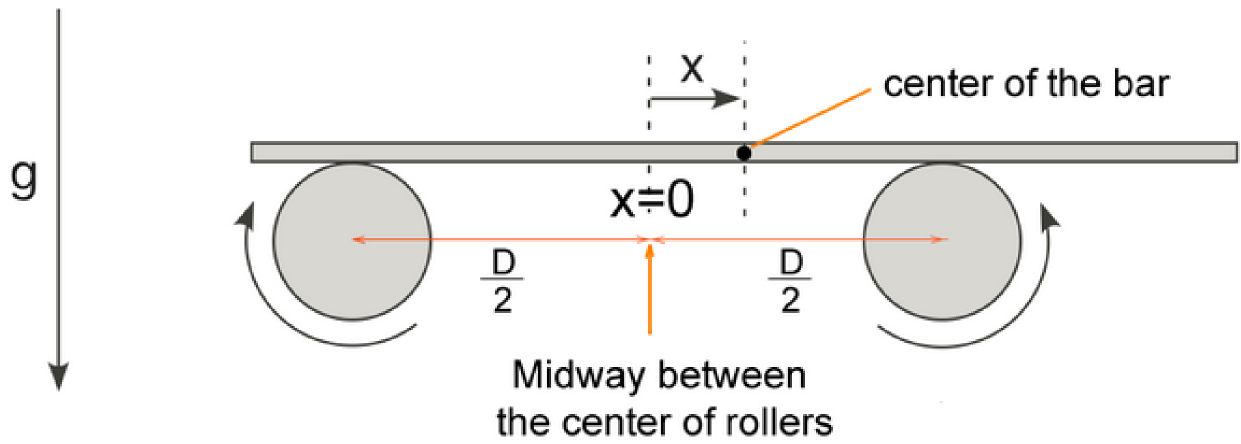
Use the following values for densities and tensile strengths:

The densities of iron and titanium are $7.8 \times 10^3\,\mathrm{kg/m^3}$ and $4.5 \times 10^3\,\mathrm{kg/m^3}$ respectively.

The breaking - ultimate tensile strength: 350 MPa for iron and 450 MPa for titanium (MPa $= 10^6$ N/m^2).

†274

61. Bar on rollers



A bar of mass m and negligible height is lying horizontally across and perpendicular to a pair of counter rotating rollers as shown in the figure. The rollers are separated by a distance D. There is a coefficient of kinetic friction μ_k between each roller and the bar. Assume that the bar remains horizontal and never comes off the rollers, and that its speed is always less than the surface speed of the rollers. Take the acceleration due to gravity to be g.

a) Find the normal forces N_L and N_R exerted by the left and right rollers on the bar when the center of the bar is displaced a distance x from the position midway between the rollers.

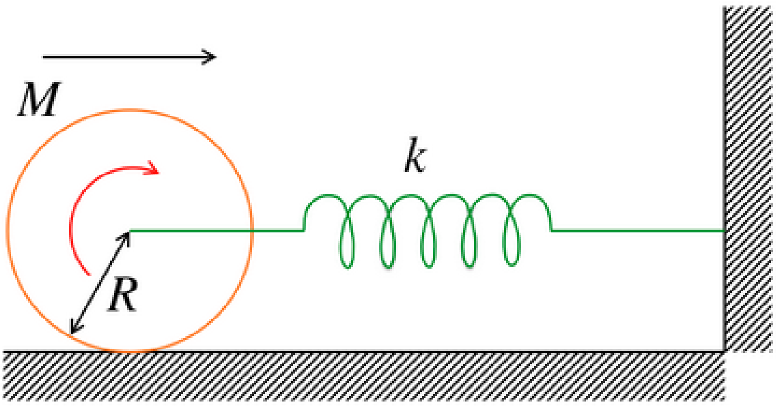
b) Find the differential equation governing the horizontal displacement of the bar $x(t)$.

c) The bar is released from rest at $x = x_0$ at $t = 0$. Find the subsequent location of the center of the bar, $x(t)$.

Express your answer in terms of x, x_0, d, m, μ_k, t and g as needed.

†274

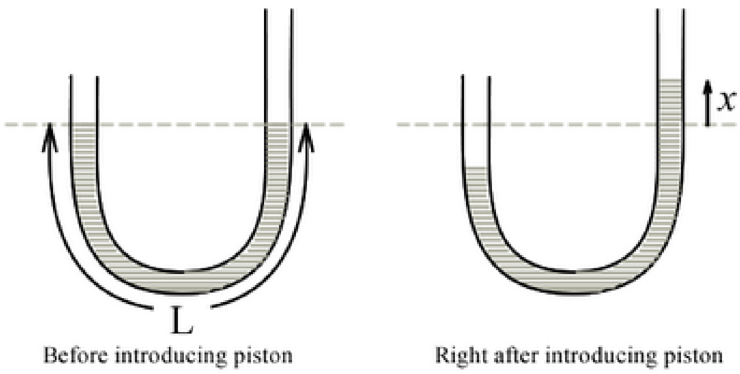
62. Table problem: Rolling solution



Attach a solid cylinder of mass M and radius R to a horizontal massless spring with spring constant k so that it can roll without slipping along a horizontal surface. If the system is released from rest at a position in which the spring is stretched by an amount x_0 what is the period T of simple harmonic motion for the center of mass of the cylinder? Express your answer in terms of M and k.

†276

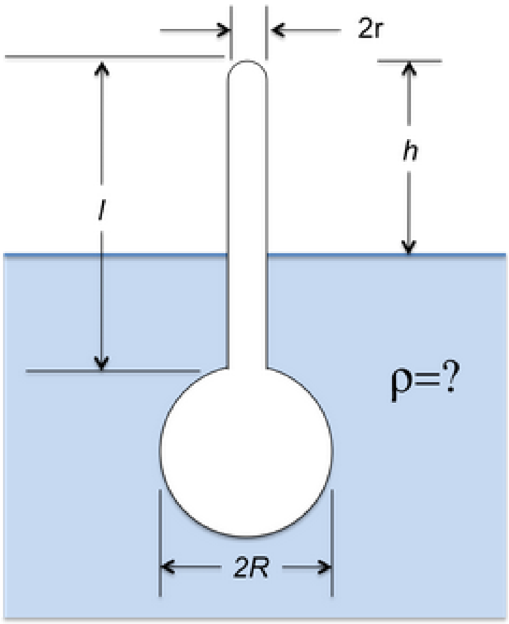
63. U-tube



A U-tube open at both ends to atmospheric pressure P_0 is filled with an incompressible fluid of density ρ. The cross-sectional area A of the tube is uniform and the total length of the column of fluid is L. A piston is used to depress the height of the liquid column on one side by a distance x_0, and then is quickly removed. What is the frequency of the ensuing simple harmonic motion? Assume streamline flow and no drag at the walls of the U-tube. (Hint: use conservation of energy). Express your answer in terms of L and acceleration due to gravity g.

†277

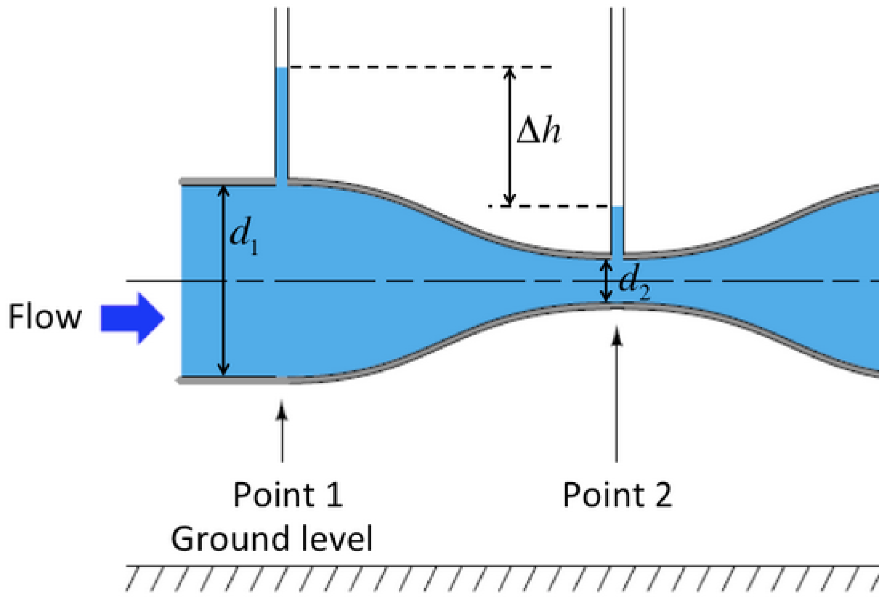
64. Liquid density



A hydrometer is a device that measures the density of a liquid. The one shown in the figure has a spherical bulb of radius R attached to a cylindrical stem of radius r and length ℓ. When placed in a liquid, the device floats as shown in the figure with a length h of stem protruding. Given that the mass of the hydrometer is M, find the density ρ of the liquid. Express your answer in terms of M ,R, r, ℓ and h.

†278

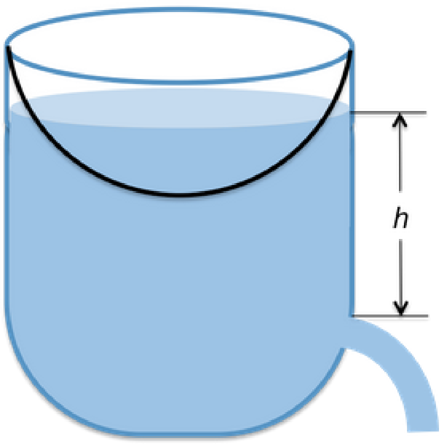
65. Venturi flow meter



A Venturi flow meter is used to measure the the flow velocity of a water main. The water main has a diameter of $d_1 = 40.0$ cm, and the constriction has a diameter of $d_2 = 20.0$ cm. The two vertical pipes are open at the top, and the difference in water level between them is $\Delta h = 2.0$ m. Find the velocity v_m (in m/s), and the volumetric flow rate Q (in $\mathrm{m^3/s}$), of the water in the main.

†278

66. Bucket with a hole



A cylindrical bucket has a small hole at the bottom. The water exiting the hole has velocity v. What is the depth, h, of the water in the bucket?

†280

67. Buoyant force of a balloon

Helium balloons are used regularly in scientific research. A typical balloon would reach an altitude of 40.0 km with an air density of $4.3 \times 10^{-3}\,\mathrm{kg/m^3}$. At this altitude the helium in the balloon would expand to $540\,000.0\,\mathrm{m}^3$. Take $g = 10\,\mathrm{m/s}^2$. Find the buoyant force on the balloon.

†281

Chapter 3

Relativity

3.1 Special relativity

When particles move Extremely Fast$^{\text{TM}}$, Newtonian Dynamics becomes inaccurate and is replaced by Einstein's Special Theory of Relativity (1905).

Its effects are noticeable only when particles approach to the speed of light,

$$c = 299\,792\,458\,\mathrm{m\,s^{-1}} \approx 3\times 10^8\,\mathrm{m\,s^{-1}}$$

This is ***really fast***.

The Special Theory of Relativity rests on the following postulate:

The laws of physics are the same in all inertial frames

This is the principle of relativity familiar to Galileo. Galilean relativity mentioned in the first chapter satisfies this postulate for dynamics. People then thought that Galilean relativity is what the world obeys. However, it turns out that there is a whole family of solutions that satisfy the postulate (for dynamics), and Galilean relativity is just one of them.

This is not a problem (yet), since Galilean relativity seems so intuitive, and we might as well take it to be the true one. However, it turns out that solving Maxwell's equations of electromagnetism gives an explicit value of the speed of light, c. This is independent of the frame of reference. So the speed of light must be the same in every inertial frame.

This is not compatible with Galilean relativity.

Consider the two inertial frames S and S', moving with relative velocity v. Then if light has velocity c in S, then Galilean relativity predicts it has velocity $c - v$ in S', which is wrong.

Therefore, we need to find a different solution to the principle of relativity that preserves the speed of light.

The Lorentz transformation

Consider again inertial frames S and S' whose origins coincide at $t = t' = 0$. For now, neglect the y and z directions, and consider the relationship between (x, t) and (x', t'). The general form is

$$x' = f(x,t), \quad t' = g(x,t),$$

for some functions f and g. This is not very helpful.

In any inertial frame, a free particle moves with constant velocity. So straight lines in (x, t) must map into straight lines in (x', t'). Therefore the relationship must be linear.

Given that the origins of S and S' coincide at $t = t' = 0$, and S' moves with velocity v relative to S, we know that the line $x = vt$ must map into $x' = 0$.

Combining these two information, the transformation must be of the form

$$x' = \gamma(x - vt), \tag{1}$$

for some factor γ that may depend on $|v|$ (***not*** v itself. We can use symmetry arguments to show that γ should take the same value for velocities v and $-v$).

Note that Galilean transformation is compatible with this – just take γ to be always 1.

Now reverse the roles of the frames. From the perspective S', S moves with velocity $-v$. A similar argument leads to

$$x = \gamma(x' + vt'), \tag{2}$$

with the same factor γ, since γ only depends on $|v|$. Now consider a light ray (or photon) passing through the origin $x = x' = 0$ at $t = t' = 0$. Its trajectory in S is

$$x = ct.$$

We want a γ such that the trajectory in S' is

$$x' = ct'$$

as well, so that the speed of light is the same in each frame. Substitute these into (1) and (2)

$$\begin{aligned} ct' &= \gamma(c - v)t \\ ct &= \gamma(c + v)t' \end{aligned}$$

Multiply the two equations together and divide by tt' to obtain

$$c^2 = \gamma^2(c^2 - v^2).$$

So

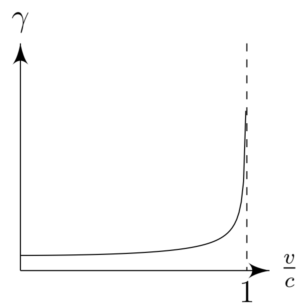
$$\gamma = \sqrt{\frac{c^2}{c^2 - v^2}} = \frac{1}{\sqrt{1 - (v/c)^2}}.$$

Definition (Lorentz factor)**.** The ***Lorentz factor*** is

$$\gamma = \frac{1}{\sqrt{1 - (v/c)^2}}.$$

Note that

- $\gamma \geq 1$ and is an increasing function of $|v|$.
- When $v \mathcal{L} c$, then $\gamma \approx 1$, and we recover the Galilean transformation.
- When $|v| \to c$, then $\gamma \to \infty$.
- If $|v| \geq c$, then γ is imaginary, which is physically impossible (or at least ***weird***).
- If we take $c \to \infty$, then $\gamma = 1$. So Galilean transformation is the transformation we will have if light is infinitely fast. Alternatively, in the world of Special Relativity, the speed of light is "infinitely fast".



For the sense of scale, we have the following values of γ at different speeds:

- $\gamma = 2$ when $v = 0.866c$.
- $\gamma = 10$ when $v = 0.9949$.
- $\gamma = 20$ when $v = 0.999c$.

We still have to solve for the relation between t and t'. Eliminate x between (1) and (2) to obtain

$$x = \gamma(\gamma(x - vt) + vt').$$

So

$$t' = \gamma t - (1 - \gamma^{-2})\frac{\gamma x}{v} = \gamma\left(t - \frac{v}{c^2}x\right).$$

So we have

Law (Principle of Special Relativity)**.** Let S and S' be inertial frames, moving at the relative velocity of v. Then

$$\begin{aligned} x' &= \gamma(x - vt)\\ t' &= \gamma\left(t - \frac{v}{c^2}x\right), \end{aligned}$$

where

$$\gamma = \frac{1}{\sqrt{1 - (v/c)^2}}.$$

This is the ***Lorentz transformations*** in the standard configuration (in one spatial dimension).

The above is the form the Lorentz transformation is usually written, and is convenient for actual calculations. However, this lacks symmetry between space and time. To display the symmetry, one approach is to use units such that $c = 1$. Then we have

$$\begin{aligned} x' &= \gamma(x - vt),\\ t' &= \gamma(t - vx). \end{aligned}$$

Alternatively, if we want to keep our c's, instead of comparing x and t, which have different units, we can compare x and ct. Then we have

$$\begin{aligned} x' &= \gamma\left(x - \frac{v}{c}(ct)\right), \\ ct' &= \gamma\left(ct - \frac{v}{c}x\right). \end{aligned}$$

Symmetries aside, to express x, t in terms of x', t', we can invert this linear mapping to find (after some algebra)

$$\begin{aligned} x &= \gamma(x' + vt') \\ t &= \gamma\left(t' + \frac{v}{c^2}x'\right) \end{aligned}$$

Directions perpendicular to the relative motion of the frames are unaffected:

$$\begin{aligned} y' &= y \\ z' &= z \end{aligned}$$

Now we check that the speed of light is really invariant:
For a light ray travelling in the x direction in S:

$$x = ct, \quad y = 0, \quad z = 0.$$

In S', we have

$$\frac{x'}{t'} = \frac{\gamma(x - vt)}{\gamma(t - vx/c^2)} = \frac{(c - v)t}{(1 - v/c)t} = c,$$

as required.
For a light ray travelling in the Y direction in S,

$$x = 0, \quad y = ct, \quad z = 0.$$

In S',

$$\frac{x'}{t'} = \frac{\gamma(x - vt)}{\gamma(t - vx/c^2)} = -v,$$

and

$$\frac{y'}{t'} = \frac{y}{\gamma(t - vx/c^2} = \frac{c}{\gamma},$$

and

$$z' = 0.$$

So the speed of light is

$$\frac{\sqrt{x'^2 + y'^2}}{t'} = \sqrt{v^2 + \gamma^{-2}c^2} = c,$$

as required.

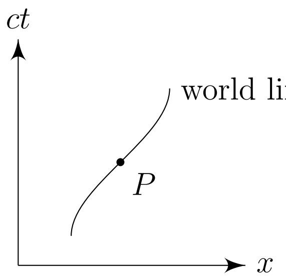
More generally, the Lorentz transformation implies

$$\begin{aligned}
c^2t'^2 - r'^2 &= c^2t'^2 - x'^2 - y'^2 - z'^2 \\
&= c^2\gamma^2\left(t - \frac{v}{c^2}x\right)^2 - \gamma^2(x-vt)^2 - y^2 - z^2 \\
&= \gamma^2\left(1 - \frac{v^2}{c^2}\right)(c^2t^2 - x^2) - y^2 - z^2 \\
&= c^2t^2 - x^2 - y^2 - z^2 \\
&= c^2t^2 - r^2.
\end{aligned}$$

We say that the quantity $c^2t^2 - x^2 - y^2 - z^2$ is ***Lorentz-invariant***.
So if $\frac{r}{t} = c$, then $\frac{r'}{t'} = c$ also.

Spacetime diagrams

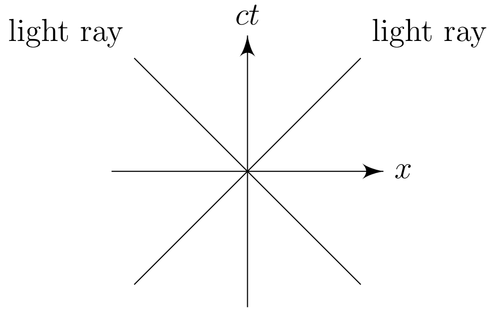
It is often helpful to plot out what is happening on a diagram. We plot them on a graph, where the position x is on the horizontal axis and the time ct is on the vertical axis. We use ct instead of t so that the dimensions make sense.



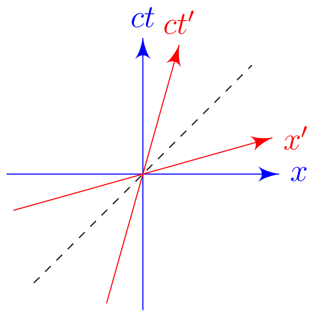
Definition (Spacetime)**.** The union of space and time in special relativity is called ***Minkowski spacetime***. Each point P represents an ***event***, labelled by coordinates (ct, x) (note the order!).

A particle traces out a ***world line*** in spacetime, which is straight if the particle moves uniformly.

Light rays moving in the x direction have world lines inclined at $45°$.



We can also draw the axes of S', moving in the x direction at velocity v relative to S. The ct' axis corresponds to $x' = 0$, i.e. $x = vt$. The x' axis corresponds to $t' = 0$, i.e. $t = vx/c^2$.



Note that the x' and ct' axes are ***not*** orthogonal, but are symmetrical about the diagonal (dashed line). So they agree on where the world line of a light ray should lie on.

Relativistic physics

Now we can look at all sorts of relativistic weirdness!

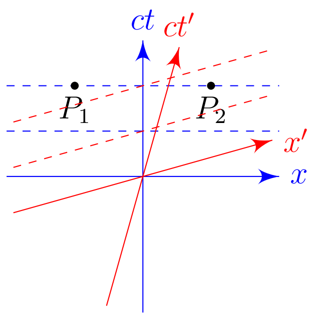
Simultaneity

The first relativistic weirdness is that different frames disagree on whether two evens are simultaneous

Definition (Simultaneous events)**.** We say two events P_1 and P_2 are simultaneous in the frame S if $t_1 = t_2$.

They are represented in the following spacetime diagram by horizontal dashed lines.
However, events that are simultaneous in S' have equal values of t', and so lie on lines

$$ct - \frac{v}{c}x = \text{constant}.$$

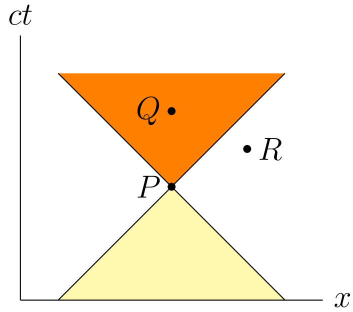


The lines of simultaneity of S' and those of S are different, and events simultaneous in S need not be simultaneous in S'. So simultaneity is relative. S thinks P_1 and P_2 happened at the same time, while S' thinks P_2 happens first.
Note that this is genuine disagreement. It is not due to effects like, it takes time for the light conveying the information to different observers. Our account above already takes that into account (since the whole discussion does not involve specific observers).

Causality

Although different people may disagree on the temporal order of events, the consistent ordering of cause and effect can be ensured.

Since things can only travel at at most the speed of light, P cannot affect R if R happens a millisecond after P but is at millions of galaxies away. We can draw a ***light cone*** that denotes the regions in which things can be influenced by P. These are the regions of space-time light (or any other particle) can possibly travel to. P can only influence events within its ***future light cone***, and ***be influenced*** by events within its ***past light cone***.



All observers agree that Q occurs after P. Different observers may disagree on the temporal ordering of P and R. However, since nothing can travel faster than light, P and R cannot influence each other. Since everyone agrees on how fast light travels, they also agree on the light cones, and hence causality. So philosophers are happy.

Time dilation

Suppose we have a clock that is stationary in S' (which travels at constant velocity v with respect to inertial frame S) ticks at constant intervals $\Delta t'$. What is the interval between ticks in S?

Lorentz transformation gives

$$t = \gamma\left(t' + \frac{v}{c^2}x'\right).$$

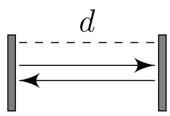
Since x' = constant for the clock, we have

$$\Delta t = \gamma \Delta t' > \Delta t'.$$

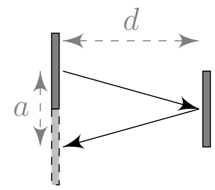
So the interval measured in S is greater! So moving clocks run slowly.

A non-mathematical explanation comes from Feynman (not lectured): Suppose we have a very simple clock: We send a light beam towards a mirror, and wait for it to reflect back. When the clock detects the reflected light, it ticks, and then sends the next light beam.

Then the interval between two ticks is the distance $2d$ divided by the speed of light.



From the point of view of an observer moving downwards, by the time light reaches the right mirror, it would have moved down a bit. So S sees



However, the distance travelled by the light beam is now $\sqrt{(2d)^2 + a^2} > 2d$. Since they agree on the speed of light, it must have taken longer for the clock to receive the reflected light in S. So the interval between ticks are longer.

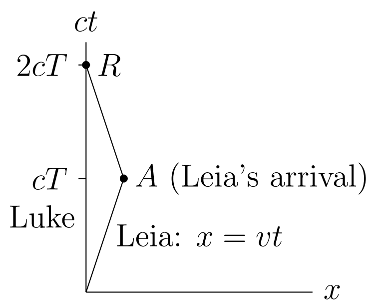
By the principle of relativity, all clocks must measure the same time dilation, or else we can compare the two clocks and know if we are "moving".

This is famously evidenced by muons. Their half-life is around 2 microseconds (i.e. on average they decay to something else after around 2 microseconds). They are created when cosmic rays bombard the atmosphere. However, even if they travel at the speed of light, 2 microseconds only allows it to travel 600 m, certainly not sufficient to reach the surface of Earth. However, we observe ***lots*** of muons on Earth. This is because muons are travelling so fast that their clocks run really slowly.

The twin paradox

Consider two twins: Luke and Leia. Luke stays at home. Leia travels at a constant speed v to a distant planet P, turns around, and returns at the same speed.

In Luke's frame of reference,



Leia's arrival (A) at P has coordinates

$$(ct, x) = (cT, vT).$$

The time experienced by Leia on her outward journey is

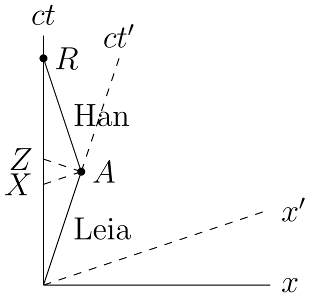
$$T' = \gamma\left(T - \frac{v}{c^2}T\right) = \frac{T}{\gamma}.$$

By Leia's return R, Luke has aged by $2T$, but Leia has aged by $\frac{2T}{\gamma} < 2T$. So she is younger than Luke, because of time dilation.

The paradox is: From Leia's perspective, Luke travelled away from her at speed and the returned, so he should be younger than her!

Why is the problem not symmetric?

We can draw Leia's initial frame of reference in dashed lines:



In Leia's frame, by the time she arrives at A, she has experienced a time $T' = \frac{T}{\gamma}$ as shown above. This event is simultaneous with event X in Leia's frame. Then in Luke's frame, the coordinates of X are

$$(ct, x) = \left(\frac{cT'}{\gamma}, 0\right) = \left(\frac{cT}{\gamma^2}, 0\right),$$

obtained through calculations similar to that above. So Leia thinks Luke has aged less by a factor of $1/\gamma^2$. At this stage, the problem ***is*** symmetric, and Luke also thinks Leia has aged less by a factor of $1/\gamma^2$.

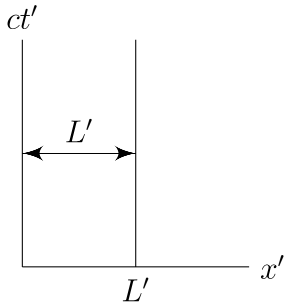
Things change when Leia turns around and changes frame of reference. To understand this better, suppose Leia meets a friend, Han, who is just leaving P at speed v. On his journey back, Han also thinks Luke ages T/γ^2. But in his frame of reference, his departure is simultaneous with Luke's event Z, not X, since he has different lines of simultaneity.

So the asymmetry between Luke and Leia occurs when Leia turns around. At this point, she sees Luke age rapidly from X to Z.

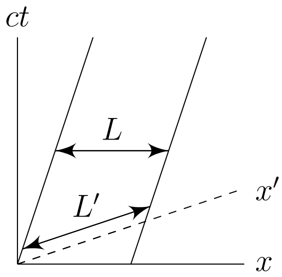
Length contraction

A rod of length L' is stationary in S'. What is its length in S?

In S', then length of the rod is the distance between the two ends at the same time. So we have



In S, we have



The lines $x' = 0$ and $x' = L'$ map into $x = vt$ and $x = vt + L'/\gamma$. So the length in S is $L = L'/\gamma < L'$. Therefore moving objects are contracted in the direction of motion.

Definition (Proper length)**.** The ***proper length*** is the length measured in an object's rest frame.

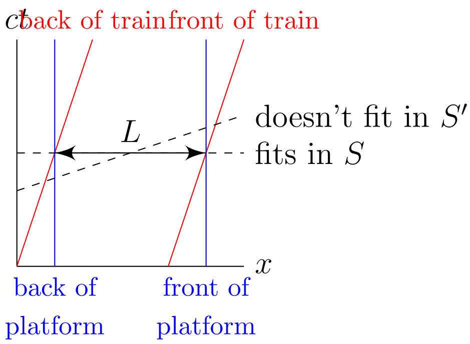
This is analogous to the fact that if you view a bar from an angle, it looks shorter than if you view it from the front. In relativity, what causes the contraction is not a spatial rotation, but a spacetime ***hyperbolic*** rotation.

Question: does a train of length $2L$ fit alongside a platform of length L if it travels through the station at a speed v such that $\gamma = 2$?

For the system of observers on the platform, the train contracts to a length $2L/\gamma = L$. So it fits.

But for the system of observers on the train, the platform contracts to length $L/\gamma = L/2$, which is much too short!

This can be explained by the difference of lines of simultaneity, since length is the distance between front and back ***at the same time***.



Composition of velocities

A particle moves with constant velocity u' in frame S', which moves with velocity v relative to S. What is its velocity u in S?

The world line of the particle in S' is

$$x' = u't'.$$

In S, using the inverse Lorentz transformation,

$$u = \frac{x}{t} = \frac{\gamma(x' + vt')}{\gamma(t' + (v/c^2)x')} = \frac{u't' + vt'}{t' + (v/c^2)u't'} = \frac{u' + v}{1 + u'v/c^2}.$$

This is the formula for the relativistic composition of velocities.
The inverse transformation is found by swapping u and u', and swapping the sign of v, i.e.

$$u' = \frac{u - v}{1 - uv/c^2}.$$

Note the following:

– if $u'v\mathcal{L}c^2$, then the transformation reduces to the standard Galilean addition of velocities $u \approx u' + v$.

– u is a monotonically increasing function of u for any constant v (with $|v| < c$).

– When $u' = \pm c$, $u = u'$ for any v, i.e. the speed of light is constant in all frames of reference.

– Hence $|u'| < c$ iff $|u| < c$. This means that we cannot reach the speed of light by composition of velocities.

Geometry of spacetime

We'll now look at the geometry of spacetime, and study the properties of vectors in this spacetime. While spacetime has 4 dimensions, and each point can be represented by 4 real numbers, this is not ordinary $\mathbb{R}^4$. This can be seen when changing coordinate systems, instead of rotating the axes like in $\mathbb{R}^4$, we "squash" the axes towards the diagonal, which is a ***hyperbolic rotation***. In particular, we will have a different notion of a dot product. We say that this space has dimension $d = 1 + 3$.

The invariant interval

In regular Euclidean space, given a vector $\mathbf{x}$, all coordinate systems agree on the length $|\mathbf{x}|$. In Minkowski space, they agree on something else.
Consider events P and Q with coordinates (ct_1, x_1) and (ct_2, x_2) separated by $\Delta t = t_2 - t_1$ and $\Delta x = x_2 - x_1$.

Definition (Invariant interval)**.** The ***invariant interval*** or ***spacetime interval*** between P and Q is defined as

$$\Delta s^2 = c^2\Delta t^2 - \Delta x^2.$$

Note that this quantity Δs^2 can be both positive or negative — so Δs might be imaginary!

Proposition. All inertial observers agree on the value of Δs^2.

Proof.

$$\begin{aligned} c^2\Delta t'^2 - \Delta x'^2 &= c^2\gamma^2\left(\Delta t - \frac{v}{c^2}\Delta x\right)^2 - \gamma^2(\Delta x - v\Delta t)^2 \\ &= \gamma^2\left(1 - \frac{v^2}{c^2}\right)(c^2\Delta t^2 - \Delta x^2) \\ &= c^2\Delta t^2 - \Delta x^2. \end{aligned}$$ □

In three spatial dimensions,

$$\Delta s^2 = c^2\Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2.$$

We take this as the "distance" between the two points. For two infinitesimally separated events, we have

Definition (Line element)**.** The ***line element*** is

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2.$$

Definition (Timelike, spacelike and lightlike separation)**.** Events with $\Delta s^2 > 0$ are ***timelike separated***. It is possible to find inertial frames in which the two events occur in the same position, and are purely separated by time. Timelike-separated events lie within each other's light cones and can influence one another.

Events with $\Delta s^2 < 0$ are ***spacelike separated***. It is possible to find inertial frame in which the two events occur in the same time, and are purely separated by space. Spacelike-separated events lie out of each other's light cones and cannot influence one another.

Events with $\Delta s^2 = 0$ are ***lightlike*** or ***null separated***. In all inertial frames, the events lie on the boundary of each other's light cones. e.g. different points in the trajectory of a photon are lightlike separated, hence the name.

Note that $\Delta s^2 = 0$ does not imply that P and Q are the same event.

The Lorentz group

The coordinates of an event P in frame S can be written as a ***4-vector*** (i.e. 4-component vector) X. We write

$$X = \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}$$

The invariant interval between the origin and P can be written as an inner product

$$X \cdot X = X^T\eta X = c^2t^2 - x^2 - y^2 - z^2,$$

where

$$\eta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

4-vectors with $X \cdot X > 0$ are called timelike, and those $X \cdot X < 0$ are spacelike. If $X \cdot X = 0$, it is lightlike or null.

A Lorentz transformation is a linear transformation of the coordinates from one frame S to another S', represented by a 4×4 tensor ("matrix"):

$$X' = \Lambda X$$

Lorentz transformations can be defined as those that leave the inner product invariant:

$$(\forall X)(X' \cdot X' = X \cdot X),$$

which implies the matrix equation

$$\Lambda^T \eta \Lambda = \eta. \tag{$*$}$$

These also preserve $X \cdot Y$ if X and Y are both 4-vectors.

Two classes of solution to this equation are:

$$\Lambda = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & & & \\ 0 & & R & \\ 0 & & & \end{pmatrix},$$

where R is a 3×3 orthogonal matrix, which rotates (or reflects) space and leaves time intact; and

$$\Lambda = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

where $\beta = \frac{v}{c}$, and $\gamma = 1/\sqrt{1-\beta^2}$. Here we leave the y and z coordinates intact, and apply a Lorentz boost along the x direction.

The set of all matrices satisfying equation $(*)$ form the ***Lorentz group*** $O(1,3)$. It is generated by rotations and boosts, as defined above, which includes the absurd spatial reflections and time reversal.

The subgroup with $\det \Lambda = +1$ is the ***proper Lorentz group*** $SO(1,3)$.

The subgroup that preserves spatial orientation and the direction of time is the ***restricted Lorentz group*** $SO^+(1,3)$. Note that this is different from $SO(1,3)$, since if you do ***both*** spatial reflection and time reversal, the determinant of the matrix is still positive. We want to eliminate those as well!

Rapidity

Focus on the upper left 2×2 matrix of Lorentz boosts in the x direction. Write

$$\Lambda[\beta] = \begin{pmatrix} \gamma & -\gamma\beta \\ -\gamma\beta & \gamma \end{pmatrix}, \quad \gamma = \frac{1}{\sqrt{1-\beta^2}}.$$

Combining two boosts in the x direction, we have

$$\Lambda[\beta_1]\Lambda[\beta_2] = \begin{pmatrix} \gamma_1 & -\gamma_1\beta_1 \\ -\gamma_1\beta_1 & \gamma_1 \end{pmatrix} \begin{pmatrix} \gamma_2 & -\gamma_2\beta_2 \\ -\gamma_2\beta_2 & \gamma_2 \end{pmatrix} = \Lambda \left[\frac{\beta_1 + \beta_2}{1 + \beta_1\beta_2}\right]$$

after some messy algebra. This is just the velocity composition formula as before.

This result does not look nice. This suggests that we might be writing things in the wrong way.

We can compare this with spatial rotation. Recall that

$$R(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

with

$$R(\theta_1)R(\theta_2) = R(\theta_1 + \theta_2).$$

For Lorentz boosts, we can define

Definition (Rapidity)**.** The ***rapidity*** of a Lorentz boot is ϕ such that

$$\beta = \tanh\phi, \quad \gamma = \cosh\phi, \quad \gamma\beta = \sinh\phi.$$

Then

$$\Lambda[\beta] = \begin{pmatrix} \cosh\phi & -\sinh\phi \\ -\sinh\phi & \cosh\phi \end{pmatrix} = \Lambda(\phi).$$

The rapidities add like rotation angles:

$$\Lambda(\phi_1)\Lambda(\phi_2) = \Lambda(\phi_1 + \phi_2).$$

This shows the close relation betweens spatial rotations and Lorentz boosts. Lorentz boots are simply ***hyperbolic*** rotations in spacetime!

Relativistic kinematics

In Newtonian mechanics, we describe a particle by its position $\mathbf{x}(t)$, with its velocity being $\mathbf{u}(t) = \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}$.

In relativity, this is unsatisfactory. In special relativity, space and time can be mixed together by Lorentz boosts, and we prefer not to single out time from space. For example, when we write the 4-vector X, we put in both the time and space components, and Lorentz transformations are 4×4 matrices that act on X.

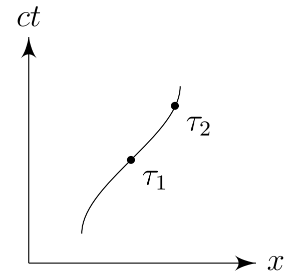
In the definition of velocity, however, we are differentiating space with respect to time, which is rather weird. First of all, we need something to replace time. Recall that we defined "proper length" as the length in the item in its rest frame. Similarly, we can define the ***proper time***.

Definition (Proper time)**.** The ***proper time*** τ is defined such that

$$\Delta\tau = \frac{\Delta s}{c}$$

τ is the time experienced by the particle, i.e. the time in the particles rest frame.

The world line of a particle can be parametrized using the proper time by $t(\tau)$ and $\mathbf{x}(\tau)$.



Infinitesimal changes are related by

$$\mathrm{d}\tau = \frac{\mathrm{d}s}{c} = \frac{1}{c}\sqrt{c^2\,\mathrm{d}t^2 - |\mathrm{d}\mathbf{x}|^2} = \sqrt{1 - \frac{|\mathbf{u}|^2}{c^2}}\,\mathrm{d}t.$$

Thus

$$\frac{\mathrm{d}t}{\mathrm{d}\tau} = \gamma_u$$

with

$$\gamma_u = \frac{1}{\sqrt{1 - \frac{|\mathbf{u}|^2}{c^2}}}.$$

The total time experienced by the particle along a segment of its world line is

$$T = \int\,\mathrm{d}\tau = \int \frac{1}{\gamma_u}\,\mathrm{d}t.$$

We can then define the ***position 4-vector*** and ***4-velocity***.

Definition (Position 4-vector and 4-velocty)**.** The ***position 4-vector*** is

$$X(\tau) = \begin{pmatrix} ct(\tau) \\ \mathbf{x}(\tau) \end{pmatrix}.$$

Its ***4-velocity*** is defined as

$$U = \frac{\mathrm{d}X}{\mathrm{d}\tau} = \begin{pmatrix} c\frac{\mathrm{d}t}{\mathrm{d}\tau} \\ \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}\tau} \end{pmatrix} = \frac{\mathrm{d}t}{\mathrm{d}\tau}\begin{pmatrix} c \\ \mathbf{u} \end{pmatrix} = \gamma_u \begin{pmatrix} c \\ \mathbf{u} \end{pmatrix},$$

where $\mathbf{u} = \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}$.

Another common notation is

$$X = (ct, \mathbf{x}), \quad U = \gamma_u(c, \mathbf{u}).$$

If frames S and S' are related by $X' = \Lambda X$, then the 4-velocity also transforms as $U' = \Lambda U$.

Definition (4-vector)**.** A ***4-vector*** is a 4-component vectors that transforms in this way under a Lorentz transformation, i.e. $X' = \Lambda X$.

When using suffix notation, the indices are written above (superscript) instead of below (subscript). The indices are written with Greek letters which range from 0 to 3. So we have X^μ instead of X_i, for $\mu = 0, 1, 2, 3$. If we write X_μ instead, it means a different thing. This will be explained more in-depth in the electromagnetism course (and you'll get more confused!).

U is a 4-vector because X is a 4-vector and τ is a Lorentz invariant. Note that $\mathrm{d}X/\mathrm{d}t$ is ***not*** a 4-vector.

Note that this definition of 4-vector is analogous to that of a tensor — things that transform nicely according to our rules. Then τ would be a scalar, i.e. rank-0 tensor, while t is just a number, not a scalar.

For any 4-vector U, the inner product $U \cdot U = U' \cdot U'$ is Lorentz invariant, i.e. the same in all inertial frames. In the rest frame of the particle, $U = (c, 0)$. So $U \cdot U = c^2$.

In any other frame, $Y = \gamma_u(c, \mathbf{u})$. So

$$Y \cdot Y = \gamma_u^2(c^2 - |\mathbf{u}|^2) = c^2$$

as expected.

Transformation of velocities revisited

We have seen that velocities cannot be simply added in relativity. However, the 4-velocity does transform linearly, according to the Lorentz transform:

$$U' = \Lambda U.$$

In frame S, consider a particle moving with speed u at an angle θ to the x axis in the xy plane. This is the most general case for motion not parallel to the Lorentz boost.

Its 4-velocity is

$$U = \begin{pmatrix} \gamma_u c \\ \gamma_u u \cos\theta \\ \gamma_u u \sin\theta \\ 0 \end{pmatrix}, \quad \gamma_u = \frac{1}{\sqrt{1 - u^2/c^2}}.$$

With frames S and S' in standard configuration (i.e. origin coincide at $t = 0$, S' moving in x direction with velocity v relative to S),

$$U' = \begin{pmatrix} \gamma_{u'} c \\ \gamma_{u'} u' \cos\theta' \\ \gamma_{u'} u' \sin\theta' \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma_v & -\gamma_v v/c & 0 & 0 \\ -\gamma_v v/c & \gamma_v & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \gamma_u c \\ \gamma_u u \cos\theta \\ \gamma_u u \sin\theta \\ 0 \end{pmatrix}$$

Instead of evaluating the whole matrix, we can divide different rows to get useful results.

The ratio of the first two lines gives

$$u' \cos\theta' = \frac{u \cos\theta - v}{1 - \frac{uv}{c^2}\cos\theta},$$

just like the composition of parallel velocities.

The ratio of the third to second line gives

$$\tan\theta' = \frac{u\sin\theta}{\gamma_v(u\cos\theta - v)},$$

which describes ***aberration***, a change in the direction of motion of a particle due to the motion of the observer. Note that this isn't just a relativistic effect! If you walk in the rain, you have to hold your umbrella obliquely since the rain seems to you that they are coming from an angle. The relativistic part is the γ_v factor in the denominator.

This is also seen in the aberration of starlight ($u = c$) due to the Earth's orbital motion. This causes small annual changes in the apparent positions of stars.

4-momentum

Definition (4-momentum)**.** The ***4-momentum*** of a particle of mass m is

$$P = mU = m\gamma_u \begin{pmatrix} c \\ \mathbf{u} \end{pmatrix}$$

The 4-momentum of a system of particles is the sum of the 4-momentum of the particles, and is conserved in the absence of external forces.

The spatial components of P are the ***relativistic 3-momentum***,

$$\mathbf{p} = m\gamma_u\mathbf{u},$$

which differs from the Newtonian expression by a factor of γ_u. Note that $|\mathbf{p}| \to \infty$ as $|\mathbf{u}| \to c$.

What is the interpretation of the time component P^0 (i.e. the first time component of the P vector)? We expand for $|\mathbf{u}|\mathcal{L}c$:

$$P^0 = m\gamma c = \frac{mc}{\sqrt{1 - |\mathbf{u}|^2/c^2}} = \frac{1}{c}\left(mc^2 + \frac{1}{2}m|\mathbf{u}|^2 + \cdots\right).$$

We have a constant term mc^2 plus a kinetic energy term $\frac{1}{2}m|\mathbf{u}|^2$, plus more tiny terms, all divided by c. So this suggests that P^0 is indeed the energy for a particle, and the remaining $\cdots$ terms are relativistic corrections for our old formula $\frac{1}{2}m|\mathbf{u}|^2$ (the mc^2 term will be explained later). So we interpret P as

$$P = \begin{pmatrix} E/c \\ \mathbf{p} \end{pmatrix}$$

Definition (Relativistic energy)**.** The ***relativistic energy*** of a particle is $E = P^0 c$. So

$$E = m\gamma c^2 = mc^2 + \frac{1}{2}m|\mathbf{u}|^2 + \cdots$$

Note that $E \to \infty$ as $|\mathbf{u}| \to c$.

For a stationary particle, we obtain

$$E = mc^2.$$

This implies that mass is a form of energy. m is sometimes called the ***rest mass***.

The energy of a moving particle, $m\gamma_u c^2$, is the sum of the rest energy mc^2 and kinetic energy $m(\gamma_u - 1)c^2$.

Since $P \cdot P = \frac{E^2}{c^2} - |\mathbf{p}|^2$ is a Lorentz invariant (lengths of 4-vectors are always Lorentz invariant) and equals m^2c^2 in the particle's rest frame, we have the general relation between energy and momentum

$$E^2 = |\mathbf{p}|^2c^2 + m^2c^4$$

In Newtonian physics, mass and energy are separately conserved. In relativity, mass is not conserved. Instead, it is just another form of energy, and the total energy, including mass energy, is conserved.

Mass can be converged into kinetic energy and vice versa (e.g. atomic bombs!)

Massless particles

Particles with zero mass ($m = 0$), e.g. photons, can have non-zero momentum and energy because they travel at the speed of light ($\gamma = \infty$).

In this case, $P \cdot P = 0$. So massless particles have light-like (or null) trajectories, and no proper time can be defined for such particles.

Other massless particles in the Standard Model of particle physics include the gluon.

For these particles, energy and momentum are related by

$$E^2 = |\mathbf{p}|^2c^2.$$

So

$$E = |\mathbf{p}|c.$$

Thus

$$P = \frac{E}{c}\begin{pmatrix} 1 \\ \mathbf{n} \end{pmatrix},$$

where $\mathbf{n}$ is a unit (3-)vector in the direction of propagation.

According to quantum mechanics, fundamental "particles" aren't really particles but have both particle-like and wave-like properties (if that sounds confusing, yes it is!). Hence we can assign it a ***de Broglie wavelength***, according to the ***de Broglie relation***:

$$|\mathbf{p}| = \frac{h}{\lambda}$$

where $h \approx 6.63 \times 10^{-34}\,\mathrm{m^2\,kg\,s^{-1}}$ is ***Planck's constant***.

For massless particles, this is consistent with ***Planck's relation***:

$$E = \frac{hc}{\lambda} = h\nu,$$

where $\nu = \frac{c}{\lambda}$ is the ***wave frequency***.

Newton's second law in special relativity

Definition (4-force)**.** The ***4-force*** is

$$F = \frac{\mathrm{d}P}{\mathrm{d}\tau}$$

This equation is the relativistic counterpart to Newton's second law.
It is related to the 3-force $\mathbf{F}$ by

$$F = \gamma_u \begin{pmatrix} \mathbf{F} \cdot \mathbf{u}/c \\ \mathbf{F} \end{pmatrix}$$

Expanding the definition of the 4-force componentwise, we obtain

$$\frac{\mathrm{d}E}{\mathrm{d}\tau} = \gamma_u \mathbf{F} \cdot \mathbf{u} \Rightarrow \frac{\mathrm{d}E}{\mathrm{d}t} = \mathbf{F} \cdot \mathbf{u}$$

and

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}\tau} = \gamma_u \mathbf{F} \Rightarrow \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = \mathbf{F}$$

Equivalently, for a particle of mass m,

$$F = mA,$$

where

$$A = \frac{\mathrm{d}U}{\mathrm{d}\tau}$$

is the 4-acceleration.
We have

$$U = \gamma_u \begin{pmatrix} c \\ \mathbf{u} \end{pmatrix}$$

So

$$A = \gamma_u \frac{\mathrm{d}U}{\mathrm{d}t} = \gamma_u \begin{pmatrix} \dot{\gamma}_u c \\ \gamma_u \mathbf{a} + \dot{\gamma}_u \mathbf{u}. \end{pmatrix}$$

where $\mathbf{a} = \frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t}$ and $\dot{\gamma}_u = \gamma_u^3 \frac{\mathbf{a} \cdot \mathbf{u}}{c^2}$.
In the instantaneous rest frame of a particle, $\mathbf{u} = \mathbf{0}$ and $\gamma_u = 1$. So

$$U = \begin{pmatrix} c \\ \mathbf{0} \end{pmatrix}, \quad A = \begin{pmatrix} 0 \\ \mathbf{a} \end{pmatrix}$$

Then $\mathbf{U} \cdot \mathbf{A} = 0$. Since this is a Lorentz invariant, we have $\mathbf{U} \cdot \mathbf{A} = 0$ in all frames.

Particle physics

Many problems can be solved using the conservation of 4-momentum,

$$P = \begin{pmatrix} E/c \\ \mathbf{p} \end{pmatrix},$$

for a system of particles.

Definition (Center of momentum frame)**.** The ***center of momentum (CM) frame***, or ***zero momentum frame***, is an inertial frame in which the total 3-momentum is $\sum \mathbf{p} = 0$.

This exists unless the system consists of one or more massless particle moving in a single direction.

Particle decay

A particle of mass m_1 decays into two particles of masses m_2 and m_2.
We have

$$P_1 = P_2 + P_3.$$

i.e.

$$\begin{aligned} E_1 &= E_2 + E_3 \\ \mathbf{p}_1 &= \mathbf{p}_2 + \mathbf{p}_3. \end{aligned}$$

In the CM frame (i.e. the rest frame of the original particle),

$$\begin{aligned} E_1 = m_1c^2 &= \sqrt{|\mathbf{p}_2|^2c^2 + m_2^2c^4} + \sqrt{|\mathbf{p}_3|^2c^2 + m_2^2c^4} \\ &\geq m_2c^2 + m_3c^2. \end{aligned}$$

So decay is possible only if

$$m_1 \geq m_2 + m_3.$$

(Recall that mass is not conserved in relativity!)

Example. A possible decay path of the Higgs' particle can be written as

$$\begin{gathered} \mathrm{h} \to \gamma\gamma \\ \text{Higgs' particle} \to \text{2 photons} \end{gathered}$$

This is possible by the above criterion, because $m_\mathrm{h} \geq 0$, while $m_\gamma = 0$.
The full conservation equation is

$$P_\mathrm{h} = \begin{pmatrix} m_\mathrm{h}c \\ \mathbf{0} \end{pmatrix} = P_{\gamma_1} + P_{\gamma_2}$$

So

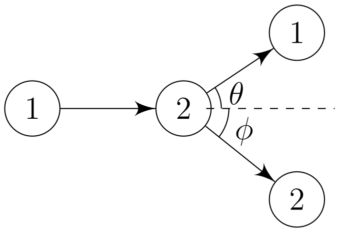
$$\begin{aligned} \mathbf{p}_{\gamma_1} &= \mathbf{p}_{\gamma_2} \\ E_{\gamma_1} = E_{\gamma_2} &= \frac{1}{2}m_\mathrm{h}c^2. \end{aligned}$$

Particle scattering

When two particles collide and retain heir identities, the total 4-momentum is conserved:

$$P_1 + P_2 = P_3 + P_4$$

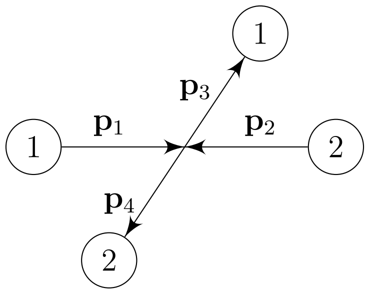
In the laboratory frame S, suppose that particle 1 travels with speed u and collides with particle 2 (at rest).



In the CM frame S',

$$\mathbf{p}_1' + \mathbf{p}_2' = 0 = \mathbf{p}_3' + \mathbf{p}_4'.$$

Both before and after the collision, the two particles have equal and opposite 3-momentum.



The scattering angle θ' is undetermined and can be thought of as being random. However, we can derive some conclusions about the angles θ and ϕ in the laboratory frame.

(staying in S' for the moment) Suppose the particles have equal mass m. They then have the same speed v in S'.

Choose axes such that

$$P_1' = \begin{pmatrix} m\gamma_v c \\ m\gamma_v v \\ 0 \\ 0 \end{pmatrix}, \quad P_2' = \begin{pmatrix} m\gamma_v c \\ -m\gamma_v v \\ 0 \\ 0 \end{pmatrix}$$

and after the collision,

$$P_3' = \begin{pmatrix} m\gamma_v c \\ m\gamma_v v\cos\theta' \\ m\gamma_v v\sin\theta' \\ 0 \end{pmatrix}, \quad P_4' = \begin{pmatrix} m\gamma_v c \\ -m\gamma_v v\cos\theta' \\ -m\gamma_v v\sin\theta' \\ 0 \end{pmatrix}.$$

We then use the Lorentz transformation to return to the laboratory frame S. The relative velocity of the frames is v. So the Lorentz transform is

$$\Lambda = \begin{pmatrix} \gamma_v & \gamma_v v/c & 0 & 0 \\ \gamma_v v/c & \gamma_v & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and we find

$$P_1 = \begin{pmatrix} m\gamma_u c \\ m\gamma_u u \\ 0 \\ 0 \end{pmatrix}, \quad P_2 = \begin{pmatrix} mc \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

where

$$u = \frac{2v}{1 + v^2/c^2},$$

(cf. velocity composition formula)
Considering the transformations of P_3' and P_4', we obtain

$$\tan\theta = \frac{\sin\theta'}{\gamma_v(1 + \cos\theta')} = \frac{1}{\gamma_v}\tan\frac{\theta'}{2},$$

and

$$\tan\phi = \frac{\sin\theta'}{\gamma_v(1 - \cos\theta')} = \frac{1}{\gamma_v}\cot\frac{\theta'}{2}.$$

Multiplying these expressions together, we obtain

$$\tan\theta\tan\phi = \frac{1}{\gamma_v^2}.$$

So even though we do not know what θ and ϕ might be, they ***must*** be related by this equation.
In the Newtonian limit, where $|\mathbf{v}|\mathcal{L}c$, we have $\gamma_v \approx 1$. So

$$\tan\theta\tan\phi = 1,$$

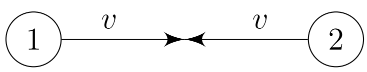
i.e. the outgoing trajectories are perpendicular in S.

Particle creation

Collide two particles of mass m fast enough, and you create an extra particle of mass M.

$$P_1 + P_2 = P_3 + P_4 + P_5,$$

where P_5 is the momentum of the new particle.
In the CM frame,



$$P_1 + P_2 = \begin{pmatrix} 2m\gamma_v c \\ \mathbf{0} \end{pmatrix}$$

We have

$$P_3 + P_4 + P_5 = \begin{pmatrix} (E_3 + E_4 + E_5)/c \\ \mathbf{0} \end{pmatrix}$$

So

$$2m\gamma_v c^2 = E_3 + E_4 + E_5 \geq 2mc^2 + Mc^2.$$

So in order to create this new particle, we must have

$$\gamma_v \geq 1 + \frac{M}{2m}.$$

Alternatively, it occurs only if the initial kinetic energy in the CM frame satisfies

$$2(\gamma_v - 1)mc^2 \geq Mc^2.$$

If we transform to a frame in which the initial speeds are u and 0 (i.e. stationary target), then

$$u = \frac{2v}{1 + v^2/c^2}$$

Then

$$\gamma_u = 2\gamma_v^2 - 1.$$

So we require

$$\gamma_u \geq 2\left(1 + \frac{M}{2m}\right)^2 - 1 = 1 + \frac{2M}{m} + \frac{M^2}{2m}.$$

This means that the initial kinetic energy in this frame must be

$$m(\gamma_u - 1)c^2 \geq \left(2 + \frac{M}{2m}\right) Mc^2,$$

which could be much larger than Mc^2, especially if $M \gg m$, which usually the case. For example, the mass of the Higgs' boson is 130 times the mass of the proton. So it would be much advantageous to collide two beams of protons head on, as opposed to hitting a fixed target.

Appendix A

Answers to selected exercises

(1) We choose a coordinate system where the origin is the center of the block, $+x$ is in the uphill direction, and $+y$ is perpendicular to the incline.

The forces we need to worry about are $\mathbf{F}$, $m\mathbf{g}$, and the normal force $\mathbf{N}$. Let's consider them together in the x direction. The object is at rest, so via Newton's second law, the net force is zero.

$$F\cos\theta = mgsin\theta$$

The normal force is perpendicular to the incline, and so the x component is zero. Meanwhile, the x component of the force F applied must balance out the gravitational force $mg\sin\theta$ in the x direction.

It's always a good idea to test the extremes and ensure the correct trig functions are used. If $\theta = 0$, we expect there to be no gravitational force at all in the x direction, and indeed $\sin(0) = 0$. Using the same argument, if $\theta = 90$, the gravitational force should be exclusively in the x direction, and again, it will be. As for $F\cos\theta$, the opposite is true, as it should be.

So far, so good. Next, let's look at the y direction. Newton's second law, again:

$$N = mg\cos\theta + F\sin\theta$$

We now have two equations, and two unknowns (F and N). Let's write the equations with the numbers substituted, and solve:

$$\begin{aligned} F\cos(38^\circ) &= (4\,\text{kg})(9.8\,\text{m/s}^2)\sin(38^\circ) \\ N &= (4\,\text{kg})(9.8\,\text{m/s}^2)\cos(38^\circ) + F\sin(38^\circ) \end{aligned}$$

The second equation alone gives us

$$N = (4\,\text{kg})(9.8\,\text{m/s}^2)\cos(38^\circ) + F\sin(38^\circ) \approx 30.89\,\text{N} + 0.6156F$$

And the first alone tells us

$$F = \frac{(4\,\mathrm{kg})(9.8\,\mathrm{m/s^2})\sin(38^\circ)}{\cos(38^\circ)} = 39.2\,\mathrm{N}\tan(38^\circ) = 30.63\,\mathrm{N}$$

So the answer for (b) is 30.63 N, while the first one is 49.75 N.

(2) First, let's identify the forces involved. There's the gravitational force mg downwards, the normal force N straight upwards, and the force from the rope, T, which will need some decomposition.

Choose a simple coordinate system where $+x$ is to the right, and $+y$ is straight upwards. The gravitational force is $-Mg$, purely in the y direction, and the acceleration is $a > 0$.

Writing Newton's second law for each of the two axes independently:

$$T_x = Ma$$
$$N + T_y = Mg$$

We know that $T_x = T\cos\theta$, so we can solve for T in terms of the acceleration and mass:

$$T = \frac{T_x}{\cos\theta} = \frac{Ma}{\cos\theta}$$

Next up, (b): $T_y = T\sin\theta$, so the third law equation becomes

$$N = Mg - T\sin\theta$$

If we substitute in the value for T, we find

$$N = Mg - \frac{Ma\sin\theta}{\cos\theta} = M(g - a\tan\theta)$$

... and we are done.

(3) Part a) As usual, let's start by looking at the forces involved. In the vertical direction, we have gravitational forces gm_1 and gm_2 acting on each of the blocks, respectively.
Block m_1 (or block 1) pushes downwards on block m_2 (or block 2) with that same force gm_1, and via Newton's third law, we find the reaction force (the normal force, in this case) from block 2 to block 1.

The total forces on block 1 are the gravitational force downwards, and the normal force upwards, from block 2 to 1. Net force: zero – as it must be, since it is at rest.

As for block 2, the downward forces are as mentioned above gm_1 from the upper block, and gm_2 from gravity on the block itself. This is cancelled out by a normal force from the ground on the block, of magnitude $g(m_1 + m_2)$. Again, the net force is zero, at it must be.

With the normal force on block 1, we know that the maximum frictional force that will oppose motion in mass m_1 is $\mu_s N = \mu_s g m_1$. As for block 2, there is no friction to the ground, so we need not worry about the maximum frictional force there.

If we write a second law equation for mass m_1 on its own, and one for the entire system, both exclusively in the x direction:

$$\begin{aligned} F - F_{Fmax} &= m_1 a \text{ (top block)} \\ F &= a(m_1 + m_2) \text{ (entire system)} \end{aligned}$$

The acceleration a as seen from an external reference frame is equal for both, since the condition is that they move together. We can solve the second equation for a, and stick it into the first, and then solve for F:

$$\begin{aligned} F - F_{Fmax} &= m_1 \frac{F}{m_1 + m_2} \\ F - F\frac{m_1}{m_1 + m_2} &= F_{Fmax} \\ F\left(1 - \frac{m_1}{m_1 + m_2}\right) &= F_{Fmax} \\ F = \frac{F_{Fmax}}{1 - \frac{m_1}{m_1+m_2}} = \frac{\mu_s g m_1}{1 - \frac{m_1}{m_1+m_2}} &\approx 9.03\,\text{N} \end{aligned}$$

Part b) We need to reverse the situation a bit. Except for the second law equations and such from above which clearly change, what else changes? The vertical forces don't; the maximum frictional force also doesn't, as it's based on the normal force, which is unchanged.
So, the force is now on m_2.

It seems like all we need to do is write a new pair of second law equations, again in the x direction only. One equation remains unchanged, the one for the entire system. However, F no longer acts on m_1!
Instead, it holds on via the frictional force, and can only accelerate together as long as that is "strong" enough.

If we push the lower block towards the right with too much force, what will happen? The upper block will glide "backwards", relative to the lower block. That means that the frictional force is now in the forward direction! Indeed, it's the only force acting on m_1 (horizontally), so we find

$$\begin{aligned} F_{Fmax} &= m_1 a \text{ (top block)} \\ F &= a(m_1 + m_2) \text{ (entire system)} \end{aligned}$$

Solving the first equation for a and substituting into the second:

$$\begin{aligned} F &= \frac{F_{Fmax}}{m_1}(m_1 + m_2) \\ &= \frac{\mu_s g m_1}{m_1}(m_1 + m_2) \\ &= \mu_s g(m_1 + m_2) \end{aligned}$$

That's the second and final answer!

(4) By symmetry the tension T in each segment is equal so since the system is in equilibrium the sum of the horizontal forces is zero:

$$2T\cos(-\alpha/2) = F$$

yielding

$$T = \frac{F}{2\cos(\alpha/2)} \approx 125.67\,\mathrm{N}$$

And we are done!

(5) see: https://www.youtube.com/watch?v=9NS0JcjNdp4

(6) Since m_1 is much greater than m_2, plus the fact that they only give us the ***kinetic*** friction coefficient, along with "and you may assume that block 1 never reaches the table", I think it's quite safe to assume the system will accelerate "counterclockwise", so that m_1 slides down towards the table.

If we draw up a free-body diagram, we find the following forces acting on block m_1, assuming a coordinate system where $+x$ is downhill and $+y$ is perpendicular to the surface (diagonally upwards to the left):

- $m_1 g\cos\theta$ acting in the $-y$ direction
- $N = m_1 g\cos\theta$ acting in the $+y$ direction, to cancel out the gravitational force
- $m_1 g\sin\theta$ acting in the $+x$ direction
- $F_f = \mu N = \mu m_1 g\cos\theta$ acting in the $-x$ direction
- T (unknown magnitude) acting in the $-x$ direction

As for the mass m_2, there are only two forces:

- $m_2 g$ acting downwards (which we call $-y$ in another coordinate system)
- T acting upwards, to counteract gravity (partially, not entirely)

In both cases, the net force must equal the object's mass times the acceleration, which will be the same for both due to the inextensible string that connects them. We can write two Newton's second law equations, and find

$$\begin{aligned} m_1 a &= m_1 g\sin\theta - T - \mu m_1 g\cos\theta \\ m_2 a &= T - m_2 g \end{aligned}$$

We can solve the second equation for T and substitute it into the first to find the acceleration:

$$\begin{aligned}
m_1 a &= m_1 g \sin\theta - (m_2 a + m_2 g) - \mu m_1 g \cos\theta \\
m_1 a + m_2 a &= m_1 g \sin\theta - m_2 g - \mu m_1 g \cos\theta \\
a(m_1 + m_2) &= m_1 g \sin\theta - m_2 g - \mu m_1 g \cos\theta \\
a &= \frac{m_1 g \sin\theta - m_2 g - \mu m_1 g \cos\theta}{m_1 + m_2} \approx 0.697\,\mathrm{m/s^2}
\end{aligned}$$

That answers part (a).

Part b)

We use the basic kinematics equation, with $x_0 = 0$ and $v_0 = 0$:

$$\frac{1}{2}at^2 = \frac{0.697\,\mathrm{m/s^2}}{2}(0.47\,\mathrm{s^2}) \approx 0.0769\,\mathrm{m} \approx 7.7\,\mathrm{cm}$$

(7) Since m_2 is much greater than m_1, m_2 will slide downhill and m_1 uphill... until they slide off each other, that is. The only other possibility is that $a = 0$ and that the system is in equilibrium, because the friction is great enough. I will assume the answer is not zero, though!

Drawing a free-body diagram (a must for most of these questions, but especially this one), we find a lot of forces.

As usual, I chose a coordinate system with x parallel to the incline, and y perpendicular. $+x$ is downhill, for no reason in particular.

On block m_1, there is friction, gravity/normal force (gravity in 2 dimensions) and tension. On block m_2, there is also gravity in two dimensions and a normal force, but we don't need to pay much attention to the y forces, since there is no friction on the ramp. We know that the normal force will cancel gravity, but that's about it for its usefulness. In addition to those, there's tension and a third law reaction force for the friction.

Let's try to write Newton's second law equations in the x direction. I will add up downhill forces, subtract uphill forces, and set it all equal to the mass times acceleration:

$$\begin{aligned}
m_1 g \sin\theta + \mu m_1 g \cos\theta - T &= -m_1 a \\
m_2 g \sin\theta - T - \mu m_1 g \cos\theta &= m_2 a
\end{aligned}$$

Not very pretty, is it? I will admit, it took me a few tries to get it right; I first forgot about the third law reaction force for the friction (there's a frictional force uphill on the second block!). As for directions, the first equation has $-m_1 a$ since the acceleration is positive downhill, but the motion will surely be uphill. The second equation has it without the minus sign, since that block will indeed move downhill.

Let's try to solve this by addition; that is, add the left sides to a new left side, and the two right sides to a new right side. The friction should cancel, so finding a should be less painful than by substitution.

$$\begin{aligned}
m_1 g \sin\theta + \mu m_1 g \cos\theta - T + m_2 g \sin\theta - T - \mu m_1 g \cos\theta &= -m_1 a + m_2 a \\
m_1 g \sin\theta - 2T + m_2 g \sin\theta &= -m_1 a + m_2 a \\
g \sin\theta (m_1 + m_2) - 2T &= a(m_2 - m_1) \\
a &= \frac{g \sin\theta (m_1 + m_2) - 2T}{m_2 - m_1}
\end{aligned}$$

Unfortunately, that doesn't quite get us all the way; we don't know T! Let's solve for it from, say, the second equation (either should work, and they're equally complex, so I just picked one). I suppose we'll do substitution after all:

$$\begin{aligned}
T &= m_2 g\sin\theta - \mu m_1 g\cos\theta - m_2 a \\
T &= g(m_2\sin\theta - \mu m_1\cos\theta) - m_2 a
\end{aligned}$$

Combining the two, we get... this monstrosity, which we need to solve for a again:

$$\begin{aligned}
a &= \frac{g\sin\theta(m_1+m_2) - 2(gm_2\sin\theta - g\mu m_1\cos\theta - m_2 a)}{m_2 - m_1} \\
a &= \frac{g\sin\theta(m_1+m_2) - 2gm_2\sin\theta + 2g\mu m_1\cos\theta + 2m_2 a}{m_2 - m_1} \\
a &= \frac{g\sin\theta(m_1+m_2) - 2gm_2\sin\theta + 2g\mu m_1\cos\theta}{m_2 - m_1} + \frac{2m_2 a}{m_2 - m_1} \\
a\left(1 - \frac{2m_2}{m_2 - m_1}\right) &= \frac{g\sin\theta(m_1+m_2) - 2gm_2\sin\theta + 2g\mu m_1\cos\theta}{m_2 - m_1} \\
a &= \frac{g\sin\theta(m_1+m_2) - 2gm_2\sin\theta + 2g\mu m_1\cos\theta}{m_2 - m_1} \cdot \frac{1}{1 - \frac{2m_2}{m_2 - m_1}}
\end{aligned}$$

Goodness, I could use Mathematica to simplify that, but it is accepted as correct!

For the sake of readability, here's a simplified version:

$$a = \frac{g(\sin\theta(m_2 - m_1)) - 2gm_1\mu\cos\theta}{m_1 + m_2}$$

(8) Okay, so let's see. The mass moves in a circle at constant speed: uniform circular motion. We don't know ω or T, though, as that's what we are looking for. We do know the angle and the rope's length, so we should be able to calculate the radius of the (horizontal) circle traced out by the mass itself, however.

In fact, if we forget about the third dimension, we have a very simple right triangle formed by the rope and the axes. We can see that

$$\begin{aligned}
\sin\beta &= \frac{r}{\ell} \\
r &= \ell\sin\beta
\end{aligned}$$

I will use cylindrical coordinates for this problem; that is, $\hat{r}$ is radially outwards, $\hat{\theta}$ is tangential to the traced out circle (positive counterclockwise, as the motion is), and $\hat{z}$ is upwards.

There is a centripetal acceleration

$$\begin{aligned}
a_c &= \omega^2(-\mathbf{r}) \\
&= \omega^2\ell\sin\beta(-\hat{r}) \\
&= \frac{4\pi^2}{T^2}\ell\sin\beta(-\hat{r})
\end{aligned}$$

towards the center of the traced circle, caused by a centripetal force m times the above.

What other forces are there? Well, there's certainly gravity, $-mg$ if we call upwards $+z$. There's the tension in the string, F_T (T is used for the period) which consists of z and r components. Let's decompose the tension.

$$F_{Tz} = F_T \cos\beta$$
$$F_{Tr} = F_T \sin\beta$$

The centripetal force is purely in the $-\hat{r}$ direction, so we don't need to decompose that. Neither do we need to decompose gravity, which is purely in the $-\hat{z}$ direction.

The net force on the mass must be the centripetal force, or there wouldn't be ***uniform*** circular motion. The z component of the tension must cancel out gravity, too, or the mass wouldn't move in a horizontal plane, as it does.

Time for Newton's second law. Let's just gather a list of the forces first, so there's no confusion while writing the equations. In the r axis, we have the centripetal force $F_r = a_c m$ inward, and the string tension also inward. In other words, the string tension ***provides*** (or ***is***, essentially) the centripetal force, and thus the cause of the centripetal acceleration.

In the z axis, there is gravity downwards, and a tension component upwards, which must cancel out to yield zero net force.

Lastly, in addition to $F_r = a_c m$, we can say that $a_c = \omega^2 r$, and we derived an expression involving the period earlier, so we find, for the r and z axes respectively,

$$a_c m = F_T \sin\beta \Rightarrow \frac{4\pi^2}{T^2}\ell m \sin(\beta) = F_T \sin\beta$$
$$mg = F_T \cos\beta$$

And we now at the point where we have two equations with two unknowns. I'll try to solve them manually. Solving the second for F_T is easy:

$$F_T = \frac{mg}{\cos\beta}$$

A-ha, nice! It's already in terms of g, m and β, so that's the finished answer for part (a)! Now, let's substitute that into the other one and solve for the period T, which was surprisingly easy:

$$\frac{4\pi^2}{T^2}\ell m \sin(\beta) = \frac{mg}{\cos\beta}\sin\beta$$
$$\frac{4\pi^2}{T^2}\ell m = \frac{mg}{\cos\beta}$$
$$\frac{2\pi\sqrt{\ell cos\beta}}{\sqrt{g}} = T$$

Tension cannot be negative, so we ignore the second solution.

(9) The block and a slab each have a gravitational force downwards, and a normal force upwards; I'll denote these by N_B for the normal force ***on*** the block (by the slab), and N_S for the normal force on the slab (by the table):

$$N_B = m_B g$$
$$N_S = g(m_B + m_S)$$

This then gives us the frictional forces F_{F1} (friction that limits the block's movement) and F_{F2} (friction that limits the slab's movement), named after the friction coefficients in the problem description:

$$F_{F1} = \mu_1 m_B g$$
$$F_{F2} = \mu_2 g(m_B + m_S)$$

What is the direction of these forces? Since the slab moves to the right relative to the table, the friction force there is to the left.
The block should also move right relative to the slab (how could the slab possibly accelerate ***faster***?), so that frictional force should also be to the left.

Do we now have all the forces? We have covered the y axis with gravitational forces and normal forces, and friction on all surfaces. Left are the third law reaction forces due to friction.

Because there is a frictional force F_{F1} by the slab (middle) on the block (top), there must be a force of equal magnitude in the opposite direction on the slab, so we have a rightwards force F_{F1} on the slab that we must not forget about.
There is also a leftwards frictional force on the slab from the table, so there is a reaction force there too, but since it's on the table, which we take to be immovable, we can ignore that force.

All in all we have, ignoring vertical forces, on the block: the external force F to the right, friction F_{F1} to the left.
On the slab, we have a reaction force F_{F1} to the ***right***, and "regular" friction with the table F_{F2} towards the left.

Let's also not forget that they don't accelerate together; the forces add up to some $m_B a_B$ and $m_S a_S$, but we can't solve for a combined a.

We can finally start writing second law equations, and substituting in the actual values. I will take $+x$ to be towards the right. First the block, then the slab:

$$F - F_{F1} = m_B a_B \Rightarrow F - \mu_1 m_B g = m_B a_B$$
$$F_{F1} - F_{F2} = m_S a_S \Rightarrow \mu_1 m_B g - \mu_2 g(m_B + m_S) = m_S a_S$$

Two equations, two unknowns (the accelerations), how unusual! However, they don't depend on each other at all, so this should be simple! Let's solve them one at a time:

$$F - \mu_1 m_B g = m_B a_B$$
$$a_B = \frac{F - \mu_1 m_B g}{m_B} \approx 12.733\,\mathrm{m/s^2}$$

$$\mu_1 m_B g - \mu_2 g(m_B + m_S) = m_S a_S$$
$$a_S = \frac{\mu_1 m_B g - \mu_2 g(m_B + m_S)}{m_S} \approx 8.829\,\mathrm{m/s^2}$$

Nice!

(10) Now that we've learned about the conservation of mechanical energy, this problem should be easier to solve than it would be with basic kinematics and friction equations. The work done by gravity should be very easy to find: the work done by gravity is the change in potential energy, which is mgh if we define h to be the height at which the block starts out, and $y = 0$ to be at the ground, so that $U = 0$ there.

We thus need to find h. The illustration makes it look a bit as if the block starts a bit down the ramp, but I assume it travels the distance L, or this would be hard to solve indeed! Via trigonometry, $\sin\theta = h/L$ so $h = L\sin\theta$. That gives us, for the work done by gravity,

$$W_g = mgL\sin\theta$$

... which answers part (b).
Next, we must find the work done by frictional forces as the block slides down. The magnitude of that force is

$$|F_f| = \mu_1 N = \mu_1 mg\cos\theta$$

We decompose the normal force, since gravity is straight downwards, while the block is on an incline.
Since the force is constant, and work is force times distance, we can find the work easily as $W_f = |F_f|L$. However, let's keep track of the signs here! The frictional force is always opposing the motion relative to the surfaces, so it is "backwards" (to the left) while the block only moves to the right. Therefore, the work is negative:

$$W_f = -(|F_f|L) = -\mu_1 mgL\cos\theta$$

... which answers part (a).
Next up is then the kinetic energy of the block as it has just reached the bottom (or end) of the incline. The kinetic energy started out at zero, and must now be at a maximum (since the potential energy is $U = 0$ at the bottom, by our definition). Without friction, it would be equal to the work gravity has done, but we must now add the work done by friction (subtract, in a way, since it is negative, but I prefer "add" to avoid confusion; subtracting a negative would give a larger value, which is clearly incorrect!).

$$\begin{aligned} K = W_f + W_g &= mgL\sin\theta - \mu_1 mgL\cos\theta \\ &= mgL(\sin\theta - \mu_1\cos\theta) \end{aligned}$$

The work-energy theorem at work... no pun intended.

Finally, part (d): how long does the block slide on the rough surface? It has a certain amount of kinetic energy, above; friction uses up a constant amount per unit length traveled, since it is constant at $\mu_2 N = \mu_2 mg$ (since the surfaces are now horizontal).

Using d for the distance traveled, the work done by friction is then $\mu_2 mgd$ ($W = Fd$). That work equals the initial kinetic energy, so we set them equal and solve for d:

$$\begin{aligned}\mu_2 mgd &= mgL(\sin\theta - \mu_1\cos\theta)\\ d &= \frac{L(\sin\theta - \mu_1\cos\theta)}{\mu_2}\end{aligned}$$

That's all!

(11) Since the spring is ideal, Hooke's law holds, and we can use the equations we found in lecture, by solving a differential equation for this simple harmonic oscillator. The equation we found was

$$x(t) = A\cos(\omega t + \varphi)$$

where A is the amplitude in meters, ω the angular frequency in radians/second, and φ the phase angle in radians. A and φ are found from the initial conditions, while ω can be found as

$$\omega = \sqrt{\frac{k}{m}}$$

The period of oscillation is

$$T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{m}{k}} = 2\pi\sqrt{\frac{7}{36}} \approx 2.77\,\mathrm{s}$$

To find the position as a function of time, we need to find the amplitude and the phase, by using the initial conditions. At $t = 0$, $x(0) = x_0 = 1.8$ meters, as given in the problem. We substitute those values into the $x(t)$ equation:

$$x_0 = A\cos(\varphi)$$

That only gets us so far, since there are two unknowns, A and φ. We can find a second equation in taking the time derivative of $x(t)$ to find $v(t)$, though, since we know the initial velocity.

$$v(t) = \frac{dx(t)}{dt} = -A\omega\sin(\omega t + \varphi)$$

At $t = 0$, this should be equal to -3 (if x_0 is positive, then $+\hat{x}$ is towards the right, but v_0 is towards the left). Combined with the equation for $x(t)$, we have these two equations:

$$\begin{aligned} x_0 &= A\cos(\varphi) \\ -v_0 &= -A\omega\sin(\varphi) \end{aligned}$$

$$\begin{aligned} -\frac{x_0}{v_0} &= -\frac{\cos(\varphi)}{\omega\sin(\varphi)} \\ \omega\frac{x_0}{v_0} &= \frac{1}{\tan\varphi} \\ \arctan\frac{v_0}{\omega x_0} &= \varphi \approx 0.6338 \text{ rad} \approx 36.31^\circ \end{aligned}$$

Solving for A should now be dead simple, using the equation $x_0 = A\cos(\varphi)$:

$$\begin{aligned} 1.8 &= 0.80578A \\ A &= 2.23\,\text{m} \end{aligned}$$

ω, using the formula above, is about 2.2678 rad/s, so all in all, the formula for $x(t)$ is

$$x(t) = 2.23\cos(2.2678t + 0.6338)$$

Evaluated at $t = 0$, this equals 1.7969 m, and the problem states $x_0 = 1.8\,\text{m}$ – close enough; it's clearly due to rounding errors.

"(c) How long will it take for the mass to first return to the equilibrium position?"

That happens when $x(t) = 0$, so we set it up and solve for t:

$$\begin{aligned} 2.33\cos(2.2678t + 0.6338) &= 0 \\ 2.2678t + 0.6338 &= \frac{\pi}{2} \text{ (by taking the arccosine of both sides)} \\ t &= \frac{\pi/2 - 0.6338}{2.2678} \approx 0.413\,\text{s} \end{aligned}$$

"(d) How long will it take for the spring to first become completely extended?"

I assume that by "completely extended", they mean when it is as long as it will ever become – since it is at its natural length at $x = 0$, which is what we found above. Since the initial velocity is in the "extending direction", this should happen the first time $v = 0$, so let's set the derivative, which we found earlier, equal to zero:

$$\begin{aligned} -A\omega\sin(\omega t + \varphi) &= 0 \\ -2.33 \cdot 2.2678\sin(2.2678t + 0.6338) &= 0 \\ \sin(2.2678t + 0.6338) &= 0 \\ 2.2678t + 0.6338 &= \pi \text{ (by taking the arcsine of both sides)} \\ t &= \frac{\pi - 0.6338}{2.2678} \approx 1.106\,\text{s} \end{aligned}$$

I chose π instead of 0 for the arcsine because choosing 0 yields a negative time, which is clearly incorrect.
Honestly, I'm not completely happy with this solution, but it worked, at least.

(12) This problem can be conceptualized similarly to problem 2, i.e. conservation of energy. The block has an initial kinetic energy of $K = \frac{1}{2}mv_i^2 = 32$ joule; by definition, that kinetic energy must go down to 0 when $v = 0$, which is of course when it first comes to a halt. Part of the kinetic energy will be eaten up by friction (turned into heat, mostly), and part will be transferred into the spring and stored there as potential energy.

The kinetic friction force is $\mu N = \mu mg$, which is constant regardless of position or velocity; the direction is opposite the motion, so to the left here, $-\hat{x}$. The spring's force is $-kx\ \hat{x}$, also to the left.

The work done by the forces together equals the sum of the forces times the distance x the block travels; this work then equals the initial kinetic energy of the block. After having set the two equal, we can solve for x, which is how far the spring has compressed (and how far the block has traveled, after the "collision" with the spring). We can either set the sum of them equal to zero, or set the two work quantities equal, which is the same thing. I chose the latter:

$$\frac{1}{2}mv_i^2 = x\left(\mu mg + kx\right)$$

Ah, but here's a snag: kx, the force from the spring, is not constant! It is 0 at the start, kx only at the end of the motion, and somewhere in between for the rest of the time. ***However***, it is linear, which is good news for us! That means we can find the average force simply as $\frac{kx}{2}$, and keep going, with no calculus:

$$\begin{aligned}
\frac{1}{2}mv_i^2 &= x\left(\mu mg + 0.5kx\right)\\
\frac{1}{2}mv_i^2 &= x\mu mg + 0.5kx^2\\
\frac{mv_i^2}{k} &= 2x\frac{\mu mg}{k} + x^2\\
x^2 + \frac{2\mu mg}{k}x - \frac{mv_i^2}{k} &= 0
\end{aligned}$$

Using the quadratic formula, $x = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$:

$$x = -\frac{\mu mg}{k} \pm \frac{\sqrt{\left(\frac{2\mu mg}{k}\right)^2 + \frac{4mv_i^2}{k}}}{2}$$

If we stick some values into that mess, we find

$$x = -2.075 \pm 2.88195$$

Since the answer is clearly positive as defined in the problem, it must be $x = -2.075 + 2.88195 = 0.80695 \approx 0.807\,\text{m}$.

(13) Okay, let's see. There is no friction, so we should be able to rely on conservation of energy to find the initial velocity from the spring's compression. It is compressed a distance d, with a spring constant k. Now, unfortunately, I don't know how to calculate the stored potential energy in a spring; it's a common formula, easy to find – but I would prefer to figure it out myself! Looking up a formula doesn't teach you much, but deriving it yourself can be very helpful indeed, especially if you've never seen it before.

So, let's take a sidestep for a moment.

Spring forces are conservative, so the amount of work done in compressing a spring should equal the amount of potential energy stored in it. We need to exert a force $F(t) = kx(t)$ to compress a spring, where $x(t)$ is the amount we have compressed it so far. The total work done, and the total energy stored, must therefore be the integral of this:

$$U_{spring} = \int_0^d F(t)\,dx = \int_0^d kx\,dx = k\left[\frac{x^2}{2}\right]_0^d = k\frac{d^2}{2} = \frac{1}{2}kd^2$$

Neat! It looks a lot like the equation for kinetic energy (and many other equations in physics, for that matter).

Now that we know how much energy is stored in the spring when the bead comes to a temporary halt, before being "shot out" again, we can find v_0, in case we need it later. The energy stored in the spring must come exclusively from the bead's kinetic energy (some of which come from gravity). If we define gravitational potential energy as 0 at the bottom, then it must be be $2mgR$ at the top of the loop.

The spring starts out with no stored energy, while the bead starts out with its kinetic energy $K_E = \frac{1}{2}mv_0^2$ and its gravitational potential energy $2mgR$. Since there is no friction or other resistive forces, the sum of all these must be conserved.

The speed at point A can be found by finding the bead's kinetic energy at that point, which is the sum of its initial kinetic energy and potential energy, minus the energy used up working against gravity, mgR, to reach point A:

$$\begin{aligned}
\frac{1}{2}mv_A^2 &= \frac{1}{2}mv_0^2 + 2mgR - mgR \\
mv_A^2 &= mv_0^2 + 2mgR \\
v_A &= \sqrt{v_0^2 + 2gR}
\end{aligned}$$

We can find v_0. When the bead has compressed the spring fully, all of the initial kinetic energy plus all of the gravitational potential energy is now stored in the spring, so we can equate them:

$$\begin{aligned}
\frac{1}{2}mv_0^2 + 2mgR &= \frac{1}{2}kd^2 \\
mv_0^2 + 4mgR &= kd^2 \\
v_0^2 &= \frac{kd^2 - 4mgR}{m} \\
v_0 &= \sqrt{\frac{kd^2 - 4mgR}{m}}
\end{aligned}$$

v_0^2 is what we need to find v_A, however:

$$v_A = \sqrt{\frac{kd^2 - 4mgR}{m} + 2gR}$$

Almost there! Now that we know the speed at A, we can apply the formula for centripetal acceleration, $|a_c| = \frac{v^2}{r}$, and then multiply by the mass m to find the centripetal force.

The normal force from the track is the only possible source for this centripetal force, which is necessary for the bead to move along the (semi)circular track. Therefore, we find the centripetal force:

$$N = m\frac{v_A^2}{R} = \frac{m}{R}\left(\frac{kd^2 - 4mgR}{m} + 2gR\right)$$

and that solves the problem!

(14) Well... Unless I'm missing something, I remember the answer from lecture! I'll still try to re-derive it, though, to make sure I fully understand the problem. If I do, this shouldn't take long.

Okay, so the track is frictionless, and we can use conservation of energy to simplify things. Since the object is released from rest, its initial potential energy is mgh, assuming $U = 0$ at $y = 0$; since that is my choice to make, I decide it shall be so.

When entering the loop, the potential energy is zero, and the object's speed is at a maximum, as is the kinetic energy. It then travels up $2R$ against gravity, which causes it to lose kinetic energy again.

Let's first find the condition for the object not falling down at the middle of the loop. $|a_c| > g$ must be the case, or the object will not move in a circle. This puts a constraint on v_{top}, the speed at the top:

$$a_{c,top} = \frac{v_{top}^2}{R} \geq g$$

Next, we need to figure out what v_{top} is, as a function of the initial height h. At that height, it will have a potential energy of $mg2R$, which is smaller than the mgh it begins with (or it will never reach that point).

$$\begin{aligned}\frac{1}{2}mv_{top}^2 &= mgh - 2mgR \\ v_{top}^2 &= 2gh - 4gR \\ v_{top} &= \sqrt{2g(h - 2R)}\end{aligned}$$

Now we just need to put the two together, and solve for h.

$$\begin{aligned}\frac{2gh - 4gR}{R} &\geq g \\ 2gh &\geq 5Rg \\ h &\geq \frac{5}{2}R\end{aligned}$$

Since the question is when it "just barely" loses contact, the answer is $h = \frac{5}{2}R$.

(15) I'll start off by finding d_1, not only because it's the first question, but because it should be independent of everything else.
For this problem, I choose a coordinate system of one axis, y, which is positive downwards, and has its origin at the spring's natural length. In other words, $y = +d_1$ when the system is at equilibrium with the mass.

Since it is in equilibrium, with the spring force upwards, and gravity downwards, with no acceleration:

$$\begin{aligned}d_1 k &= mg \\ d_1 &= \frac{mg}{k} \approx 0.19818 \text{ m}\end{aligned}$$

Now then, onto the rest of the problem. I will use the same coordinate system, by the way.

In a horizontal oscillator (as in lecture), there is only one horizontal force, which is that of the spring. I know (from a quick and dirty test) that the period is the same for this vertical oscillator, but how can we show that to be the case, now that gravity is present along the oscillating axis? If this were an exam question, I would ***not*** have wasted a try on that assumption!

We can actually show that this system is equivalent to the horizontal one.
We've just shown that the "new" equilibrium position is at $y = d_1$. However, we can re-define y instead, so that $y = 0$ at that point. Why? Because the block will oscillate around that point, moving equal amounts up as down from the new zero point, which is not the case for the old one. In other words, we will get a symmetrical problem if we change the zero point, so we do just that.

The spring force is upwards, in magnitude $k(d_1 + y)$ in this case, now. At $y = 0$, it should be kd_1, and for greater values of y (further down), it should be greater, so that looks about right. Gravity is mg, always downwards. Putting this all together, $a = \ddot{y}$ being positive downwards, we set $m\ddot{y}$ equal to the net force, adding the downwards force (gravity) and subtracting the upwards force (spring force):

$$m\ddot{y} = mg - k(d_1 + y)$$

However, note that since $d_1 = \frac{mg}{k}$, $mg = kd_1$, we can replace mg by kd_1:

$$\begin{aligned}m\ddot{y} &= kd_1 - k(d_1 + y) \\ m\ddot{y} &= -ky \\ \ddot{y} &= -\frac{k}{m}y\end{aligned}$$

$$\ddot{y} + \frac{k}{m}y = 0$$

A-ha! This is clearly the exact same differential equation we had earlier in lecture, only we call the axis y instead of x, so we can safely use the same solutions! That it,

$$\begin{aligned}
\omega &= \sqrt{\frac{k}{m}} \\
T &= 2\pi\sqrt{\frac{m}{k}} \\
y &= A\cos(\omega t + \varphi) \\
\dot{y} &= -A\omega\sin(\omega t + \varphi) \\
f &= \frac{\omega}{2\pi}
\end{aligned}$$

We have already solved (a), so let's calculate the frequency for part (b). Using the above formulas, we find $\omega \approx 7.03562$ rad/s, so $f \approx 1.12\,\text{Hz}$.

Next, the spring's length when the block reaches its highest point. The amplitude of the oscillation is d_2, the amount we extended it from the (new, with the mass) equilibrium point, so the answer is the spring's original length plus d_1, which is the new equilibrium point, minus the amplitude d_2. All in all, $\ell_{top} = \ell + d_1 - d_2 \approx 1.098\,\text{m}$.

Finally, the maximum speed of the block. The velocity is given by $\dot{y}(t)$ above, which is clearly maximized when the sine function is 1. We don't care when that happens, only that the speed at that point is the magnitude of the function's value when the sine term is 1, i.e. $A\omega = d_2\omega \approx 2.814\,\text{m/s}$, and that's it for this question!

(16) The day's length is $0.7 \cdot 24$ hours $= 16.8$ hours, or 60480 seconds. This must then be the orbital period of the satellite, since it is supposed to remain over the same point at all times.

I don't recall the exact formulas we learned from lecture (and if I did, I likely wouldn't a year from now), but I do remember that the total mechanical energy is exactly $\frac{1}{2}U$. The mechanical energy is then the sum of the current kinetic energy, and the gravitational potential energy:

$$\begin{aligned}
\frac{1}{2}\left(-\frac{Gm_p m_s}{r}\right) &= \frac{1}{2}m_s v_{orb}^2 - \frac{Gm_p m_s}{r} \\
\frac{Gm_p m_s}{r} &= m_s v_{orb}^2 \\
\frac{1}{r} &= \frac{v_{orb}^2}{Gm_p} \\
r &= \frac{Gm_p}{v_{orb}^2}
\end{aligned}$$

We can then write v_{orb}, the tangential velocity of the satellite, in terms of r and T:

$$\begin{aligned}
v_{orb} &= \frac{2\pi r}{T} \\
v_{orb}^2 &= \frac{4\pi^2 r^2}{T^2}
\end{aligned}$$

Substitute into r (by multiplying by the reciprocal, instead of having a 3-layer fraction):

$$
\begin{aligned}
r &= Gm_p \cdot \frac{T^2}{4\pi^2 r^2} \\
r^3 &= \frac{Gm_p T^2}{4\pi^2} \\
r &= \left(\frac{Gm_p T^2}{4\pi^2}\right)^{1/3}
\end{aligned}
$$

Next, part (b): what is the escape velocity at this distance r from the planet?
I could re-derive the expression for the escape velocity as well, which wasn't that hard, but I recall that $v_{esc} = \sqrt{2} \times v_{orb}$, and we already have an expression for v_{orb}. Multiplying v_{orb} by $\sqrt{2}$ and then simplifying:

$$
\begin{aligned}
v_{orb} &= \frac{2\pi}{T}\left(\frac{Gm_p T^2}{4\pi^2}\right)^{1/3} \\
v_{esc} &= \sqrt{2}\left(\frac{2\pi G m_p}{T}\right)^{1/3}
\end{aligned}
$$

However, they want the answer in ***km/sec***, so we need to divide that by 1000.

(17) Hitting the water at, say, 0.1 m/s will surely not be lethal, but I assume the condition is that she doesn't touch the water whatsoever, or we can't find an exact answer to the question.

I will use a coordinate system where y increases downwards, and is centered on the bridge; thus the water is at $y = h$.
Also, I will use conservation of energy to solve this problem. My first solution was to find the total energy at $y = L$, after a period of free fall, and then the total energy at $y = h$, solving for L that way. I realized later, reading the forums, that this is unnecessarily complex, so my much simpler solution is below.

The kinetic energy is zero both just as you jump (since it is done with zero speed) and as you almost reach the water: the velocity vector reverses at that point, so $v = 0$ at the lowest point (which is $y = h$).

The change in gravitational potential energy is mgh, and all of that goes into the spring. (That's the only possibility other than kinetic energy, which we already ruled out).

The energy stored in the spring is given by $\frac{1}{2}kx^2$, where x in this case is $h - L$, the distance the cord is stretched beyond its natural length of L. (It is the distance to the water, from the natural length.)

We set the two equal, and solve for L:

$$
\begin{aligned}
mgh &= \frac{1}{2}k(h-L)^2 \\
2mgh &= k(h^2 - 2hL + L^2) \\
0 &= h^2 - 2hL + L^2 - \frac{2mgh}{k} \\
0 &= L^2 - (2h)L - \left(\frac{2mgh}{k} - h^2\right)
\end{aligned}
$$

We use the quadratic formula:

$$\begin{aligned}
L &= h \pm \frac{1}{2}\sqrt{(-2h)^2 + 4\left(\frac{2mgh}{k} - h^2\right)} \\
L &= h - \frac{1}{2}\sqrt{\frac{8mgh}{k}} \\
L &= h - \sqrt{\frac{2mgh}{k}} \approx 61.9942\,\text{m}
\end{aligned}$$

The plus-solution gives $L > h$, so that is clearly not the solution we want, so I got rid of that one between steps 1 and 2.

Next, part (b).

Same as last week: the spring's natural length is L, but at equilibrium, it is stretched a bit further due to the downwards force mg balancing out with the upwards force kx (where x how far it has stretched beyond its natural length L). We simply set them equal:

$$\begin{aligned}
kx &= mg \\
x &= \frac{mg}{k}
\end{aligned}$$

So the equilibrium point is at $L + \frac{mg}{k} \approx 69.22\,\text{m}$. The distance left down to the water is then $h - 69.22\,\text{m} \approx 30.78\,\text{m}$.

Full disclosure: my initial solution, which ***was*** marked as correct, was actually invalid. The reason I tried the energy approach later despite the green checkmark was because the equation I got was way too complex for it to make sense – but that was due to a bit of a miss on my side: I used both g and the value 10 instead of g, and tried to simplify... 10 and g didn't cancel, of course, so it turned out very complex... until I realized, used g everywhere, and it was only slightly more complex than the answer above.

Anyway, my process there was to treat it as a spring oscillator, like last week's problem 7. The problem with that is, I realized, that this cord only acts as a spring when ***stretched***, not otherwise. I'm not 100% sure why that affects the answer even when we ***only*** consider the way down, but the answer was about 0.7 meters greater. (Close enough to be considered correct!)

The larger the mass is, the further apart the two solutions become. The symbolic solution I got there was

$$L = h - \frac{\sqrt{2gmhk - g^2m^2}}{k} \text{ (invalid!)}$$

(18) Alright, let's start by identifying the forces on the bead. Gravity and spring forces are quite obvious, but is there anything else? Yes, there is: a normal force by the hoop itself – which they ask for in part (b).

The centripetal force required for this motion is still $m\frac{v^2}{R}$ at all times, but v is not a constant in this problem (since both gravity and the spring will change the bead's speed), so the centripetal force will vary, too.

Since there is no friction, and gravitational forces and spring forces are both conservative, let's try conservation of energy.

The initial energy is all either gravitational potential energy or and spring potential energy. Let's set $U_g = 0$ at the center of the circle; in that case, the initial gravitational potential energy is mgR, and the final, at $\theta = 90°$ is 0 by our definition.
There is no initial kinetic energy, since the initial speed was negligible.
What about the spring? It is stretched a distance R beyond its natural length (total length $2R$, natural length R) so it stores a potential energy $U_s = \frac{1}{2}kR^2$ at the top.

$$E = mgR + \frac{1}{2}kR^2$$

At $\theta = 90°$, all gravitational potential energy, and part of the spring's, will have turned into kinetic energy in the bead.

Here, the kinetic energy is $\frac{1}{2}mv^2$. The spring's stored energy is related to how far it is stretched beyond R; how far is that, at this point?
If we draw this up, with a θ as a right angle, and we draw a triangle with the spring length as the hypotenuse, the left and top sides of the triangle are both R in length, so the hypotenuse (the spring's current length) is $x = \sqrt{2R^2} = \sqrt{2} \times R$. It is then stretched $d = R\sqrt{2} - R = R(\sqrt{2} - 1)$ beyond its natural length. That gives it a potential energy of $U_s = \frac{1}{2}kR^2(2 - 2\sqrt{2} + 1)$.

Adding it all up, and setting it equal to E above, which is the total energy at all times:

$$\begin{aligned}
\frac{1}{2}mv^2 + \frac{1}{2}kR^2(2 - 2\sqrt{2} + 1) &= mgR + \frac{1}{2}kR^2 \\
mv^2 &= 2mgR + kR^2 - kR^2(2 - 2\sqrt{2} + 1) \\
mv^2 &= 2mgR + kR^2(1 - (2 - 2\sqrt{2} + 1)) \\
mv^2 &= 2mgR + kR^2(-2 + 2\sqrt{2}) \\
v &= \sqrt{\frac{2mgR + kR^2(-2 + 2\sqrt{2})}{m}}
\end{aligned}$$

Next, we need to find the magnitude of the normal force from the hoop on the bead.

The ***radial*** force (inwards) must always add up to the centripetal force, so we can decompose the forces and set that equal to $\frac{mv^2}{R}$.

Gravity at $\theta = 90°$ is clearly purely tangential; there's no left-or-right force due to gravity. In other words, we can ignore gravity for this part.

The spring force, on the other hand, clearly has components both tangential (up/down) and radial (left/right) at this point.

The total spring force is proportional to its extension past R (its natural length), which we found earlier, so

$$F_{spr} = k(\sqrt{2R^2} - R) = k\sqrt{2}R - kR = kR(\sqrt{2} - 1)$$

The above is the total spring force; we only want the radial component, which is $1/\sqrt{2}$ times that, or $F_{spr,rad} = kR(1 - 1/\sqrt{2})$.

The normal force is then the centripetal force $\frac{mv^2}{R}$, minus the force in that direction that the spring provides. (That is, the hoop must provide all the necessary force that the spring isn't.)

$$N + kR(1 - \frac{1}{\sqrt{2}}) = \frac{mv^2}{R}$$
$$N + kR(1 - \frac{1}{\sqrt{2}}) = 2mg + kR(2\sqrt{2} - 2)$$

$$N = 2mg + kR\left(2\sqrt{2} - 2 - 1 + \frac{1}{\sqrt{2}}\right)$$
$$N = 2mg + kR\left(\frac{5}{\sqrt{2}} - 3\right)$$

That's it!

(19) The moon is about 300 times more massive than the planet; I will assume that makes it valid to use the formulas we've already used (that are not valid if the masses are close to each other; more on that and center on mass very soon – in the next problem).

As with the previous problem regarding orbit, I will use $E = \frac{1}{2}U$ here – it's easy to remember, so why not?

$$K_E + U = \frac{1}{2}U$$
$$K_E + \frac{1}{2}U = 0$$
$$\frac{1}{2}m_m v_{orbit}^2 - \frac{1}{2}\frac{Gm_p m_m}{R} = 0$$
$$v_{orbit}^2 - \frac{Gm_p}{R} = 0$$
$$v_{orbit} = \sqrt{\frac{Gm_p}{R}}$$

The period is then simply the distance divided by the velocity:

$$T = \frac{2\pi R}{v_{orbit}} = 2\pi R\sqrt{\frac{R}{Gm_p}}$$
$$T = 2\pi\sqrt{\frac{R^3}{Gm_p}} = 2\pi\frac{R^{3/2}}{\sqrt{Gm_p}} \approx 637\,374\,\mathrm{s}$$

Finally, we just need to divide this by one "Earth day" of 86400 seconds, so the answer is $637374/86400 \approx 7.38$ days.

(20) Ah, a possibly-scary problem. The concept of ***center of mass*** should make it easy, though, especially since the period is the same for both stars.

The center of mass of a system is a point around which ***both*** stars orbit. (In our solar system, the center of mass is inside the Sun, since it's such a dominant mass, but it's not at the Sun's center – so the Sun actually makes a tiny orbit around the center of mass).

Apparently, in the case of two bodies, $m_1 r_1 = m_2 r_2$ will hold. Combined with $s = r_1 + r_2$ where s is a given, we already have two equations and two unknowns. Too easy.

We can solve the second equation to give $r_1 = s - r_2$ and substitute into the first, to give one equation with one unknown:

$$\begin{aligned} m_1(s - r_2) &= m_2 r_2 \\ m_1 r_2 + m_2 r_2 &= m_1 s \\ r_2(m_1 + m_2) &= m_1 s \end{aligned}$$

$$r_2 = \frac{m_1 s}{m_1 + m_2}$$

We can then find r_2 easily, and $r_1 = s - r_2$ as mentioned, so that too is easy. For the given values,

$$\begin{aligned} r_1 &= 1.411 \times 10^{18}\,\mathrm{m} \\ r_2 &= 1.909 \times 10^{18}\,\mathrm{m} \end{aligned}$$

Now, we just need to find the period. If the bodies orbit as shown, the gravitational attraction between them is always towards the center of mass. We can find ω this way, by equating the centripetal force $m|a_c| = m\omega^2 r$ with the gravitational force on one of the masses:

$$\begin{aligned} m_1 \omega^2 r_1 &= \frac{G m_1 m_2}{s^2} \\ \omega^2 &= \frac{G m_1 m_2}{m_1 r_1 s^2} \\ \omega &= \sqrt{\frac{G m_2}{r_1 s^2}} \end{aligned}$$

Finally, $T = \dfrac{2\pi}{\omega}$:

$$T = 2\pi \sqrt{\frac{r_1 s^2}{G m_2}} \approx 7.505 \times 10^{17}\,\mathrm{s} \approx 23.8\,\mathrm{billion years}$$

This is, incredibly enough, correct. The staff admitted in a forum post that the value for the distances was way, way larger than what is realistic (by 6 orders of magnitude), and so the period grew to about 10^9 times larger than expected!

(21) Well, (b) is easy from the graph – it is at $x = 1$. But let's avoid getting ahead of ourselves.

The important thing to remember here is that $\dfrac{dU}{dx} = -F_x$. So far part (a), we need to find the derivative of $U(x)$, and then remember to negate the answer. Using the chain rule,

$$\begin{aligned}
\frac{dU}{dx} &= 4(x^2-1)(2x) = 4(2x^3-2x) = 8x^3-8x \\
F_x = -\frac{dU}{dx} &= 8x - 8x^3 = 8(x-x^3)
\end{aligned}$$

For a more rigorous solution of part (b), we can find where $F_x = 0$, and only look at the cases where $x > 0$, which is the condition given:

$$\begin{aligned}
8(x-x^3) &= 0 \\
x^3 &= x \\
x^2 &= 1 \\
x &= \pm\sqrt{1}
\end{aligned}$$

For $x > 0$, the only solution is $x = 1$. As a last step, we can confirm whether this is a stable equilibrium point, or an unstable one. It's clear from the graph that it's stable (if there is a small amount of force on the body, it will tend to roll back down from the "hills", rather than roll away, as it would from one of the peaks).

Mathematically, the condition here is that the second derivative of U is positive; that makes the curve "concave upward", i.e. looks like a U shape, so that things tend to stay inside. If $\frac{d^2U}{dx^2} < 0$, the opposite is true, and we are at a peak.

We calculate the second derivative, and stick $x = 1$ in there:

$$\begin{aligned}
24x^2 - 8 &\stackrel{?}{>} 0 \\
16 &> 0
\end{aligned}$$

The second derivative is positive, and so this is indeed a ***stable*** equilibrium point. If we try this at $x = 0$, we find -8, less than zero, and indeed, that is an unstable equilibrium point according to the graph.

Next, part (c), which asked

"c) Suppose the body starts with zero speed at $x = 1.5$ m. What is its speed (in m/s) at $x = 0$ m and at $x = -1$ m?"

Okay, so what does this imply? It starts at rest (zero kinetic energy), and we can easily calculate $U(1.5)$. We can then easily calculate $U(0)$, subtract the two, and we know the change in kinetic energy, and can solve for v.

$$\begin{aligned}
K_E(0) &= U(1.5) - U(0) \\
\frac{1}{2}mv^2 &= 2(1.5^2-1)^2 - 2(-1)^2 \\
\frac{1}{2}mv^2 &= 3.125 - 2 = 1.125 \\
v &= \sqrt{2.25} = 1.5\,\text{m/s}
\end{aligned}$$

For $x = -1$, we simply do the same thing, but use $U(-1)$ instead of $U(0)$.

$$\begin{aligned} K_E(0) &= U(1.5) - U(-1) \\ \frac{1}{2}mv^2 &= 3.125 \\ v &= \sqrt{6.25} = 2.5\,\mathrm{m/s} \end{aligned}$$

(22) Ah, I actually solved this on the forum last week or so, using Gauss's law. I will try to do it this way instead, here, though.

At the surface,

$$F_g = mg = \frac{Gmm_e}{r_e^2}$$

The fraction of Earth's mass inside this smaller radius r is just the ratio of the volume of r to the volume of r_e. I will call this mass m_r, so

$$\begin{aligned} m_r &= \frac{4/3\pi r^3}{4/3\pi r_e^3} m_e \\ m_r &= \frac{r^3}{r_e^3} m_e \end{aligned}$$

$$\begin{aligned} F_i &= \frac{Gm}{r^2}\frac{r^3}{r_e^3} m_e \\ F_i &= \frac{Gmr}{r_e^3} m_e \end{aligned}$$

Almost there... we need to get rid of that G, and write it in terms of g instead. We have an equation for $F_g = mg$ in terms of G and so on above, so we solve that one for G, and substitute in in here:

$$G = \frac{gr_e^2}{m_e}$$

$$\begin{aligned} F_i &= \frac{mr}{r_e^3} m_e \left(\frac{gr_e^2}{m_e}\right) \\ F_i &= \frac{mgr}{r_e} \end{aligned}$$

Got it!

Next, part (b): "How long would it take for this object to reach the other side of the earth? Express your answer in terms of the gravitational constant at the surface of earth g, m, and r_e as needed."

Okay, so the force experienced by the mass, at all times, is the force shown above. We can find the acceleration simply by dividing out m. If the acceleration were constant, we could use a simple kinematics equation here... but it's not constant. The velocity will not be constant, either, so we can't simply find a value for the velocity and calculate the time from knowing distance and velocity.

However...! The force is in the form $F = kr$, where $k = \frac{mg}{r_e}$ is a ***constant***, in newtons per meter. In other words, this looks like a spring problem, in a way. Not exactly, perhaps, but close enough: consider a spring of near-zero natural length, attached at the center of the Earth. It will always have an inwards force, which is proportional to r, the distance you've stretched it beyond its original zero length.

Once you've passed the center, it will still be an inwards force, that is now trying to make you stop and reverse. One full oscillation of this system will then bring you all the way to the other side, and then back, in a symmetric motion. Therefore, the answer is half the period.

$$T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{m}{k}} = 2\pi\sqrt{\frac{m}{mg/r_e}} = 2\pi\sqrt{\frac{r_e}{g}}$$

Half this is then simply

$$\frac{T}{2} = \pi\sqrt{\frac{r_e}{g}}$$

(23) All right, let's see. The spring is compressed, so as we start this experiment, block 2 will accelerate towards the right. The blocks are "non-identical", so we can't say anything qualitative about the center of mass, other than that it must be somewhere between the blocks (possibly part-way inside one of them).

This is an easy problem, IF you approach it correctly. If you don't, it's very easy to get it wrong. The approach that is way easier than the others is to consider conservation of momentum. In the beginning of the problem, there is a net external force on the system – the normal force from the wall pushing towards the right. Net force means acceleration, so to begin with, there is an ***acceleration towards the right***, while $x_2 < d$ (the spring is compressed), so option (d) is correct.

When block 2 passes $x_2 > d$, the spring starts to pull together, which moves block 1 towards the right. When it moves away from the wall, there is ***no longer a net external force*** in the horizontal direction, and we can (and should) apply the conservation of momentum to consider what may happen next. No matter what the masses of the two blocks are, momentum must be conserved!

The net momentum of the system is $p_{tot} = m_{tot}v_{cm}$. The mass is not changing, and p_{tot} must be held constant and so v_{cm} ***is a constant*** after this; option (b) is also correct. All options except (b) and (d) are thus incorrect.

This was demonstrated in lecture, with an extremely similar system, of two air track-carts and a spring. After the system had been set in motion, the center of mass held a constant velocity, despite the oscillating behavior of the two masses. That is exactly what will happen here.

Since the center of mass will hold a constant velocity towards the right, the system will keep moving towards the right until it hits an obstacle (given that we ignore friction).

(24) This a bit nonintuitive, based an experience – but it's important to note the force F is the same in all cases.

We have found previously that the momentum of a system can be found as $m\mathbf{v_{cm}}$, where m is the total mass:

$$\mathbf{p_{tot}} = m_{tot}\mathbf{v_{cm}}$$

If we take the time derivative of this equation, we find

$$\frac{dp_{tot}}{dt}F_{ext} = m_{tot}\mathbf{a_{cm}}$$

The change in momentum of the entire system is the same as the net external force, which is the same as the mass-acceleration product of the center of mass. That gives us, for the acceleration, $a_{cm} = \dfrac{F_{ext}}{m_{tot}}$. If $F = F_{ext}$ is constant, as it is, and m_{tot} is also constant, then clearly the only possible answer is that the acceleration is the same for all points, the fourth option.

http://www.youtube.com/watch?v=vWVZ6APXM4w has a great demonstration of this effect. Make sure you watch the follow-up video http://www.youtube.com/watch?v=N8HrMZB6_dU and the explanation video http://www.youtube.com/watch?v=BLYoyLcdGPc too. They are a bit less than 15 minutes combined, but the effect is quite nonintuitive and so the videos are rather interesting.

(25) Well, let's see. Since the force is constant, the impulse is simply given by FT. However, I think we should solve this is a different manner than impulse.

The movement of the center of mass is given by $F_{ext} = ma_{cm}$. With a constant force, and thus a constant acceleration, we can use $\Delta y = \frac{1}{2}at^2$, with $a = a_{cm}$ and $t = T$.
However, let's not forget about gravity. $F_{net} = F - mg$, so $a_{cm} = F/m - g$. That gives us, for the displacement

$$\Delta y = \frac{1}{2}(\frac{F}{m} - g)T^2$$

(26) First, just in case we need them, let's write m_2 and m_3 in terms of m_1:

$$m_2 = \frac{m_1}{4}$$
$$m_3 = \frac{3m_1}{4}$$

Okay, so what do we know? Ignoring air drag, momentum is conserved in the x direction. After the explosion, $m_2v_2^{'} + m_3v_3^{'} = m_1v_1$.
$v_1 = x_m/t$, but we don't know t. However, we do also know (see below) that $v_2^{'} = -v_1$.

The smaller piece has a certain momentum after the launch, and the exact opposite momentum the other way back. Why? Because $p = mv$, and since it returns to exactly its launch point along the same path, the v must be the same both ways, only in opposite directions. With no air drag, it takes the same amount of time to fall from the top down to the ground, and it must traverse the same horizontal distance back as it did in getting to the top during that same time, which implies having the same horizontal velocity, which for a given mass implies the same momentum (as far as magnitude goes).

The time t taken for m_3 to hit the ground is exactly the same as that of m_2, since there is no air drag that could cause any difference in timing. Using conservation of momentum (equation one), substituting in $v_1 = x_m/t$ (equation two), substituting in the masses (equation three) and finally substituting in $v_3^{'} = (x_f - x_m)/t$:

$$\begin{aligned}
-m_2 v_1 + m_3 v_3^{'} &= m_1 v_1 \\
-m_2(\frac{x_m}{t}) + m_3 v_3^{'} &= m_1 \frac{x_m}{t} \\
-\frac{m_1}{4}(\frac{x_m}{t}) + \frac{3m_1}{4} v_3^{'} &= m_1 \frac{x_m}{t} \\
-\frac{m_1}{4}(\frac{x_m}{t}) + \frac{3m_1}{4}\frac{(x_f - x_m)}{t} &= m_1 \frac{x_m}{t}
\end{aligned}$$

All that remains is simplification. First we can eliminate t, followed by m_1 and multiplying it all by 4:

$$\begin{aligned}
-\frac{m_1}{4}(x_m) + \frac{3m_1}{4}(x_f - x_m) &= m_1 x_m \\
-(x_m) + 3(x_f - x_m) &= 4x_m
\end{aligned}$$

And the remainder doesn't need much explanation:

$$\begin{aligned}
3x_f &= 8x_m \\
x_f &= \frac{8}{3}x_m
\end{aligned}$$

Just as I hoped, all terms could be written in terms of t, so that it could be eliminated, leaving only known values m_1 (which also cancelled) and x_m, plus the unknown x_f.
Quite a nice result!

(27) We choose a coordinate system centered at the center of the Earth, which is clearly the simplest choice. The definition of the center of mass is then

$$r_{cm} = \frac{\sum_i m_i r_i}{\sum_i m_i} = \frac{m_e(0) + m_m r_{em}}{m_m + m_e} \approx 4656.2\,\mathrm{km} = 4.6562 \times 10^6\,\mathrm{m}$$

The term that is zero is the distance from the center of the coordinate system to the center of the Earth, which is obviously zero given the choice of coordinate system.

(28) The velocity just prior to the collision can be find in several ways, e.g. kinematics or conservation of energy. I will use the latter.
If we choose $U = 0$ at the ground, the initial potential energy is mgh_i, all of which becomes kinetic energy. We set the two equal and solve for v:

$$\frac{1}{2}mv^2 = mgh_i$$
$$v = \sqrt{2gh_i}$$

The magnitude of the momentum prior to the collision just $p = m\sqrt{2gh_i}$, then.

What about after the collision? Since it returns to a lower height than it was let go from, the collision must have been partially inelastic, so that kinetic energy was lost. The initial kinetic energy must be mgh_f, however. We can then find the new velocity by relating the new kinetic energy and that:

$$\frac{1}{2}m(v')^2 = mgh_f$$
$$v' = \sqrt{2gh_f}$$

The magnitude of the momentum is then $p' = mv' = m\sqrt{2gh_f}$.

The impulse is just the difference between these, $I = p_f - p_i$; however, since we have magnitudes, we need to consider that the final momentum is really in the opposite direction of the initial momentum. This turns this subtraction into an addition.

$$I = m(\sqrt{2gh_f} + \sqrt{2gh_i}) = m\sqrt{2g}(\sqrt{h_f} + \sqrt{h_i})$$

Finally, the magnitude of the average force of the ground on the ball. First, we note that $\langle F \rangle = \dfrac{\Delta p}{\Delta t}$, so the average force due to the collision is just the above answer divided by t_c. However, there is a second force involved! Gravity is pulling the ball down with a force mg, and because it is in contact with the floor, there is a normal force mg, also upwards. The answer is the sum of the two:

$$\left|\langle F \rangle\right| = \frac{m\sqrt{2g}(\sqrt{h_f} + \sqrt{h_i})}{t_c} + mg$$

(29) Well, with no other source of energy, we can rule out options (d) and (e) at once. We should also be able to rule out (c) since this is an inelastic collision. However, let's do the math.

Momentum is conserved: $m_A v_A + m_B v_B = (m_A + m_B)v'$. However, $v_B = 0$, so

$$v' = \frac{m_A v_A}{m_A + m_B}$$

The initial kinetic energy is

$$K = \frac{1}{2}m_A v_A^2$$

while final kinetic energy is

$$K' = \frac{1}{2}(m_A + m_B)(v')^2 = \frac{1}{2}(m_A + m_B)\frac{m_A^2 v_A^2}{(m_A + m_B)^2} = \frac{m_A^2 v_A^2}{2(m_A + m_B)}$$

The ratio between the two is $K'/K = \frac{m_A}{m_A+m_B}$. However, because $m_B = m_A$, we find that the kinetic energy is ***half*** of the initial, the first choice.

(30) So this is an interesting problem. It's easy to say that the answer is obviously (c), given that it may appear that all forces are internal, which is in fact not the case. It would be the case if he caught the ball, but he doesn't!

As he throws a ball, it gains momentum towards the left, while he (and the cart, via friction in his shoes) gains momentum towards the right, so that momentum is conserved. Shortly thereafter, the ball bounces, and gives momentum to the cart towards the left, and the ball momentum to the right – except that this change is ***twice as large*** as when he throw the ball. In throwing it, he changed the ball's momentum from 0 to mv, while the bounce changed it from mv to $-mv$, a change of $2mv$.

Defining the positive direction to be towards the left:
Before the throw, the cart and ball both have 0 momentum.
After the throw, the ball has momentum mv and therefore the cart $-mv$, so that the sum is zero.
After the bounce, the ball has momentum $-mv$ and therefore the cart $+mv$, so that, again, the sum is zero.

That's when the problem ends – the ball exits the system, and the momentum is never cancelled out, so the cart gains a velocity towards the left.
If he caught the ball, we could add:
After the catch, the ball transfers its momentum $-mv$ to the cart, which then gets a momentum $mv - mv = 0$, and we are back where we began.

A simpler analysis:
Initial momentum of the system is zero, and final momentum of the ball is towards the right. That ***must*** mean that there is an equal amount of momentum towards the ***left*** of the cart, or momentum would not be conserved!

(31) Spoiler alert: most of the text in this problem is justifying why the solution works, and is only necessary if you don't realize it at once. (I didn't, until it was "too late" to use the simple solution; I'd already solved it in more complex way.)

Considering the two as a system, there are no external forces, so momentum must be conserved. Momentum is a vector though, so we need to be careful with signs. If we take v_{1i} to be positive, the initial velocity of Saturn is negative, and both velocities on the right-hand side are negative.

$$m_1 v_{1i} - m_s v_s = -m_1 v_{1f} - m_s v_{sf}$$

We don't know v_{1f} and we don't know v_{sf} (the final velocity of Saturn). The latter must change, even if by an absolutely imperceptible amount.
With two unknowns, we need a second equation.
What more can we say and express as an equation? The total mechanical energy of the system should certainly be constant, since gravity is a conservative force. The mechanical energy here is

$$K_{m1} + U_{m1} + K_s + U_s = K'_{m1} + U'_{m1} + K'_s + U'_s$$

Before we try to calculate this, which will clearly not be pretty, let's try to simplify it. Gravitational potential energy depends on two things: the two masses, and the distance between them. (Plus G, which is a constant, of course.) Therefore, if the problem starts and ends at the same distance r, or it starts and ends where r is large enough that $U \approx 0$ (keep in mind that gravitational potential energy is always negative), we can assume that either that $U_{m1} = U'_{m1}$ and $U_s = U'_s$, or that all of those terms are practically zero. This simplifies things a great deal:

$$K_{m1} + K_s + = K'_{m1} + K'_s$$

So now, the condition is that the sum of the kinetic energies are the same before and after, i.e. the increase in kinetic energy in the spacecraft comes from a decrease in Saturn's kinetic energy.

With momentum and kinetic energy both conserved, we can solve this in a very simple way: this is an elastic collision. It doesn't matter that the force involved is gravity, instead of contact forces (that are mostly electromagnetic, in the end).

The mass of Saturn is about 10^{23} times greater than that of the satellite, so to an extremely good approximation, a reference frame centered on Saturn is the center of mass frame. For the same reason, the velocity of the COM frame is the velocity of Saturn – the error here is so small that a pocket calculation would round it away entirely; in fact, I couldn't get Mathematica to give me an exact answer! All I can say is that it is much, much less than 1 nanometer per second, which it gives me for m_1 as large as 10^{11} kg. I think we'll be OK with this "approximation"!

All we need to do, then, is transfer ourselves into the center of mass frame, by subtracting the center of mass velocity, find the velocity u_{1f} (using u instead of v in the COM frame), and transfer back. We transfer into it by subtracting the COM velocity:

$$u_{1i} = v_{1i} - v_{cm} = v_{1i} - v_s = 3 \times 10^3\,\mathrm{m/s} + 9.6 \times 10^3\,\mathrm{m/s} = 12.6 \times 10^3\,\mathrm{m/s}$$

Since Saturn's velocity is negative in our coordinate system, the subtraction becomes an addition. This makes sense, too: the center of mass, inside Saturn, sees the planets heading towards each other, so the net speed is larger than either of the individual speeds.

Next, we find the velocity after the collision. In the center of mass frame, this is just too easy: the signs flip. $u_{1f} = -u_{1i} = -12.6 \times 10^3\,\mathrm{m/s}$.

Finally, we convert back to the reference frame of the outside observer by ***adding*** the COM velocity of $-9.6 \times 10^3\,\mathrm{m/s}$, and end up with

$$v_{1f} = u_{1f} + -9.6 \times 10^3\,\mathrm{m/s} = -22.2 \times 10^3\,\mathrm{m/s}$$

They ask for the speed, though, so we need to drop the minus sign, and we are done.

(32) I have to say that 2 kg for the entire gun and the car seems ridiculously low! If the projectile flies away at 130 m/s, via conservation of momentum, the rail car will move backwards with a speed of at least about a quarter of that (that's just guesswork), which is crazy fast, about the speed of a car on a freeway. I can't see it being less than a tenth, at least. I suppose we'll see soon enough.

Intuitively, I have to admit I thought $\phi = \theta$ and $v_p = v_0$, and thought of the recoil as separate thing, which is clearly not correct. Let's look at a proper analysis.

Clearly, conservation of momentum will be the main way we approach this problem.
Since this is a two-dimensional problem, there will be a bit more work than in the problems we've seen earlier on.

Momentum will be conserved in the x direction, which will be quite a useful fact. What about the y direction? Well, the shell clearly gains upwards momentum, but what about the car/gun? It is pushed down, but can't move downwards. Instead, the momentum is transferred to the Earth. After the launch, gravity acts on the shell, and so the y component of its momentum will change.

Let's first think about this from the reference frame of the car. Not many strange things happen here: the shell launches at an angle θ, and moves away from you at v_0 ($v_0 \cos\theta$ in the horizontal direction, and $v_0 \sin\theta$ upwards). So far, so good.

What happens according to an observer on the ground? The vertical component of the shell's motion is unchanged, since the car is stationary along the y axis. In other words, this observer sees the shell move upwards at $v_0 \sin\theta$ m/s, same as someone on the car.
What about the horizontal component? I find it helpful to take things to extremes (even if they are unrealistic). What if the recoil speed of the car was greater than the shell's speed?
The horizontal component as seen from the ground would shrink, and since the vertical component is unchanged, the angle grows, and v_p moves closer to $v_0 \sin\theta$.
This implies that $\phi > \theta$, and of course that $v_p < v_0$.

What about a more quantitative analysis? Let's first look at the reference frame of the rail gun. The equations for the shell is

$$\begin{aligned} v_{0x} &= v_0 \cos\theta \\ v_{0y} &= v_0 \sin\theta \end{aligned}$$

Nothing strange going on there.
In the reference frame of an outside observer, standing still on the ground, things change. Such an observer would see the gun speeding towards the left at the same time the shell starts flying to the right. To him, it is clear that the gunner would see the shell move ***faster*** (in the x direction) than what he sees. In fact, in the limit where the speed of the gun and the speed of the shell are equal, the shell would move straight up to the outside observer.

The relevant equations here are also not very strange, but we can relate the two sets soon. First, the easy part:

$$\begin{aligned} v_{px} &= v_p \cos\phi \\ v_{py} &= v_p \sin\phi \end{aligned}$$

To the outside observer (and to the gunner), the rail gun is stationary along the y axis. Therefore, $v_0 \sin\theta = v_p \sin\phi$: the two agree on the vertical component. That gives us one useful relationship.

Next, we can relate the x components. The difference there is a simple reference frame shift. As mentioned above, the outside observer sees the shell having a lower speed along the x axis. The difference between the two frames is v_r.

$$v_p \cos\phi = v_0 \cos\theta - v_r$$

Next, we can relate the momenta of the two objects. The initial momentum is zero, in both reference frames. Let's write a conservation equation in the outside frame.

$$mv_p \cos\phi - Mv_r = 0$$

Since v_r is a speed in the opposite direction, we need a minus sign there. (Both terms will be positive, and their difference is zero.)

The final answer for v_p has the form

$$v_p = (v_p \sin\phi)\hat{x} + (v_0 \sin\theta)\hat{y}$$

... since the y component is the same in both reference frames. However, we only need to find the x component, and then calculate ϕ from that; so we don't really need to think of ϕ as an unknown, as far as solving the system goes. All we need is the x component of the shell, as seen from the outside reference frame.

We have

$$mv_p \cos\phi - Mv_r = 0$$

But $v_p \cos\phi = (v_0 \cos\theta - v_r)$, so

$$\begin{aligned} m(v_0 \cos\theta - v_r) - Mv_r &= 0 \\ mv_0 \cos\theta &= v_r(M + m) \\ \frac{mv_0 \cos\theta}{M + m} &= v_r \end{aligned}$$

We know all of those variables, so we can find that $v_r = 30.641\,29\,\mathrm{m/s}$. That means we can find the x component:

$$v_{px} = v_0 \cos\theta - v_r = 61.283\,\mathrm{m/s}$$

We already had v_{py} in terms of knowns, $v_0 \sin\theta$:

$$v_{py} = v_0 \sin\theta = 91.9239\,\mathrm{m/s}$$

We can then finally find v_p and the angle ϕ:

$$\begin{aligned} v_p &= \sqrt{v_{px}^2 + v_{py}^2} = 110.479\,\mathrm{m/s} \\ \phi &= \arctan\frac{v_{py}}{v_{px}} = 56.31^\circ \end{aligned}$$

(33) We can use the simple equations for rotational kinematics which is

$$f_0 = \frac{1}{T} = \frac{\omega}{2\pi}$$

This implies that $\omega = 2\pi f_0$, which answers part (a).

For part (b), we use $\alpha = \dfrac{\omega_1 - \omega_0}{\Delta t}$, where ω_0 is the initial angular velocity, and ω_1 the final angular velocity.

In this case, $\omega_0 > \omega_1$, so α is negative. However, they asked for the magnitude, so we drop the sign, and it comes to a complete halt at time t, so $\omega_1 = 0$:

$$\alpha = \frac{\omega_0}{t} = \frac{2\pi f_0}{t}$$

Finally, for part (c), we can use $\Delta\theta = \omega_0 t + \frac{1}{2}\alpha t^2$, derived from $\theta = \theta_0 + \omega_0 t + \dfrac{1}{2}\alpha t^2$. α is negative, so the addition becomes a subtraction:

$$\Delta\theta = 2\pi f_0 t - \frac{1}{2}\left(\frac{2\pi f_0}{t}\right) t^2 = 2\pi f_0 t - (\pi f_0)\, t = \pi f_0 t$$

That's it!

(34) Hmm, I wonder if there is a particular reason why part (d) is not in joules. The dimension is equivalent, but they didn't state "in joules" for whatever reason.

Anyhow, let's see. Unless otherwise specified, I will consider torques and angular momentum relative to the center of mass – though angular momentum should be the same for all points, since this is a rotation about the center of mass, so that disclaimer is probably not necessary.

To start with, we need to calculate the moment of inertia, since we only know it without the children being included.

Considering them as point masses, the total moment of inertia is just the sum of I_{cm} plus mr_o^2 for each of the children:

$$I = I_{cm} + 2mr_o^2 = 5098\,\text{kgm}^2$$

Now, then. The rotational analogue of $F = ma$ is $\tau = I\alpha$. We can find α very easily if we only find the torque relative to the center of mass.

The torque is given by $\tau = \mathbf{R} \times \mathbf{F}$, where $\mathbf{R}$ is the position vector from the origin to point where the force is applied. The force is specified as "tangential", so there is always a right angle between the two, and $\mathbf{R} \times \mathbf{F} = RF$, since $\sin(\pi/2) = 1$. The angular acceleration is then

$$\alpha = \frac{\tau}{I} = \frac{RF}{I} = 0.0785\,\text{rad/s}^2$$

Using that, we can find the final angular velocity very easily, using $\omega = \omega_0 + \alpha t$. t is given as 10 seconds in this case, so

$$\omega_{final} = 0 + \alpha t = 0.785\,\mathrm{rad/s}$$

In more familiar units, this is 8.00 seconds per rotation (0.125 Hz or 7.5 rpm).

What is the average power of the person pushing the merry-go-round? We should be able to use $W = \mathbf{F} \cdot \mathbf{v}$ here, where v is the tangential velocity, $v = \omega R$. The two are always parallel, and so

$$P = Fv = F\omega R$$

We could find the average using an integral:

$$P_{avg} = \frac{1}{t_b - t_a} \int_{t_a}^{t_b} P(t)\,dt$$

... but surely there is a better way. I looked up the relationship for power and torque, and found that $P = \boldsymbol{\tau} \cdot \boldsymbol{\omega}$, which also would require an integration (in fact, it would be the same integral), since ω is constantly changing.

I'm not sure if there is an easier way, but this integral shouldn't be very bad, so let's do it.

$$\begin{aligned}
P_{avg} &= \frac{1}{\Delta t} \int_0^{\Delta t} FR\omega(t)\,dt \\
&= \frac{FR}{\Delta t} \int_0^{\Delta t} \alpha t\,dt \\
&= \frac{FR\alpha}{\Delta t} \left(\frac{(\Delta t)^2}{2} \right) \\
&= \frac{FR\alpha \Delta t}{2}
\end{aligned}$$

For these values, $P_{avg} = 157$ watts. Quite reasonable.

And at last, the final rotational kinetic energy. The book proves that the work-energy theorem is applicable to rotational energy, so all the work done ($W = P_{avg}\Delta t$) is turned into rotational kinetic energy, so the answer is

$$W = P_{avg}\Delta t = 1570\,\mathrm{J}$$

As as update after the staff solutions are out, this was technically incorrect – but was accepted anyway. They wanted the rotational kinetic energy of the ***merry-go-round alone, without the children***, but I don't think that was too clear.

We can find the kinetic energy of that alone as $\frac{1}{2} I_{cm}\omega^2 \approx 1540$ J, instead. Not a lot harder, but I do think the question is a bit vague. Since the previous three questions were all found by considering the children's moments of inertia, I just assumed we should do here, too.

(35) Ah, interesting stuff: a non-massless pulley! Granted, we don't allow for any slipping, and the string is still of negligible mass... but this is still a considerable step towards some realism. The string is ***not massless***, however. Remember that when a string is massless, we can prove that the tension at two different points along the string must have the same magnitude... but in this case, a difference in tension is the cause of the torque that rotates the pulley! More on that in a second.

First off, since the string is inextensible, the acceleration and velocity of both masses and the rope (and the pulley, i.e. the tangential velocity at its edge) must all be the same.
Since $m_1 > m_2$, the system will accelerate such that m_1 goes downwards, m_2 upwards, and the pulley rotates counterclockwise, as seen from the direction we see it. This means $\boldsymbol{\omega}$ for the pulley will be out of the screen.

The forces on each block are easy to find. Each block has gravity and tension acting on it. I will take downwards to be the positive direction for block 1, and upwards for block 2, which then yields a common acceleration a without any trouble with signs and directions.
Let's then write Newton's second law equations for the two blocks:

For block 1, $m_1 a = m_1 g - T_1$.
For block 2, $m_2 a = T_2 - m_2 g$.

The differences in tension will cause a tangential force on the pulley, which causes a torque relative to its center, which I will call point C. This torque causes a rotation via $\tau_C = I\alpha_C$.
$a = \alpha_C R$ (this is just the time derivative of $v = \omega R$), so we can also say that $\tau = \frac{Ia}{R}$, so that $a = \frac{\tau R}{I}$. The dimension works out to be that of acceleration, which is always a good sign!
What is the torque, then? Well, the tension is tangential, and so the moment arm into the center becomes the radius R, and the angle is always 90 degrees. That gives us $\tau_C = (T_1 - T_2)R$.
We know that the rotation will be counterclockwise, so the torque must be directed out of the screen. Using $\mathbf{R} \times \mathbf{F}$ where $F = T_1 - T_2$ with a leftwards direction, the direction of positive torque, according to the cross product, is out of the screen – as it should be! (That is assuming that $T_1 > T_2$, which it should be in this case.)
This means we have three equations and three unknowns:

$$\begin{aligned} m_1 a &= m_1 g - T_1 \\ m_2 a &= T_2 - m_2 g \\ a &= \frac{2(T_1 - T_2)}{m_p} \end{aligned}$$

I substituted in $I = \frac{1}{2}m_p R^2$ in the last equation, which removed the dependence on R.
I'm never a fan of solving systems of three equations. Can we simplify the task? Solving the first two for T_1 and T_2 respectively, we can find $T_1 - T_2$ by subtracting the other sides of those equations:

$$T_1 - T_2 = m_1 g - m_1 a - m_2 a - m_2 g$$

We can then stick this into the third equation, and solve for a:

$$\begin{aligned}
a &= 2\frac{g(m_1 - m_2) - a(m_1 + m_2)}{m_p} \\
a\left(1 + \frac{2(m_1 + m_2)}{m_p}\right) &= 2\frac{g(m_1 - m_2)}{m_p} \\
a &= 2\frac{g(m_1 - m_2)}{m_p}\frac{1}{\left(1 + \frac{2(m_1+m_2)}{m_p}\right)} \\
a &= 2\frac{g(m_1 - m_2)}{m_p + 2(m_1 + m_2)}
\end{aligned}$$

Not bad!

Next up, how long does it take to move a distance d? a is clearly constant, since there are only constants in the above equation. Therefore, we can use $d = v_0 t + \frac{1}{2}at^2$ here. $v_0 = 0$, so the first term disappears. We solve the rest for t:

$$\begin{aligned}
\frac{1}{2}at^2 &= d \\
t^2 &= \frac{2d}{a} \\
t &= \sqrt{\frac{2d}{a}}
\end{aligned}$$

All that remains is then to stick the above, semi-complex expression into the square root:

$$\begin{aligned}
t &= \sqrt{\frac{2d}{2\frac{g(m_1-m_2)}{m_p+2(m_1+m_2)}}} \\
t &= \sqrt{d \cdot \frac{m_p + 2m_1 + 2m_2}{g(m_1 - m_2)}}
\end{aligned}$$

(36) Hmm, this looks as if it should be easier than the previous problem.

The cable has negligible mass, so the tension ought to be zero without the mass there. Therefore, the tension is all due to gravity acting on the block.

Newton's second low on the block, taking downwards to be positive, is

$$m_2 a = m_2 g - T$$

The tension then acts on the pulley, in a tangential fashion (as in the last problem, though on the side this time, instead of the top), so that the torque relative to its center is $\tau_C = \mathbf{R} \times \mathbf{T} = RT$, with the direction being out of the screen (since the rotation will be counterclockwise). $\mathbf{R}$ is then the position vector from the center to the point where the force acts, so the $\sin\theta$ term is again always 1, due to the 90 degree angle between the two vectors.

This torque causes an acceleration of the pulley via

$$\tau_C = I_c\alpha \Rightarrow \alpha = \frac{\tau_C}{I_c} = \frac{RT}{I_c}$$

$a = \alpha R \Rightarrow \alpha = \dfrac{a}{R}$, so

$$\begin{aligned}
\frac{a}{R} &= \frac{RT}{I_c} \\
a &= \frac{R^2 T}{I_c}
\end{aligned}$$

Two equations, with a and T as two unknowns. We can solve both for a, sen them equal, and find T:

$$\begin{aligned}
g - \frac{T}{m_2} &= \frac{R^2 T}{I_c} \\
\frac{T}{m_2} &= g - \frac{R^2 T}{I_c} \\
T + \frac{m_2 R^2 T}{I_c} &= m_2 g \\
T\left(1 + \frac{m_2 R^2}{I_c}\right) &= m_2 g \\
T &= m_2 g \frac{1}{\left(1 + \frac{m_2 R^2}{I_c}\right)} \\
T &= \frac{m_2 g}{1 + \frac{m_2 R^2}{I_c}} \\
T &= \frac{m_2 g I_c}{I_c + m_2 R^2}
\end{aligned}$$

That then answers part (b). Let's stick in into the other equation and find a:

$$\begin{aligned}
a &= \frac{R^2}{I_c} \frac{m_2 g I_c}{I_c + m_2 R^2} \\
a &= g \frac{m_2 R^2}{I_c + m_2 R^2}
\end{aligned}$$

Finally, the speed of the object as it hits the floor.
As previously, $v_0 = 0$ and a is a constant, so we can use basic kinematics equations... only that those involve both t and h.
We can solve $v = at$ for t, and find $t = \dfrac{v}{a}$. Substitute that into the one that relates acceleration to distance:

$$\begin{aligned}
\frac{1}{2} a \left(\frac{v}{a}\right)^2 &= h \\
\frac{1}{2} \frac{v^2}{a} &= h
\end{aligned}$$

$$v = \sqrt{2ha} = \sqrt{2hg\frac{m_2R^2}{I_c + m_2R^2}}$$

And that's it for this one!

(37) Hmm. Well, unfortunately I don't have a yo-yo (nor anything similar, like a spool of sewing thread), so I can't really try it out! I also have no real intuition of how it behaves here, though I do know that it rolls away when β is large, and towards you when β is small.

Since $\sin\beta \approx \beta$ for small angles, option 1 cannot be true; it is more likely to roll to the left if $\sin\beta$ is large.
Option 2 could be true.
$\cos\beta$ becomes smaller as the angle grows. Larger angle means more likely to roll to the left, so smaller cosine also means that. This means we can rule out option 4.

Left are options 2 and 3, though I don't see any obvious way to choose between the two without actually making the calculations! Let's have a look at that.

What can we say about the yo-yo? There are external forces, which also causes external torques. R_1 acts as a moment arm for our pull, for the torque relative to the center of the yo-yo.
There is also the force due to friction. Friction acts along R_2, and also causes a torque on the yo-yo, in the opposite direction to the torque due to the pull.

I will use a coordinate system where leftwards motion is positive.

If we draw a free-body diagram (considering only the center of mass; we should not do this for torques, since distances matter there) and use P to notate the force due to our pull, we find $P\cos\beta$ in the rightwards direction (negative, in this coordinate system), and F_{fr} towards the left.

Using Newton's second law, we can write for the center of mass,

$$ma_{cm} = F_{fr} - P\cos\beta$$

We can then calculate the torque due to this pulling force, as $\mathbf{R} \times \mathbf{P}$; the angle to the position vector is always 90 degrees, and $\mathbf{R}$, the moment arm, is R_1, since the string is wrapped around R_1:

$$\tau_P = R_1 P \text{ (direction: out of the screen / causes CCW rotation)}$$

There is also a torque due to friction. Again, the angle is always 90 degrees, so $\mathbf{R} \times \mathbf{F_{fr}}$ is just the magnitude of the two multiplied together, where the moment arm is now R_2 (friction acts on the outside of the yo-yo):

$$\tau_{fr} = R_2 F_{fr} \text{ (direction: in to the screen / causes CW rotation)}$$

When $\tau_{fr} > \tau_P$, there is clockwise rotation, and the yo-yo rolls towards the right. When τ_P wins, it moves towards the right.
Since the torque must reverse direction between these two cases, there is also the possibility that the net torque is zero.

Net torque (CCW/left): $\tau = \tau_P - \tau_{fr} = R_1 P - R_2 F_{fr}$
Using the condition that the torque is zero, we can relate F_{fr} to P:

$$R_1 P - R_2 F_{fr} = 0$$
$$F_{fr} = \frac{R_1}{R_2} P$$

Making this substitution into the Newton's second law equation:

$$a = \frac{P}{m}\left(\frac{R_1}{R_2} - \cos\beta\right)$$

This acceleration is positive when the yo-yo accelerates to the left, due to the choice of coordinate system, so the condition for moving towards the left is that the above expression is greater than zero. We set up the inequality and solve:

$$\frac{P}{m}\left(\frac{R_1}{R_2} - \cos\beta\right) > 0$$

That happens when

$$\frac{R_1}{R_2} > \cos\beta$$

which of course is the same as one of the answer options,

$$\cos\beta < \frac{R_1}{R_2}$$

So the answer is option 3,
Yo-Yo rolls to the left if $\cos\beta < \frac{R_1}{R_2}$, and to the right if $\cos\beta > \frac{R_1}{R_2}$.

(38) Torque is the force times the moment arm length, which is easy in this case. The relevant force is mg, which acts on the center of mass. Since the stick is uniform, the center of mass is at $\frac{\ell}{2}$. The torque relative to point B is simply

$$\tau_B = mg\frac{\ell}{2}$$

since the angle between the two vectors is 90 degrees (just after the stick is released).
We can now find the angular acceleration by knowing the torque, via $\tau = I\alpha$. The moment of inertia in question is the one for the rod, about its end, since that is the pivot point.
Since the pivot point is clearly not at the center of mass in this case, we need to use the parallel axis theorem.

I remember that $I_c = \frac{1}{12}m\ell^2$ for a rod, but we then need to add a term due to the parallel axis theorem. The distance between the center of mass and this new axis is half the rod's length, so via the parallel axis theorem,

$$I_B = \frac{1}{12}m\ell^2 + m\left(\frac{\ell}{2}\right)^2$$

Now, using $\tau_B = I_B\alpha$, we can solve for α:

$$\begin{aligned}
mg\frac{\ell}{2} &= \alpha\left(\frac{1}{12}m\ell^2 + m\left(\frac{\ell}{2}\right)^2\right) \\
g\frac{\ell}{2} &= \alpha\frac{1}{12}\ell^2 + \alpha\frac{\ell^2}{4} \\
g &= \alpha\frac{1}{6}\ell + \alpha\frac{\ell}{2} \\
g &= \alpha\ell\left(\frac{1}{6} + \frac{1}{2}\right) \\
\frac{g}{\ell} &= \alpha\left(\frac{4}{6}\right) \\
\frac{3g}{2\ell} &= \alpha
\end{aligned}$$

Next, they want to know the vertical acceleration of the center of mass. α describes the angular acceleration about point B, that the center of mass undergoes. We can use the relationship $a = \alpha R$, and in this case, $R = \frac{\ell}{2}$.

$$a = \frac{3g}{2\ell}\frac{\ell}{2}$$

$$a = \frac{3g}{4}$$

Finally, what is the magnitude of the vertical component of the hinge force at B?
Well, first up, what is hinge force? I haven't seen that term before, but I assume it is the normal force from the table on the end of the stick, especially as they give it as N.

It's clearly not zero, or the stick would just fall right through.
What we do here is to remember the videos and demonstration of an impulse on a ruler. No matter ***where*** on the ruler the force is exerted, the acceleration of the center of mass is affected in the same way. Therefore, we can use Newton's second law to relate the net downwards force, $mg - N$, with the mass times acceleration of the stick:

$$\begin{aligned}
mg - N &= ma \\
N &= m(g - a)
\end{aligned}$$

We know a from above, so we can substitute than in there:

$$N = m(g - \frac{3g}{4})$$
$$N = \frac{mg}{4}$$

(39) In part (a), the disk is fixed. We begin by calculating the total moment of inertia for rotating about what I will call point P, the point where the rod is mounted to the roof.

For the rod, we use the parallel axis theorem:

$$I_{rod,end} = \frac{1}{12}m\ell^2 + m\left(\frac{\ell}{2}\right)^2 = \frac{m\ell^2}{3}$$

We also use the parallel axis theorem for the disc. About the disc's own center of mass, $I_{cm} = \frac{1}{2}mR^2$. We need to add to that the distance to the new axis, which is ℓ away.

$$I_{disc} = \frac{1}{2}mR^2 + m\ell^2 = m\left(\frac{R^2}{2} + \ell^2\right)$$

The total moment of inertia for rotation about the pivot point, for the combination is then

$$I_P = \frac{1}{6}m(8\ell^2 + 3R^2)$$

Let's now consider the torque (relative to the pivot point, P). There is a torque due to the rod (because of gravity acting on its center of mass), and a torque due to the wheel (again, due to gravity acting on its center of mass). These torques depend on the moment arm length, the force of gravity, and the sine of the angle between the two, via the cross product definition:

$$\tau_{P,rod} = \mathbf{r_P} \times \mathbf{F_g} = \frac{\ell}{2}mg\sin\theta$$
$$\tau_{P,disc} = \mathbf{r_P} \times \mathbf{F_g} = \ell mg\sin\theta$$
$$\tau_P = \tau_{P,rod} + \tau_{P,disc} = \frac{3}{2}mg\ell\sin\theta$$

This is a restoring torque, that is always trying to get things back to equilibrium. Using Newton's second law, or perhaps rather its rotational equivalent $\tau = I\alpha$, only with a negative sign in front since it is a restoring torque:

$$\alpha = -\frac{3}{2I_P}mg\ell\sin\theta$$

Using $\alpha = \ddot{\theta}$, and a small angle approximation $\sin\theta \approx \theta$, we get

$$\ddot{\theta} + \frac{3}{2I_P}mg\ell\theta = 0$$

... which is simple harmonic oscillator. This is of course what we wanted all along. The period is then given by $\frac{2\pi}{\omega}$, where ω^2 is the stuff multiplying the square root. We flip that upside down, take the square root, and multiply by the 2π:

$$T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{2I_P}{3mg\ell}}$$

$$\begin{aligned} T &= 2\pi\sqrt{\frac{m(8\ell^2+3R^2))}{9mg\ell}} \\ &= \frac{2\pi}{3}\sqrt{\frac{8\ell^2+3R^2}{g\ell}} \end{aligned}$$

What happens for part (b)? When the disc is free to spin, it is also free to stay stationary, so to speak. That is, when it is ***fixed***, it is ***forced*** to rotate along with the motion. If we made a vertical mark at the top of the disk, that mark would turn at an angle θ together with the rod and the rest of the disc.
Because of this, it has a spin component of moment of inertia of $I_{cm,disk} = \frac{1}{2}mR^2$, in addition to the orbital component of mR^2.

With a frictionless bearing, on the other hand, that vertical mark on the disk would be vertical at all times, which means it is not spinning any more.
There is no torque acting on the disc: gravity acts equally on all points, and since it is attached in the center with ***no friction***, there can be no torque due to the pin there, either.

This means that the term for the disc's moment of inertia that is due to the spin disappears, and $I_{disc} = mR^2$ – only the orbital part remains.
So we can think of the motion of the disc as having two components: one "orbital", and one "spin". In the previous case, both were present. In this case, when the disc can stay stationary (have no spin motion at all), only the orbital motion remains, and so only the orbital part of the moment of inertia remains.

The torque is unchanged, since we calculated that based on the center of mass. What changes is I_P; the part due to the rod is unchanged, but that due to the disc changes, so that

$$I_{disc} = m\ell^2$$

The total moment of inertia about the pivot point is again the sum of the two moments of inertia:

$$I_P = \frac{m\ell^2}{3} + m\ell^2 = \frac{4m\ell^2}{3}$$

That is the only thing that changes, so we stick that into the equation for the period:

$$T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{2I_P}{3mg\ell}}$$

$$T = 2\pi\sqrt{\frac{2(\frac{4m\ell^2}{3})}{3mg\ell}}$$
$$T = 2\pi\sqrt{\frac{8m\ell^2}{9mg\ell}}$$
$$T = \frac{2\pi}{3}\sqrt{\frac{8\ell}{g}}$$

That solves this problem!

(40) It is quite remarkable that ω_1 and ω_2 are independent of μ, and it is also independent of the time it takes for the equilibrium to be reached (i.e independent of how hard one pushes the disks against each other)."

Unlike most cases, I'm writing almost all the text for this problem after having solved it. (I usually write while solving, then clean up the text when I have everything correct, and feel I understand the solution fully.)
This problem was certainly the most confusing of the week for many, including myself until I thought about it for quite a long while, while following the forum discussions.

First: angular momentum will ***not be conserved***! This is an extremely important point, of course – solving this by assuming it ***is*** conserved does not work. (Except a side note, below.)

It is clear that there is friction between the disks, or they could not affect one another. Friction is proportional to the normal force, but since the disks are at the side of one another, there is no natural force to push them together.
This force must be provided by something ***external*** to the system, such as a person holding the two axles.
In addition, the force due to friction acts "upwards" and "downwards" on the two disks, respectively (in the order shown in the figure). With a net force upwards or downwards on an object, the center of mass must accelerate upwards! $a_{cm} = \frac{F_{ext}}{m}$ must hold for the center of mass. Therefore, in order for the disks to stay where they are, another ***external force*** comes in: the leftwards disk must be forced down, and the rightwards disk must be forced up, or they will not stay put.

Now, in a bit of a freak coincidence, the correct solution ***can*** be found by assuming angular momentum is conserved, and by assuming that $\frac{m_1}{m_2} = \frac{R_1}{R_2}$, which is incorrect! Since mass is proportional to volume, and volume is $\pi R_i^2 h$, the correct equation is $\frac{m_1}{m_2} = \frac{R_1^2}{R_2^2}$.

Combining this ***correct*** equation with conservation of momentum, and you can find an answer which looks like the correct ones, only that all exponents (on R_1 and R_2) are one too large! If you then also use the incorrect formula for the masses above, ***the error cancels out***, and you find the correct answer!
To be clear, this does not imply that the ***method*** is correct – it is trivial to show that the total angular momentum must change! See the end notes below, after my solution.

My solution

Okay, so let's consider this in more detail. To begin with, note that below, any time I say the leftmost disk, I mean the leftmost disk in the figure above, which is disk 2 (since it has radius R_2 and ends up spinning at ω_2, I call it disk 2). The rightmost disk is disk 1.

Okay. First, we can write two equations regarding the change in angular momentum of each disk on its own. By the way, because we also deal with objects spinning about an axis through their center of mass, we don't need to specify the point relative to which we find the angular momentum, as the answer is the same for all such points.
The two equations relating these changes are

$$\Delta L_1 = I_1(\omega_1 - \omega)$$
$$\Delta L_2 = I_2(\omega_2 - 0)$$

Disk 2 starts with 0 initial angular momentum, so its final angular momentum $I_2\omega_2$ equals the change.

The most important forces involved will be the frictional forces due to the contact of the two disks. The magnitude of these forces is unknown (they depend on how hard the disks are pushed together, which we are not told), but that doesn't matter for the solution, as the problem sort-of states.

Disk 1 spins clockwise to begin with. When it comes in contact with disk 2, there is a frictional force on disk 1, due to disk 2. This frictional force must oppose the relative motion, and so it acts downwards (counterclockwise) on disk 1, slowing its rotation. (Anything else would be crazy!)
Via Newton's third law, there is an equal but opposite force on disk 2 (which is still stationary), due to disk 1. This means that force is upwards, i.e. causes counterclockwise rotation.

These forces must cause torque on the two disks, or their rotation would be unaffected (since torque causes change in rotational motion, just as force causes change in linear motion).
For disk 1, there is friction on the left side, acting downwards tangentially along the disk. The torque caused by this, relative to the disk's center, is the cross product of the position vector from the center and the friction vector:

$$\tau_1 = \mathbf{R_1} \times \mathbf{F_{fr}} = -R_1 F_{fr}$$

As for direction, via the right-hand rule, it is out of the screen, i.e. acts counterclockwise. Again, anything else would be crazy, since the opposite torque would speed the disk's rotation up.
I notate this with a minus sign, as I use clockwise rotation (into the screen) as positive. That is the initial rotation, so I figured it would make sense to call that positive.

For disk 2, we do the same process. Friction is on the right side, acting upwards, tangentially. The torque relative to this disk's center is

$$\tau_2 = \mathbf{R_2} \times \mathbf{F_{fr}} = -R_2 F_{fr}$$

The direction of this torque is also out of the screen, i.e. it acts counterclockwise. This is also clear if you consider the direction of the motion; the disk starts to spin such that the tangential velocity is reduced, so that slipping is reduced. This is only possible if it spins up counterclockwise.

Note that both torques act counterclockwise, which means angular momentum is increasing in the CCW direction for both disks, and therefore for the system of the two disks combined. This can clearly not be the case if angular momentum is conserved/held constant; if it were held constant, the increase in one disk must be matched by a decrease in the other.

I used F_{fr} for both frictional forces, since they have the same magnitude via Newton's third law. Their directions do differ, however.

Say that this frictional force acts for an unknown time Δt. We can then also write the changes in angular momenta as

$$\begin{aligned}\Delta L_1 &= -F_{fr}\Delta t R_1\\ \Delta L_2 &= -F_{fr}\Delta t R_2\end{aligned}$$

using the relationship $\dfrac{dL}{dt} = \tau$, which becomes $\Delta L = \tau\Delta t$ if we bring it out of the differential form, and rearrange.

So, we have four equations; two per disk, both of which define the change in angular momentum. If we set them equal in pairs, we get two equations, with many unknowns (F_{fr}, Δt, I_1, I_2, ω_1 and ω_2 – wow).

Not to worry, as we can eliminate many of those. First, we can eliminate I_2 by writing it in terms of I_1. It is specified that the disks have the same density and thickness, so we can relate their masses and/or moments of inertia by comparing the radii.

The mass of a disk with some density ρ is $\pi R_i^2 h\rho$. The moment of inertia is then $\frac{1}{2}mR_i^2 = \frac{1}{2}(\pi R_i^2 h\rho)R_i^2$, and the ratio of the two moments of inertia becomes

$$\frac{I_2}{I_1} = \frac{\frac{1}{2}(\pi R_2^2 h\rho)R_2^2}{\frac{1}{2}(\pi R_1^2 h\rho)R_1^2} = \frac{R_2^4}{R_1^4}$$

which gives us $I_2 = I_1\dfrac{R_2^4}{R_1^4}$. It is proportional to R^4 because both the mass and the moment of inertia are, on their own, proportional to R^2.

Combining the two pairs of ΔL equations, and making the substitution for I_2 using the relationship above, we have

$$\begin{aligned}I_1\omega_1 - I_1\omega &= -F_{fr}\Delta t R_1\\ I_1\frac{R_2^4}{R_1^4}\omega_2 &= -F_{fr}\Delta t R_2\end{aligned}$$

We can divide the two equations – note how this gets rid of F_{fr}, Δt and I_1 all at once!

$$
\begin{aligned}
\frac{I_1\omega_1 - I_1\omega}{I_1\frac{R_2^4}{R_1^4}\omega_2} &= \frac{-F_{fr}\Delta t R_1}{-F_{fr}\Delta t R_2} \\
R_1^4\frac{\omega_1 - \omega}{R_2^4\omega_2} &= \frac{R_1}{R_2} \\
R_1^3\frac{\omega_1 - \omega}{R_2^3\omega_2} &= 1 \\
R_1^3\omega_1 - R_1^3\omega &= R_2^3\omega_2 \\
R_1^3\omega_1 - R_2^3\omega_2 &= R_1^3\omega
\end{aligned}
$$

A bit of a prettier way to write this would be to consider the relative magnitudes of the two torques instead (the torques are ***not*** the same in magnitude, but the frictional force that causes them ***are***). The end result is the same; it is simply a different way to write the equations.

Another relationship we can use is that of the linear velocities of the two disks, which need to match for there to be no slipping.

$$\omega_1 R_1 = -\omega_2 R_2$$

We then have two equations and two unknowns:

$$
\begin{aligned}
R_1^3\omega_1 - R_2^3\omega_2 &= R_1^3\omega \\
\omega_1 R_1 &= -\omega_2 R_2
\end{aligned}
$$

The solutions are

$$
\begin{aligned}
\omega_1 &= \frac{R_1^2\omega}{R_1^2 + R_2^2} \\
\omega_2 &= -\frac{R_1^3\omega}{R_2(R_1^2 + R_2^2)}
\end{aligned}
$$

They asked for the magnitudes, though, so we need to drop the minus sign in front of ω_2.

Aftermath

So with the solutions in mind, what happens in terms of angular momentum?

$$L_{initial} = I_1\omega$$

$$L_{final} = I_1\omega_1 + I_2\omega_2$$

... keeping in mind that ω_2 is negative. We know that $\omega > \omega_1$, and that the moments of inertia don't change. The change in angular momentum is

$$\Delta L_{sys} = L_{final} - L_{initial} = (I_1\omega_1 + I_2\omega_2) - I_1\omega$$

Which is, using the expressions for the solutions ω_1 and ω_2, and using $I_2 = I_1\dfrac{R_2^4}{R_1^4}$:

$$\Delta L_{sys} = -\frac{I_1 R_2^2(R_1 + R_2)}{R_1(R_1^2 + R_2^2)}\omega$$

Not a very pretty expression (I think simplification might have made it uglier), but we can consider the simpler case when $R_2 = R_1$:

$$\Delta L_{sys,R_1=R_2} = -I_1\omega$$

In this special case, the change in angular momentum is exactly the negative of the ***initial*** angular momentum: the net angular momentum is ZERO afterwards.
This does actually make a whole lot of sense. If the disks are identical (same thickness, radii and density implies same mass and same moment of inertia), they will rotate at the same angular speed... but opposite directions! Since $L_C = I_c\omega$, and both disks have the same magnitude (but opposite direction) of ω, and the same I_c, the angular momentum of disk 1 is exactly the opposite of disk 1, and the sum is zero.

The solution for angular velocities in this special case is $\omega_1 = \omega/2$ and $\omega_2 = -\omega/2$, so.

$$L_{final,R_1=R_2} = I\frac{\omega}{2} + I\left(-\frac{\omega}{2}\right) = 0$$

(41) The motion of the center of mass is very easy to derive. Say the rod is hit by an impulse I. It has zero momentum to begin with, so its new total momentum is I.
$\mathbf{p_{tot}} = m_{tot}\mathbf{v_{cm}}$ must hold, and so

$$v_{cm} = \frac{I}{m} = 4/3\,\mathrm{m/s}$$

In the absence of external forces, this is held constant.
Part (c) is also extremely simple, then:

$$D = v_{cm}t = (4/3\,\mathrm{m/s})(8\,\mathrm{s}) = 32/3\,\mathrm{m}$$

The rotational motion is bit more tricky.
We can choose to consider torques relative to the center of mass, or relative to a point along the line of the impulse. (We ***can*** choose differently, but why would we?)
I'm not sure which is easier in the end, but I find it easier to visualize it relative to the center of mass, point C.

The torque is then $\tau_C = Fd$, where F is the magnitude of the force, and d the distance between C and X. It we multiply both sides by the (unknown) impact time, we get $\tau_C\Delta t = (F\Delta t)d$, which is the same as saying $L_C = Id$. The initial angular momentum relative to point C is zero, so this is the total angular momentum after the hit.

The angular momentum relative to point C is about the center of mass; so $L_C = I_c\omega$ also holds (where I_c is the moment of inertia of the rod around the center). Setting the two equal,

$$I_c\omega = Id$$

$$\omega = \frac{Id}{I_c} = \frac{Id}{\frac{1}{12}m\ell^2}$$

For the numbers given, $\omega = 9.6$ rad/s.

After 8 seconds, it has rotated 76.8 radians, which about 12.22 rotations; the angle should be a bit less than 90 degrees (0.22 radians), in other words.
To find the angle,

$$\theta = 76.8 \bmod (2\pi) = 1.402\,\text{rad} = 80.32^\circ$$

where mod gives the remainder after a division. The result is the same as $76.8 - (2\pi \times \lfloor \frac{76.8}{2\pi} \rfloor)$.
Finally, the total kinetic energy. This is simply the sum of the translational (linear) kinetic energy, and the rotational:

$$K = \frac{1}{2}mv_{cm}^2 + \frac{1}{2}I_c\omega^2 = 2.6667\,\text{J} + 2.88\,\text{J} = 5.5467\,\text{J}$$

(42) (a) What is the speed v_0 (in m/s) of the earth in its orbit of radius $R = 1.5 \times 10^{11}$ m around the sun with a mass $M = 1.99 \times 10^{30}$ kg? Take the gravitational constant $G = 6.674 \times 10^{-11}\,\text{m}^3\text{kg}^{-1}\text{s}^{-2}$.

First, if we treat the orbit as circular (as they clearly want us to: it is "near circular", and they ask for ***the*** orbital speed; elliptical orbits don't have a single speed, but one that varies over time.

I tend to not always remember the equation here, but I do always remember that the total mechanical energy is $K_e + U = \frac{1}{2}U$. We can rearrange that prior to substitution of the actual values, and then solve for v_0:

$$K_e + U = \frac{1}{2}U$$
$$K_e = -\frac{1}{2}U$$

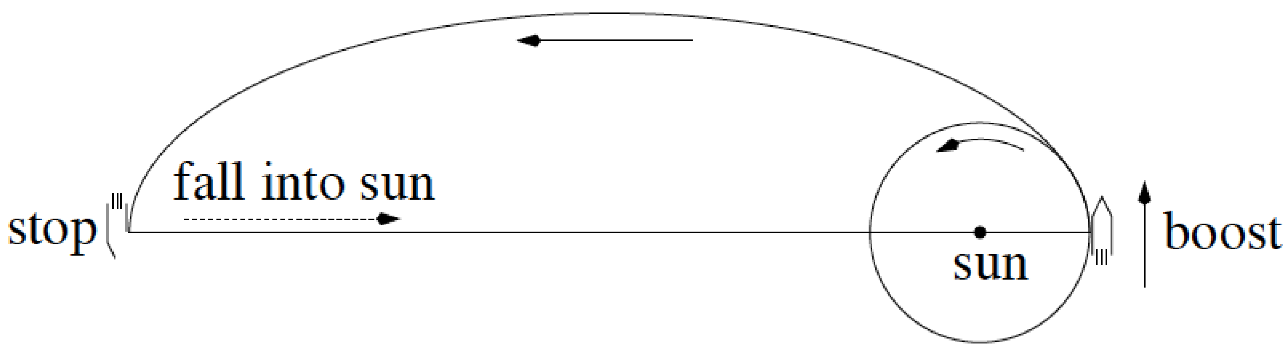
$$\frac{1}{2}mv_0^2 = \frac{mMG}{2R}$$
$$v_0 = \sqrt{\frac{MG}{R}} \approx 29\,756\,\text{m/s}$$

“We want the spacecraft to fall into the sun. One way to do this is to fire the rocket in a direction opposite to the earth's orbital motion to reduce the spacecraft's speed to zero (relative to the sun).

(b) What is the total impulse I_0 that would have to be given by the rocket to the spacecraft to accomplish this? You may ignore the effect of the earth's gravitation as well as the orbital speed of the spacecraft around the earth as the latter is much smaller than the speed of the earth around the sun. Thus, you may assume that the spacecraft, before the rocket is fired, has the same speed in its orbit around the sun as the earth. Express your answer in terms of m and v_0.”

Given that we can neglect almost everything, this is very easy. We have an initial momentum mv_0 (if we indeed neglect the orbital speed around the Earth), and we need to get that speed down to zero, which implies getting our momentum to zero. The change is simply $-mv_0$.
The answer that is accepted is mv_0, however. A bit strange, to me – they don't ask for any magnitudes, and since mv_0 is clearly the current momentum, I would argue that an impulse of $I = mv_0$ would double the current momentum (and thus speed) in the direction the spacecraft is currently moving.
Ah well.

“We will now show that there is a more economical way of doing this (i.e., a much smaller rocket can do the job). By means of a brief rocket burn the spacecraft is first put into an elliptical orbit around the sun; the boost is provided tangentially to the earth's circular orbit around the sun (see figure). The aphelion of the new orbit is at a distance r from the sun. At aphelion the spacecraft is given a backward impulse to reduce its speed to zero (relative to the sun) so that it will subsequently fall into the sun.



(c) Calculate the impulse I_1 required at the first rocket burn (the boost). Express your answer in terms of I_0, R and r.”

Okay, so aphelion is the furthest it ever comes from the Sun (perihelion is the closest). If we call aphelion point A, perihelion point P and the Sun point Q, then we have AQ + PQ = $2a$, where a is the orbit's semi-major axis.
If the distance AQ is r, and the current distance PQ from us to the Sun is R, then via the diagram provided, clearly $2a = R + r$, where a is the semi-major axis of the new, elliptical orbit.

Combined with the next question, we need to find the impulse required to move into an elliptical orbit with new speed v_1, such that $a = \dfrac{R + r}{2}$.

We make a burn so that the new speed is v_1, and the new (linear) momentum mv_1. The impulse is then $m(v_1 - v_0)$, but we don't know v_1 yet.
We can figure out v_1 by conservation of energy. ***After*** the burn, energy is conserved (but not

during, of course). The new kinetic energy, plus the new (same as before) potential energy must equal half of the potential energy of the new, elliptical orbit:

$$\frac{1}{2}mv_1^2 - \frac{mMG}{R} = -\frac{mMG}{2a} v_1^2 = \frac{2MG}{R} - \frac{MG}{a}$$
$$v_1^2 = 2MG\left(\frac{1}{R} - \frac{1}{2a}\right)$$
$$v_1 = \sqrt{2MG\left(\frac{1}{R} - \frac{1}{R+r}\right)} = \sqrt{\frac{2GMr}{R(R+r)}}$$

Now, here's the slightly tricky part... We know that $v_0 = \sqrt{\frac{MG}{R}}$, and we need to write the above in terms of v_0. Thankfully, with the simplification done, that is in fact now the opposite of tricky. It could have been! We simply remove those variables from inside the square root, and tack on v_0 outside:

$$v_1 = v_0\sqrt{\frac{2r}{(R+r)}}$$

Next, we need to write this in terms of impulse. $I_0 = v_0/m$, and $I_1 = m(v_1 - v_0)$.

$$I_1 = mv_0\sqrt{\frac{2r}{(R+r)}} - mv_0 = I_0\left(\sqrt{\frac{2r}{(R+r)}} - 1\right)$$

"(d) What is the speed v_2 of the spacecraft at aphelion? Express your answer in terms of v_0, R and r."

Finally, we need to convert v_1 into v_2. v_1 at perihelion, and v_2 is at aphelion. The speed at perihelion is much greater than that at aphelion.

Angular momentum is the same at both locations. Therefore, $Rmv_1 = rmv_2$, or $Rv_1 = rv_2 \Rightarrow v_2 = \frac{R}{r}v_1$.

$$v_2 = \frac{R}{r}v_0\sqrt{\frac{2r}{R+r}}$$

"(e) Calculate the impulse I_2 required at the second rocket burn (at aphelion). Express your answer in terms of I_0, R and r."

This shouldn't be too bad now. We need to bring v_2 down to zero, so

$$I_2 = mv_2 = mv_0\frac{R}{r}\sqrt{\frac{2r}{R+r}} = I_0\frac{R}{r}\sqrt{\frac{2r}{R+r}}$$

Again, they want a positive value.

"(f) Compare the impulse under b) with the sum of the impulses under c) and e) (i.e find $I_0 - (I_1 + I_2)$), and convince yourself that the latter procedure is more economical. Express your answer in terms of I_0, R and r."

I will call this ΔI for a lack of a better name.

$$\Delta I = I_0 - \left(I_0 \left(\sqrt{\frac{2r}{(R+r)}} - 1 \right) + I_0 \frac{R}{r} \sqrt{\frac{2r}{R+r}} \right)$$
$$\Delta I = I_0 - I_0 \left(\sqrt{\frac{2r}{(R+r)}} - 1 + \frac{R}{r} \sqrt{\frac{2r}{R+r}} \right)$$

To convince ourselves, we need to find that the expression in parenthesis is always such that $\Delta I > 0$ (otherwise, it's equally or even less efficient).

$$\Delta I = I_0 - I_0 \left(\sqrt{\frac{2r}{(R+r)}} \left(1 + \frac{R}{r} \right) - 1 \right)$$
$$\Delta I = 2I_0 - I_0 \left(\sqrt{\frac{2r}{(R+r)}} \left(1 + \frac{R}{r} \right) \right)$$
$$\Delta I = I_0 \left(2 - \sqrt{2} \sqrt{\frac{R+r}{r}} \right)$$

Finally, we can truly convince ourselves by solving this for r manually:

$$2 - \sqrt{2} \sqrt{\frac{R+r}{r}} > 0$$
$$\sqrt{2} \sqrt{\frac{R+r}{r}} < 2$$
$$2 \frac{R+r}{r} < 4$$
$$2R < 2r$$
$$R < r$$

So indeed, for ***any*** chosen $r > R$, this is more efficient. Of course, we need to remain in orbit for the result to be useful; we could of course make a ridiculous burn to reach an extremely high speed and escape, which would be less efficient, but in that case, we would have any r as we would not be in an elliptical orbit.

Phew! This took a very long time for me – a while to figure out how to solve part (c), and a ***very*** long time to figure out where I was going wrong. I got v_0 correct at once, but then accidentally wrote down an incorrect expression in my notes: $v_0 = \sqrt{\frac{2MG}{R}}$. You may notice that is the escape velocity for Earth's orbit, not v_0 – I did too, only the day after I started working on this problem. Once I noticed, everything else went rather smoothly.

(43) Well, part (a) is easy, at least. We even saw that expression, exactly as-is, during the lecture (indeed, in the part about Cygnus X-1, i.e. this system).

We use the equation for periods of elliptical orbits, sometimes known as Kepler's third law (though Kepler only said $T^2 \propto a^3$; the rest was calculated later), only we substitute in $m_1 + m_2$ for the mass, and $r_1 + r_2$ for the orbital radius:

$$T = \sqrt{\frac{4\pi^2(r_1 + r_2)^3}{G(m_1 + m_2)}}$$

For part (b), they tell us the period T, and the velocity v_2. Finding r_2 is a piece of cake, then, if we don't get wrapped up in complex thinking!

$$\frac{2\pi r_2}{T} = v_2$$

$$r_2 = \frac{v_2 T}{2\pi} = \frac{(148 \times 10^3\,\mathrm{m/s})(5.6\,\mathrm{days})}{2\pi} \approx 1.1397 \times 10^{10}\,\mathrm{m}$$

We now know T and r_2, but not m_1, m_2 or r_1. For finding r_1, they gave us a hint, though:

"Hint: Your calculations will be greatly simplified if instead of r_1 you set up your equations in terms of r_1/r_2, and using some relation between the distances and the masses. Once you express your equation in terms of r_1/r_2, you will find a third order equation in r_1/r_2. Only one solution is real; the other two are imaginary. There are various ways to find an approximation for r_1/r_2. You can find the solution by trial and error using your calculator, or you can plot the function."

Hmm. Well, via the center of mass definition,

$$m_1 r_1 = m_2 r_2$$

We can certainly find r_1/r_2 from that:

$$\frac{r_1}{r_2} = \frac{m_2}{m_1}$$

They also tell us that $m_2 = 30M_{Sun}$.

$$r_1 = \frac{30M_{Sun}}{m_1} r_2$$

Here is where we need to start applying the hint given. I will copy the staff solution a bit here (i.e. I'm writing this part after the deadline has passed to clean up). We can assign a variable $x = r_1/r_2$.

This then implies that $m_1 = m_2/x$ using the above relationships.

We can now start rewriting the period equation. First, we square it to get rid of the square root on the right-hand side. Then, we factor out r_2^3 and m_2, respectively, out of the parenthesis, to get the insides in fraction form:

$$T^2 = \frac{4\pi^2 r_2^3(\frac{r_1}{r_2}+1)^3}{Gm_2(\frac{m_1}{m_2}+1)}$$

Next, we write this in terms of x:

$$T^2 = \frac{4\pi^2 r_2^3(x+1)^3}{Gm_2(\frac{1}{x}+1)}$$

Finally, we isolate x on the right hand side:

$$\frac{Gm_2T^2}{4\pi^2 r_2^3} = \frac{(x+1)^3}{\frac{1}{x}+1}$$

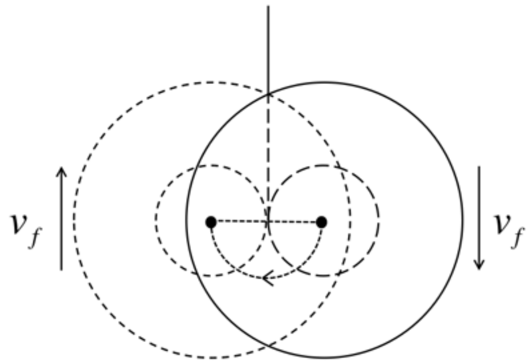
We can now approximate this function. We know everything on the left-hand side: $m_2 = 30M_{Sun} = 30 \times 2 \times 10^{30}$ kg, $T = 5.6$ days times 86400 seconds and $r_2 = 1.1397 \times 10^{10}$ m. The left-hand side is approximately equal to 16.04.

We can then plot the two functions

$$y = 16.04$$
$$y = \frac{(x+1)^3}{\frac{1}{x}+1}$$

and see where they intersect. That happens at approximately $x = 1.9031$.

With that value in hand, we can now find $r_1 = xr_2 = 2.169 \times 10^{10}$ m and $m_1 = m_2/x = 15.764$ solar masses.



(44)

Let's see. First, we can write an equation for the acceleration of the center of mass, in terms of string tension acting upwards, and gravity acting downwards. We choose downwards to be the positive direction, and find

$$ma = mg - T$$

Next, we can consider the torque. I will do so considering the center of the yo-yo, call it point C:

$$\tau_C = I_C \alpha$$

The torque is due to the tension acting on the inner spool of radius b, and is $\tau_C = Tb$.
We can also use the relationship $a = \alpha R$, which holds if there is no slipping. With these two things in mind, we can rewrite the above equation as

$$Tb = I_C \frac{a}{b}$$

We can solve for the tension by solving these for A and setting them equal.

$$\begin{aligned} a &= g - \frac{T}{m} \\ a &= \frac{Tb^2}{I_C} \end{aligned}$$

$$\begin{aligned} g - \frac{T}{m} &= \frac{Tb^2}{I_C} \\ g &= T\left(\frac{b^2}{I_C} + \frac{1}{m}\right) \\ \frac{g}{\frac{b^2}{I_C} + \frac{1}{m}} &= T \end{aligned}$$

We are given that $I_C = \frac{1}{2}mR^2$, so we can stick that in there and simplify to find the tension in terms of the variables they want:

$$\begin{aligned} \frac{g}{\frac{b^2}{(1/2)mR^2} + \frac{1}{m}} &= T \\ \frac{mg}{2\frac{b^2}{R^2} + 1} &= T \\ \frac{mgR^2}{2b^2 + R^2} &= T \end{aligned}$$

Next, they want the angular speed when the Yo-Yo reaches the bottom.
Now, we have a situation equivalent to pure roll, which means that the tangential velocity is always equal to the velocity of the center of mass.
We can therefore solve this more easily (I believe it's easier, anyway) by using a, using that to find the velocity of the center of mass, which then is equal to the tangential velocity, and converting that to angular speed.
We have an expression for the acceleration as a function of T, and we know T, so

$$a = g\left(1 - \frac{R^2}{2b^2 + R^2}\right)$$

Since the acceleration is clearly constant in time, the velocity as a function of acceleration is just $v = at$, but we don't know t.
We can use a second constant acceleration kinematics equation, though: $\ell = \frac{1}{2}at^2$. We solve that one for t:

$$\ell = \frac{1}{2}at^2$$
$$\sqrt{\frac{2\ell}{a}} = t$$

Combining the two,

$$v_f = a\sqrt{\frac{2\ell}{a}} = \sqrt{2\ell a} = \sqrt{2\ell}\sqrt{g\left(1 - \frac{R^2}{2b^2 + R^2}\right)}$$

Finally, to convert to angular speed, we simply use $v_f = \omega b$, so $\omega = \frac{v_f}{b}$:

$$\omega = \frac{\sqrt{2\ell}}{b}\sqrt{g\left(1 - \frac{R^2}{2b^2 + R^2}\right)}$$

This can be simplified quite a bit further (I used Mathematica for this one):

$$\omega = 2\sqrt{\frac{g\ell}{2b^2 + R^2}}$$

We should be able to solve the last part in terms of impulse. If the speed v_f going back up is the same as the speed down, as the diagram shows, the impulse is $2mv_f$.
The average force acting on the yo-yo is found via

$$I = \langle F \rangle \Delta t$$

However, the average force and the average tension are not the same thing. Regardless of the tension, there is clearly a constant downwards force mg acting on the yo-yo, due to gravity. Let's take care of that last.

$$\frac{2mv_f}{\Delta t} = \langle F \rangle$$

Of course, this causes a new problem: what is Δt? We know the speed v_f just prior to and just after, but what about during this turnaround?
Because the angular velocity is about the same for the entire turnaround (it doesn't switch directions), v_f is also approximately constant, since the two are linearly proportional.
In that case, $\Delta t = d/v_f$, where d is the distance traveled during this time. So what is ***that***, then? I would think it is half the circumference of the inner spool, which is πb. We can then

find the time as the distance divided by the tangential velocity, $\Delta t = (\pi b)/v_f$, so using that, plus our expression of the velocity v_f as the string is unwrapped:

$$\frac{2mv_f^2}{\pi b} = \langle F \rangle$$
$$\frac{2m}{\pi b} 2\ell g \left(1 - \frac{R^2}{2b^2 + R^2}\right) = \langle F \rangle$$

(Side note added afterwards: we can just as easily, probably more easily, consider that it moves π radians about the inner spool, and use ω at the turnaround point to calculate the time taken.)

Let's now not forget that $\langle F \rangle$ is the average ***net*** force on the object. Gravity is pulling it down, which the tension is trying to counteract. Therefore, we ***add*** mg, $braF_{gravity}\rangle$ (which is thankfully a constant) to the above to find the average tension:

$$mg + \frac{2m}{\pi b} 2\ell g \left(1 - \frac{R^2}{2b^2 + R^2}\right) = \langle T_r \rangle$$
$$mg + \frac{4m\ell g}{\pi b} \left(1 - \frac{R^2}{2b^2 + R^2}\right) = \langle T_r \rangle$$

Again, this can be simplified quite a bit, and again, I used Mathematica for that part:

$$\langle T_r \rangle = mg \left(\frac{8b\ell}{2\pi b^2 + \pi R^2} + 1\right) + m$$

(45) All right, time to look at some forces, to begin with!

Block 2 has four forces acting on it: mg downwards, $N = mg$ upwards (since there is no acceleration along the y axis, they must cancel), a tension T_2 towards the right, and friction $F_f = \mu_k m_2 g$ towards the left.
Block 1 has only two: mg downwards, and T_1 upwards.

Newton's second law for the two gives us, taking downwards (block 1) = rightwards (block 2) as positive:

$$m_2 a = T_2 - \mu_k m_2 g$$
$$m_1 a = m_1 g - T_1$$

Next, we can consider the torque and angular acceleration of the pulley. Relative to the center C of the pulley, the torque is $I_c\alpha$. As usual, we use $a = \alpha R$ to rewrite this in terms of the linear acceleration a, and assume there is no slipping or such going on.

$$\tau_C = I_c \frac{a}{R}$$

So what is the torque? Well, we can write it as the torque due to T_1 (which causes clockwise rotation) minus the torque due to T_2. Both act at 90 degree angles with the center, so

$$R(T_1 - T_2) = I_c \frac{a}{R}$$

We now have three equations and three unknowns: a, T_1 and T_2. If we solve the tension equations for T_2 and T_1 respectively, we can find $T_1 - T_2$ easily, and therefore a.
First, I will solve the above equation for a:

$$a = \frac{R^2}{I_c}(T_1 - T_2)$$

Solving the two is also easy:

$$\begin{aligned} m_2 a + \mu_k m_2 g &= T_2 \\ m_1 g - m_1 a &= T_1 \end{aligned}$$

All that remains is to combine the three as mentioned, and solve for a:

$$\begin{aligned} a &= \frac{R^2}{I_c}(m_1 g - m_1 a - m_2 a - \mu_k m_2 g) \\ a &= \frac{R^2}{I_c}(m_1 g - \mu_k m_2 g) - \frac{R^2}{I_c} m_1 a - \frac{R^2}{I_c} m_2 a \\ a\left(1 + \frac{R^2 m_1}{I_c} + \frac{R^2 m_2}{I_c}\right) &= \frac{R^2}{I_c}(m_1 g - \mu_k m_2 g) \\ a &= \frac{\frac{R^2}{I_c}(m_1 g - \mu_k m_2 g)}{1 + \frac{R^2 m_1}{I_c} + \frac{R^2 m_2}{I_c}} \\ a &= \frac{g R^2 (m_1 - \mu_k m_2)}{I_c + R^2(m_1 + m_2)} \end{aligned}$$

Well then! Let's see about part (b).
Is the acceleration constant? Yes, it is; nothing in there should change over time. Therefore, we can answer this one using some very basic kinematics:

$$d = \frac{1}{2} a t_1^2 = \frac{1}{2} \frac{g R^2 (m_1 - \mu_k m_2)}{I_c + R^2(m_1 + m_2)} t_1^2$$

(46) They then ask for a, T_1 (tension at m_1) and T_2 (tension at m_2).
This certainly looks like the slightly more complex brother of the previous problem!

To begin with, we can't know which of the masses will "win", if any. If static friction wins, then $a = 0$, which is the trivial solution and one that I will not even attempt to submit. What happens otherwise? Well, getting the sign correct is guesswork, as far as I can tell; according to forum discussions, this seems to be the consensus. I will call downwards (for m_1) and uphill (for m_2) positive in this solution.

Well then! We yet again have a bunch of forces. The forces on the hanging mass are unchanged, so we get the same equation there:

$$m_1 a = m_1 g - T_1$$

Block 2 changes the game a little. We have the same four forces, but we now need to decompose the gravitational force into the normal force component and the "downhill" component. Performing the decomposition, we find the normal force as $m_2 g \cos\theta$, which the downhill force is $m_2 g \sin\theta$. The incline is frictionless, so gravity is the only downhill force. There is a tension T_2 uphill, however, All in all, the normal force cancels out the component of gravity perpendicular to the incline, while T_2 uphill and $m_2 g \sin\theta$ battles where the block should move.
Using the directions I chose,

$$m_2 a = T_2 - m_2 g \sin\theta$$

Two equations, three unknowns. We need to consider the pulley next, as usual in these problems. As in the previous problem, the tensions cause a torque, and both are perpendicular to the center of the wheel. The torque relative to the pulley's center is $\tau_C = R(T_1 - T_2)$, which again is equal to $I_{cm}\alpha = I_{cm}\frac{a}{R}$. We were given I_{cm} in terms of mass and radius:

$$\begin{aligned} R(T_1 - T_2) &= \frac{1}{2} m_p R a \\ \frac{2(T_1 - T_2)}{m_p} &= a \end{aligned}$$

I will again find $T_1 - T_2$ by solving those two equations individually and subtracting them:

$$\begin{aligned} m_1 g - m_1 a &= T_1 \\ m_2 a + m_2 g \sin\theta &= T_2 \end{aligned}$$

$$T_1 - T_2 = m_1 g - m_1 a - m_2 a - m_2 g \sin\theta$$

Substitute that in to the torque equation:

$$\begin{aligned} a &= \frac{2}{m_p}\left(m_1 g - m_1 a - m_2 a - m_2 g \sin\theta\right) \\ a &= \frac{2}{m_p}\left(m_1 g - m_2 g \sin\theta\right) - \frac{2m_1}{m_p} a - \frac{2m_2}{m_p} a \\ a\left(1 + \frac{2m_1}{m_p} + \frac{2m_2}{m_p}\right) &= \frac{2}{m_p}\left(m_1 g - m_2 g \sin\theta\right) \\ a &= \frac{\frac{2}{m_p}\left(m_1 g - m_2 g \sin\theta\right)}{1 + \frac{2m_1}{m_p} + \frac{2m_2}{m_p}} \\ a &= \frac{2g\left(m_1 - m_2 \sin\theta\right)}{m_p + 2m_1 + 2m_2} \end{aligned}$$

We can then find the tensions easily, since we solved for them earlier. Whether it will be pretty is a different matter!

$$T_1 = m_1 \left(g - \frac{2g\,(m_1 - m_2 \sin\theta)}{m_p + 2m_1 + 2m_2} \right)$$

$$T_2 = m_2 \left(g \sin\theta + \frac{2g\,(m_1 - m_2 \sin\theta)}{m_p + 2m_1 + 2m_2} \right)$$

(47) Okay, so let's see. First, what happens with zero friction? Clearly, there is no rolling at all, since there will be no torque on the cylinder.

For there to be pure roll, a condition that must be fulfilled is that the tangential speed ωR is the same as the velocity at the center of mass. Since the angular acceleration is $\alpha = a/R$, we must have $\alpha R = a$ for pure roll to hold.

Let's start out with part (a) and see where that leads.

If there is is no slipping, then there is no kinetic friction. There is, however, ***static*** friction. Without that, the cylinder would slide down without turning at all.

If we choose a coordinate system where y is perpendicular to the incline, while x is downhill, we can write the the normal component of gravity as $Mg\cos\theta$, while the downhill component is $Mg\sin\theta$. Static friction acts upwards: the friction must be such that the torque causes clockwise rotation (or it would roll the wrong way!). This implies an uphill frictional force, $\mu_s N = \mu_s Mg\cos\theta$. (Another way to think of it is that the cylinder wants to slide downhill. Therefore, static friction acts uphill, since friction always ***opposes*** relative motion between surfaces.)

$$Ma = Mg\sin\theta - \mu_s Mg\cos\theta$$
$$a = g(\sin\theta - \mu_s\cos\theta)$$

This would seem to answer part (a), but we're not allowed to use μ_s in the answer, so we need to keep working.

As mentioned earlier, in this analysis, the cylinder rolls due to the torque caused by friction. Friction acts uphill, and the magnitude of the torque, relative to the center of the cylinder, is $RF_f = RMg\mu_s\cos\theta$.

A useful relation is then that $\tau = I_{cm}\alpha = I_{cm}\dfrac{a}{R}$ (the latter part holds for pure roll only), and we are given that $I_{cm} = MR^2$, so

$$RMg\mu_s\cos\theta = aMR$$

M and R both cancel.

We can solve this for μ_s:

$$g\mu_s\cos\theta = a$$
$$\mu_s = \frac{a}{g\cos\theta}$$

This gives us the acceleration, now that we can write μ_s in terms of g and $\cos\theta$:

$$
\begin{aligned}
a &= g(\sin\theta - \frac{a}{g\cos\theta}\cos\theta) \\
a &= g\sin\theta - a \\
a &= \frac{g\sin\theta}{2}
\end{aligned}
$$

Next, we substitute this back into μ_s to get it in terms of θ:

$$
\begin{aligned}
\mu_s &= \frac{\frac{g\sin\theta}{2}}{g\cos\theta} \\
\mu_s &= \frac{\tan\theta}{2}
\end{aligned}
$$

Very nice and simple! This is the ***minimum*** amount of friction required for pure roll. More friction wouldn't hurt; as the acceleration equation shows, more friction doesn't cause less acceleration, but it does prevent sliding.

Finally, what is the velocity of the center of mass as it reaches the bottom? Well, we know the acceleration. We could use the work-energy theorem, but there will be both linear kinetic energy and rotational kinetic energy, so that seems like it would be harder. Then again, we don't know the ***time*** which we need for kinematics, so I will go the energy route anyway.
The final velocity v causes an angular velocity $v = \omega/R$ with no sliding. The total kinetic energy can be written down as being equal to Mgh, which is the total energy available to be converted:

$$
\begin{aligned}
K_{lin} + K_{rot} &= Mgh \\
\frac{1}{2}Mv^2 + \frac{1}{2}I_{cm}\omega^2 &= Mgh \\
\frac{1}{2}Mv^2 + \frac{1}{2}(MR^2)\left(\frac{v}{R}\right)^2 &= Mgh \\
\frac{1}{2}Mv^2 + \frac{1}{2}Mv^2 &= Mgh \\
v^2 &= gh \\
v &= \sqrt{gh}
\end{aligned}
$$

Nice! The intermediate results were semi-complex at times, but the answers are all dead simple.

There are several interesting things in this result, at least two of which I didn't realize until a few days after solving this. One is that the rotational kinetic energy in this case is exactly equal to the linear kinetic energy – the expression on the right in equation 4 above simplifies to $\frac{1}{2}Mv^2$! Without this term, the velocity would be $\sqrt{2gh}$ instead, i.e. exactly a factor $\sqrt{2}$ greater, regardless of much of anything else.
(I rewrote that equation after realizing this; I previously had it in a form which made this hard to see.)

Second, I chose to analyze this relative to the center, which means that static friction provides a torque. How can there be an increase in rotational kinetic energy without a torque that does positive work? As far as I know, there certainly cannot. Therefore, according to this analysis, static friction provides this positive work!

However, it still does no net work, which is the amazing thing: it is a linear force uphill, which therefore fights with $Mg\sin\theta$ about the linear acceleration. It therefore acts to reduce the final linear velocity by a factor $1/\sqrt{2}$, and the final kinetic energy by a factor $1/2$; this is instead turned into rotational kinetic energy here.

So while static friction appears to does positive work increasing the rotational kinetic energy, it appears to do an equal amount of negative work in the linear motion, for a net of zero work – as it must be.

If we instead analyze this problem relative to the point of contact, friction can provide no torque (as it acts through that point), and we will instead find gravity providing the torque and therefore doing the work that gets the cylinder rolling. For other points, where both forces can cause a torque relative to the center, we should find some combination of the two effects, but with the same end result.

(48) I wonder how realistic the answers will be – a piece of debris with negligible speed (relative to the Earth) wouldn't stay in place for very long!

The satellite begins with linear momentum mv_0. After the hit, the mass doubles, and so velocity is cut in half. Call this post-hit velocity v_a (a for apogee); using conservation of (linear) momentum, we then have $mv_0 + 0 = 2mv_a$, so indeed $v_a = \dfrac{v_0}{2}$.

We could also find this relationship using conservation of angular momentum relative to the center of the Earth, by the way.

Given this "initial" velocity v_a and the initial distance to the Earth, we could find the orbital parameters for the new elliptical orbit, but I don't believe we will need all of them. Clearly, the apogee distance r_a is simply the initial radius of the circular orbit, which is even given in the problem, only they don't mention it explicitly, but use the same variable for the two (and draw the graphic showing the two are equal).

Now, then. How can we calculate v_0? Well, we know the orbital radius, and for a circular orbit, each orbital radius has unique velocity. This velocity can be derived by remembering that the total mechanical energy is always $\frac{1}{2}U$, but I'm confident that I remember the quite simple velocity equation, so:

$$v_0 = \sqrt{\frac{MG}{r_a}}$$

$r_a + r_p = 2a$, where a is the elliptical orbit's semi-major axis. Via an energy calculation, we can relate the new velocity v_a plus the current potential energy with the total mechanical energy for an elliptical orbit, which depends on $2a$, so we can find a.

$$\begin{aligned}
\frac{1}{2}(2m)v_a^2 - \frac{2mMG}{r_a} &= -\frac{2mMG}{2a} \\
v_a^2 - \frac{2MG}{r_a} &= -\frac{2MG}{2a} \\
\frac{1}{v_a^2 - \frac{2MG}{r_a}} &= -\frac{a}{MG} \\
\frac{MG}{\frac{2MG}{r_a} - v_a^2} &= a \\
\frac{r_a MG}{2MG - r_a v_a^2} &= a
\end{aligned}$$

Since we know that $r_a + r_p = 2a$, the above must be equal to $\frac{r_a + r_p}{2}$. We can also substitute in the value for $v_a = v_0/2 = \frac{1}{2}\sqrt{\frac{MG}{r_a}}$:

$$\begin{aligned}
\frac{r_a MG}{2MG - r_a \frac{MG}{4r_a}} &= \frac{r_a + r_p}{2} \\
\frac{4r_a MG}{8MG - MG} &= \frac{r_a + r_p}{2} \\
\frac{8r_a}{7} - r_a &= r_p \\
\frac{r_a}{7} &= r_p
\end{aligned}$$

Well, that sure became simple. The ratio is then $r_a/r_p = 7$ – showing that the apogee is at a (much) greater than distance than the perigee, as one would expect.

And that's it for this problem!

(49) Time for a short break to read the textbook! I'm unsure whether we can use $\Omega = \dfrac{\tau}{I_c\omega}$ here (after vector decomposition), or not.
They indeed seem to consider that we can ignore any angular momentum due to the orbital motion, and therefore, this approximation should be valid. Very well, then.
My solution will be less rigorous than the quite technical discussion in the textbook; if you want more detail, I recommend having a look there. Actually, I would recommend that either way!

The torque relative to what I will call point P, the pivot point where the axle meets the stand, is $\tau_P = \mathbf{d} \times \mathbf{F_{gr}} = (\mathbf{d} \times \mathbf{g})M$. Unlike what we have seen previously, the angle is not 90 degrees. Gravity is always straight downwards, of course, but as the angle the axle makes with the horizontal grows (downwards), the torque goes down. It is at a maximum with $\theta = 0$, and zero when the axle is pointing straight down (which makes sense: the two vectors are then anti-parallel, so the cross product must be zero). The equation then becomes

$$\tau_p = \mathbf{d} \times \mathbf{F_g} = dMg\cos\theta$$

(where θ is the angle that is marked as 30 degrees).

Why a cosine, in a cross product? Because the angle ***between the two vectors*** is not equal to the 30° degrees shown, but instead is $90° - 30°$ degrees. It makes intuitive sense that when the

angle shown is zero, the torque is at a maximum, and when the axle is vertical, there is zero torque.
We could write the cross product as $dMG \sin\alpha$, where α is the angle between the vectors, followed by $\alpha + \theta = 90°$. This then makes it clear that we need $\sin\alpha = \sin(90° - \theta) = \cos\theta$. I will write it in terms of the cosine of θ, since that gives us a simple expression in terms of the given variables ($\theta = 30°$).

The spin angular momentum due to the disk spinning about its center of mass can be written as $I_c\omega$, where $I_c = \frac{1}{2}MR^2$ for a solid disk.
The direction of this is "inwards" along the axle, no matter the axle's angle; so radially inwards and partially upwards, in this case.

We now know torque and the spin angular momentum. The spin angular momentum needs to be decomposed, though, as only the radial portion matters for the precession.
Consider the time when the system has rotated such that the view from the angle the figure is shown is now such that the axle is in the plane of the page, and we see the disk head-on, on the right side of the pivot point.
The torque is then pointed into the page, while spin angular momentum points left/upwards, at an angle with the horizontal due to the non-horizontal axle.

Left/upwards in more mathematical terms would mean $-\hat{r}$ (left) and $+\hat{k}$ (upwards), using cylindrical coordinates, where $+\hat{\theta}$ is into the page.

As the disk/gyroscope precesses, only the direction of the radial component changes, with the center of mass of the disk tracing out a circle in a horizontal plane. The angle, and therefore the upwards/z component does not change as long as ω (the disk's spin angular velocity) is held constant. Neither does the magnitude of the spin angular momentum change; the only change in its direction, as mentioned.
The time derivative of $\hat{r}$ is given as $\frac{d\theta}{dt}\hat{\theta} = \Omega\hat{\theta}$, i.e. into the page. However, if we treat this more rigorously, we will find that Ω is negative, and so the system will move "towards us" as seen here (clockwise as seen from above).

For a more rigorous treatment, see chapter 22 in the textbook (the end of page 22-14 and onwards).

All in all, we have

$$\begin{aligned} |\Omega| = \frac{|\tau|}{|L_{spin}|} &= \frac{dMg\cos\theta}{0.5MR^2\omega\cos\theta} \\ &= \frac{2dg}{R^2\omega} \end{aligned}$$

For these values, $\Omega = \frac{1}{3}$ rad/s (using $g = 10\,\mathrm{m/s^2}$, which it appears we are supposed to), which is accepted as correct!
I was a bit worried when the cosines cancelled out, as I expected the angle to matter. Apparently, the effect is indeed cancelled out, as both the torque component and the spin angular momentum component that matter are smaller (by the same factor).

(50) Having read the section in the book (chapter 22) on exactly this problem, I feel like I'm cheating here! I will do what I can to derive everything I use, in order to ensure I understand it all, at least.

All right. The first part is rather easy, at least: the center of mass of the wheel must move with speed $v_{cm} = \omega b$ if there is no slip (this is a condition of pure roll). Meanwhile, the entire wheel

is also rotating about the center axis with angular speed Ω, which can be used to find $v_{cm} = \Omega R$ separately from the previous relationship.
We can then simply set the two equal and solve for ω, since Ω is allowed in the answer:

$$\omega b = \Omega R$$
$$\omega = \frac{\Omega R}{b}$$

Next, the horizontal component of the angular momentum relative to point P. Given the hint in the problem, this is very easy. The angular momentum about the axle's axis due to the rotation (about the wheel's center of mass) is just $I_c\omega$, where we use $I_c = \frac{1}{2}Mb^2$ for a solid disk of radius b (not R in this problem!):

$$L_P = I_c\omega = (\frac{1}{2}Mb^2)\left(\frac{\Omega R}{b}\right) = \frac{\Omega RMb}{2}$$

Part (c) is regarding the magnitude of the torque about the center axle (point P is not in the figure, but it is in the book; it is where the axle connects to the vertical bar, at the hinge).

Well, what forces could cause a torque? Gravity acting on the wheel certainly counts; the torque at P (see above) due to gravity acting on the wheel is $\tau_{P,gravity} = (\mathbf{R} \times \mathbf{g})M = RMg$ (there is a 90 degree angle, so $\sin\theta = 1$), with the direction being into the page (causing rotation as shown for Ω).
Next, there is the normal force $N = 2Mg$ causing a torque $\mathbf{R} \times \mathbf{N} = RN = 2RMg$, with the direction being out of the page, opposing the previous torque.
If the axle is taken to be massless, there are no other forces that act such that they cause a torque relative to point P.

The net torque, or at least the magnitude of it, is just the torque due to the normal force minus the torque due to gravity:

$$|\tau_{P,net}| = RN - RMg$$

We could write this in terms of $2Mg$ instead of N, but the grader really wants it in terms of N, according to the forum discussions. I submitted the above as the first attempt, and it was indeed accepted.

Finally, what is Ω, in terms of only b and g, at a time where $N = 2Mg$?
This is the precession frequency – note how the system looks a lot like a gyroscope. (It's even as an example in the gyroscope section in the book.)

We learned in lecture that $\Omega = \dfrac{\tau}{L_{spin,cm}}$, but this only holds if $L_{spin,cm} \gg L_{orbital}$, which doesn't appear to be the case here. In the case of a typical gyroscope, the spin could be several thousand rpm (200π rad/s or more), while the orbits was closer to 5 per minute or even less.
Here, the two are much closer together.

We can solve this in (at least?) two ways. One is, in fact, to use the above equation:

$$\Omega = \frac{2RMg - RMg}{\frac{1}{2}\Omega RMb}$$

$$\Omega^2 = \frac{g}{\frac{1}{2}b}$$

$$\Omega = \sqrt{\frac{2g}{b}}$$

The second is to find the torque as $\frac{dL}{dt}$ (i.e. take the time derivative of L above) and set that equal to the torque we found earlier. However, to do this properly, we need to consider the directions properly too. Check the book (chapter 22) for a proper derivation. The result is:

$$\frac{dL}{dt} = \frac{\Omega^2 RMb}{2} = RMg$$

$$\Omega^2 b = 2g$$

$$\Omega = \sqrt{\frac{2g}{b}}$$

The source of the extra Ω is tricky, since I have not written all this in terms of components and unit vectors. The source of it is due to the differentiation of the $\hat{r}$ unit vector:

$$\frac{d\hat{r}}{dt} = \frac{d\theta}{dt}\hat{\theta}$$

where θ is the position along the circle, and $\hat{\theta}$ is the unit vector in the azimuthal direction (in cylindrical coordinates). Ω is just the time rate of change of this angle, by definition, so that $\Omega = \frac{d\theta}{dt}$. Therefore, in terms of vectors,

$$\mathbf{L}_{spin,cm} = \frac{\Omega RMb}{2}(-\hat{r})$$

$$\frac{d\mathbf{L}_{spin,cm}}{dt} = \frac{\Omega RMb}{2}\left(-\frac{d\hat{r}}{dt}\right) = \frac{\Omega RMb}{2}\left(-\Omega\hat{\theta}\right)$$

The magnitude is therefore multiplied by Ω in this differentiation. I apologize for the sloppiness here; again, check the book if you're looking for a rigorous treatment of this problem.

(51) Because drum B (the one at the top) is free to rotate, this problem is not quite as easy as it might look to begin with. We must assume that it too rotates, and that the tape is unrolled from ***both*** drums at the same time.

Okay then, let's see. First, let's consider the linear acceleration of drum A, which will certainly give us more than one unknown. Using downwards as the positive direction,

$$Ma = Mg - T$$

The string (tape?) will unroll, which means we can also consider the angular acceleration, due to the torque provided by this tension. The torque relative to the center of drum A $\tau_A = I_c\alpha$, which is also simply $\mathbf{R} \times \mathbf{T}$, where $\mathbf{R}$ is the position vector from the center (since we take that as our origin for the torque) to the edge of drum A.

$$RT = \left(\frac{1}{2}MR^2\right)\alpha_A$$

Here is where we must be very careful. We can ***not*** use $a = \alpha R$ here! That holds when the drum unrolls such that 100% of the added length of tape comes from the drum – but both drums are unrolling at the same time! In other words, we don't have pure roll in this situation. Instead, we must consider the torque and angular acceleration of drum B. Since both radius, mass and tension are all the same, we find

$$RT = \left(\frac{1}{2}MR^2\right)\alpha_B$$

By comparing these two last equations, we don't even need to solve either so find $\alpha_A = \alpha_B$; everything except those variable names are the same in both equations.

Finally, we can consider the position (and change in position) considering how much tape is unrolled. Following the book's approach, an amount $R\Delta\theta_A$ is unrolled from the first drum in some time Δt, and the same thing except with a B index holds for drum B. The distance fallen for drum A is the sum of the two, i.e. the total amount of tape unwound. If we take the time derivative of these expressions, we get

$$\frac{dy}{dt} = R\frac{d\theta_A}{dt} + R\frac{d\theta_B}{dt}$$

... and again:

$$\frac{d^2y}{dt^2} = a = R(\alpha_A + \alpha_B)$$

The values of $\alpha_A = \alpha_B$ in terms of the tension is

$$\frac{2T}{MR} = \alpha_A = \alpha_B$$

And using the first equation we found, $T = M(g - a)$, so

$$\alpha_A = \alpha_B = \frac{2}{MR}M(g - a)$$
$$\alpha_A = \alpha_B = \frac{2}{R}(g - a)$$

So at this stage, we have two equations:

$$\alpha_A = \alpha_B = \frac{2}{R}(g - a)$$
$$a = R(\alpha_A + \alpha_B)$$

Substitute the top one into the lower one:

$$a = 2R(\frac{2}{R}(g - a))$$
$$a = 4g - 4a$$
$$a = \frac{4}{5}g$$

The acceleration is higher than the $\frac{2}{3}g$ we find if the top drum cannot spin, and we therefore assume pure roll.

(52) Let's first consider the vertical forces on the beam. We have three weights, balanced by the same tension in two places; the tensions need to be decomposed, though. If the angle was 90 degrees, the vertical component of the tension would clearly be at a maximum, so we need a sine in there (which drawing it out and doing the trigonometry confirms):

$$g(m_A + m_B + m_C) = 2T \sin\theta$$

We only need to divide both sides by $2\sin\theta$, and we have the answer to part (a):

$$\frac{g(m_A + m_B + m_C)}{2\sin\theta} = T$$

For part (b), we need to consider the torque on the system. We can calculate torques relative to any point of our choosing, but what point would make things the easiest? If we choose $x = 0$, the torque due to mass C disappears. The same argument holds for other points and other masses. Just below the cable, between the two tensions, the torques due to both tensions cancel out.

Because the answer doesn't allow θ and doesn't allow g, we should choose the point where the tensions cause no torque. That way, all disallowed variables should either not enter the equation (θ) or cancel (g).

I will call that point $b = \ell_1 + (\ell_2 - \ell_1)/2 = \dfrac{\ell_1 + \ell_2}{2}$, to reduce clutter in the torque equation. I use out of the screen as the positive direction.

$$\tau_b = bgm_C - (x - b)gm_B - (y - b)gm_A$$

This must be equal to zero. g cancels, as hoped for/expected.

$$
\begin{aligned}
0 &= bm_C - (x-b)m_B - (y-b)m_A \\
xm_B &= bm_C + bm_B - ym_A + bm_A \\
x &= \frac{b(m_C + m_B + m_A) - ym_A}{m_B} \\
x &= \frac{\frac{\ell_1+\ell_2}{2}(m_C + m_B + m_A) - ym_A}{m_B}
\end{aligned}
$$

We would, in the end, find the same answer if we calculated the torque relative to any other point.

(53) There are four forces on the beam (with 1 or 2 components each): normal force (2 components) at the hinge, gravity acting purely downwards at the center of mass ($L/2$), gravity acting purely downwards at d and the tension (2 components) at the end of the beam.

The tension clearly acts upwards and inwards, so the normal force must act outwards (towards the right), as they are the only two horizontal forces. Whether the normal force acts upwards or downwards I don't know however, since there is also gravity in the mix. I will guess that it acts upwards, and so if it turns out negative, I guessed wrong.

For the tensions, we have

$$
\begin{aligned}
T_x &= -T\cos\theta \\
T_y &= T\sin\theta
\end{aligned}
$$

using a coordinate system where $+x$ is towards the right. We can now calculate the sum the forces in the vertical direction to zero:

$$
N_y + T\sin\theta - g(m_1 + m_2) = 0
$$

One equation, two unknowns. Next, we can consider torque. The net torque relative to any point must be zero. If we choose the point right at the hinge, the unknown normal force doesn't cause a torque, so we get

$$
g\left(\frac{L}{2}m_1 + dm_2\right) - LT\sin\theta = 0
$$

The horizontal forces also cannot cause at torque relative to this point. We now have two equations and two unknowns, though we also need to find N_x later on. That turns out to be trivial, however, so let's begin with T and N_y.

Note that T is the only unknown in this second equation, so we start by finding that:

$$
\begin{aligned}
g\left(\frac{L}{2}m_1 + dm_2\right) &= LT\sin\theta \\
\frac{g\left(\frac{L}{2}m_1 + dm_2\right)}{L\sin\theta} &= T
\end{aligned}
$$

For the given values, $T = 1691.5$ newton. We can then find N_y by solving the previous equation for that, and sticking in this value of T.

$$N_y = g(m_1 + m_2) - T\sin\theta$$
$$N_y = g(m_1 + m_2) - \frac{g\left(\frac{L}{2}m_1 + dm_2\right)}{L}$$

For the given values, $N_y = 1087.8$ newton.

As for N_x, it and T_x are the only two horizontal forces. Therefore, they must be equal in magnitude, and so $N_x = T\cos\theta = 1385.6$ N.

(54) The vertical forces consist of the normal force where the ladder touches the ground (I call this point Q), gravity due to the person at $d/3$ along the length, and gravity at the ladder's center of mass $d/2$ along the length. Therefore,

$$N_Q = g(m_1 + m_2)$$

In the horizontal direction, we have the normal force from the wall (point P) N_P towards the left, and a frictional force $f_s \le \mu N_Q$ at point Q towards the right (since the ladder wants to slip towards the left).
This gives us, just at the edge of slipping ($f_s = \mu_s N_Q$, i.e. the maximum friction possible):

$$N_P = f_s = \mu_s N_Q$$

Next, we can consider the torque. I will calculate them relative to point Q, so that two out of the five forces/force components "disappear" (they can't cause torque through that point). I will use into the screen (clockwise rotation) as positive, since that is how the ladder wants to rotate.

Now, these cross products depend on the angle, but the angle between the position vector from Q to where gravity acts, and the gravitational force vector, is not θ. Indeed, it's easy to see that if $\theta = 0$, the angle between them would be 90 degrees. The relevant angle is $90^\circ - \theta$, so that is what we need for the cross products; also, $\sin(90^\circ - \theta) = \cos(\theta)$.
θ is the relevant angle for the normal force at P, however, so that one remains a sine.
Alternatively, we can try to find the perpendicular distance of either vector, and multiply that by the full magnitude of the other, which is the same thing.

$$\tau_Q = \frac{d}{3}m_2 g\cos\theta + \frac{d}{2}m_1 g\cos\theta - dN_P\sin\theta$$

This needs to be equal to zero. We can set it equal to zero, solve for N_P (which we earlier said was equal to f_s in magnitude) and find the answer for part (a):

$$0 = \frac{d}{3}m_2 g\cos\theta + \frac{d}{2}m_1 g\cos\theta - dN_P\sin\theta$$
$$N_P = \frac{\frac{d}{3}m_2 g\cos\theta + \frac{d}{2}m_1 g\cos\theta}{d\sin\theta}$$
$$f_s = N_P = g\cot\theta\left(\frac{m_2}{3} + \frac{m_1}{2}\right)$$

(Since $f_s = N_P$.)
All variables above are known, so we can calculate $f_s = 418.5$ N.

Next, we need to find μ_s. $f_s = \mu_s N_Q$, and we know N_Q to be the sum of the two weights, $g(m_1 + m_2)$.

$$\mu_s = \frac{1}{g(m_1+m_2)} g \cot\theta \left(\frac{m_2}{3} + \frac{m_1}{2}\right)$$
$$\mu_s = \frac{1}{m_1+m_2} \cot\theta \left(\frac{m_2}{3} + \frac{m_1}{2}\right)$$
$$\mu_s = \frac{cot\theta(2m_2+3m_1)}{6(m_1+m_2)}$$

In terms of numbers, $\mu_s \geq 0.427$ will meet this condition, so that there is no sliding.

Next, they want the magnitude and angle of the contact force. $N_Q = g(m_1+m_2) = 980$ N, and $f_s = 418.5N$. In terms of unit vectors,

$$C_{ladder,ground} = f_s\hat{x} + N_Q\hat{y}$$

The magnitude of this vector is $C_{ladder,ground} = \sqrt{418.5^2 + 980^2} = 1065.6$ N. The angle α must be less than 90 degrees, or the friction would point towards the left. It is found as $\alpha = \arctan \frac{N_Q}{f_s}$, which is about 1.167 radians, or 66.88 degrees.

(55) The problem description certainly sounds complex, but given the diagram and even a free body diagram, this should be one of the easier problems of the week. I choose a coordinate system with $x = 0$ and $y = 0$ at the elbow joint, with $+x$ to the right and $+y$ upwards (which I just noticed is marked in the free body diagram).

We need a net force of zero in the vertical direction, which gives us our first equation (equating upwards and downwards forces):

$$T = F + m_1g + m_2g$$

where m_2g is equal in magnitude to the normal force from the hand to the ball.

Next, the torques must be zero, relative to any point of our choosing. I choose the center of the coordinate system, so that F causes no torque. Downwards forces then cause a counterclockwise (into the screen) torque, which I denote as positive.

$$\tau = -dT + sm_1g + 2sm_2g$$

This must be equal to zero; we can set it as such and solve for T:

$$0 = -dT + sg(m_1 + 2m_2)$$
$$T = \frac{sg(m_1+2m_2)}{d}$$

This answers part (a); for part (b), we solve the force equation for F and substitute in T.

$$F = T - g(m_1 + m_2)$$
$$F = \frac{sg(m_1 + 2m_2)}{d} - g(m_1 + m_2)$$

Indeed quite easy compared to the previous ones.

(56) The possibly relevant values in the handout are (all values for steel, of course):
$Y = 20 \times 10^{10}\,\mathrm{N/m^2}$
Ultimate tensile strength $= 5.2 \times 10^8\,\mathrm{N/m^2}$
Density: $\rho = 8 \times 10^3\,\mathrm{kg/m^3}$

This problem is fairly similar to problem 9, which I solved prior to this one.

First, we need to calculate the tension at the center. The book has a derivation in chapter 9. The result is

$$T(r) = \frac{m\omega^2}{2L}(L^2 - r^2)$$
$$T(0) = \frac{1}{2}Lm\omega^2$$

as r is the distance from the center. (m is the total mass of the rod, while L is the length ***assuming we rotate it about its end***.)

We can write for the total mass $m = AL\rho$, where A is the unknown cross-sectional area of the stick. That gives us, for the tension at the center,

$$T(0) = \frac{1}{2}AL^2\rho\omega^2$$

The ultimate tensile stress is a pressure, $P_{ult} = F/A$. We need to multiply it by the cross-sectional area to get a force, that we can compare with the tension. We can then set the two equal and solve for ω.

$$\frac{1}{2}AL^2\rho\omega^2 = P_{ult}A$$

A cancels, and we can solve to find

$$\omega = \sqrt{\frac{2P_{ult}}{L^2\rho}}$$

However, in this equation, L is not the one meter length of the meter stick! It is half that: it is the length that sticks out from the center, and since we rotate the stick about its midpoint, we get half a meter for L. This then gives $\omega \approx 721$ rad/s, which is about 6900 rpm.

(57) The stick is leaning towards a wall on the left, and θ is measured between the vertical and the stick, so that it would be 0 if the stick was upright.

This problem is very similar to the one with the leaning ladder, only that there is now a frictional force along the wall also.
I will use the same naming scheme of point Q touching the ground (normal force N_Q) and point P touching the wall (normal force N_P). As for friction, I will use F_Q and F_P.
Aside from those four, there is only one force remaining: gravity, acting on the center of mass. Apparently, this must cancel out (the mass is not given), but I will call it m while solving.

The frictional force on the wall must be upwards, since the stick wants to slide down. The frictional force on the floor is towards the left, since the stick wants to slide to the right. I will use a standard coordinate system with $+x$ being towards the right and $+y$ being upwards.
The problem notes that the stick is just about to slide at the wall, so $F_P = \mu_s N_P$ holds there.

However, how could it slide at the wall without also sliding on the floor? It's a rigid stick; unless it goes off into the third dimension, it cannot slide at the wall while staying in place on the floor. Not only that, but this might just be a statically indeterminate problem if we don't consider it to be about to slip in both places at once. That is, if we don't assume that, we will have more unknowns than equations, and need extra information. We haven't learned about those in the course, so in short, I assume that it is about to slip in ***both*** places, so that also $F_Q = \mu_s N_Q$ holds, rather than the general case $F_Q \le \mu_s N_Q$ which doesn't help us a whole lot.

First off, we need the sum of forces in both directions to be zero. Starting with the vertical forces,

$$\begin{aligned} F_P + N_Q - mg &= 0 \\ \mu_s N_P + N_Q &= mg \end{aligned}$$

Next, the horizontal forces:

$$\begin{aligned} N_P - F_Q &= 0 \\ N_P &= \mu_s N_Q \end{aligned}$$

And finally, the torque, relative to point Q (or any other point, but I choose point Q), must be zero.
F_Q and N_Q act through this point, and cannot cause any torque relative to it. The torque due to gravity is $(\ell/2)mg\sin\theta$; the others are in the opposite direction, with $F_P = \mu_s N_P$ causing a torque $\ell\mu_s N_P \sin\theta$, and N_P causing a torque $\ell N_P \cos\theta$.

$$(\ell/2)mg\sin\theta - \ell N_P(\mu_s \sin\theta + \cos\theta) = 0$$

So, three equations, with μ_s, N_P and N_Q as unknowns. We only really care about μ_s, though. We can eliminate N_P using $N_P = \mu_s N_Q$, which leaves two equations and two unknowns:

$$\begin{aligned} (\ell/2)mg\sin\theta - \ell\mu_s N_Q(\mu_s \sin\theta + \cos\theta) &= 0 \\ \mu_s^2 N_Q + N_Q &= mg \end{aligned}$$

We can solve the second one for N_Q:

$$\begin{aligned}
\mu_s^2 N_Q + N_Q &= mg \\
N_Q(1+\mu_s^2) &= mg \\
N_Q &= \frac{mg}{1+\mu_s^2}
\end{aligned}$$

We can then combine the two equations; in the second equation below, mg cancels, ℓ cancels, and we can divide through by $\sin\theta$. The rest is just simplification to get it into a standard form for a quadratic:

$$\begin{aligned}
\frac{\ell mg}{2}\sin\theta - \frac{\ell \mu_s mg}{1+\mu_s^2}(\mu_s \sin\theta + \cos\theta) &= 0 \\
\frac{1}{2} - \frac{\mu_s}{1+\mu_s^2}(\mu_s + \cot\theta) &= 0 \\
\frac{1}{2} - \frac{\mu_s^2 + \mu_s\cot\theta}{1+\mu_s^2} &= 0 \\
\frac{1-\mu_s^2 - 2\mu_s\cot\theta}{2(1+\mu_s^2)} &= 0 \\
1-\mu_s^2 - 2\mu_s\cot\theta &= 0 \\
\mu_s^2 + 2\cot(\theta)\mu_s - 1 &= 0
\end{aligned}$$

Finally, after all that massaging, we can solve this for μ.

$$\begin{aligned}
\mu_s &= \frac{-2\cot\theta \pm \sqrt{4\cot^2\theta + 4}}{2} \\
\mu_s &= -\cot\theta \pm \sqrt{\cot^2\theta + 1} \\
\mu_s &= -\cot\theta + \frac{1}{\sin\theta}
\end{aligned}$$

Only the positive root gives a meaningful answer (the other one gives $\mu_s < 0$ which is unphysical). We can simplify this even one step further:

$$\mu_s = \tan\frac{\theta}{2}$$

Lots of work if you do the math manually (unless I missed some obvious simplifications), but the result is certainly very elegant!

Sidenote: this problem was graded incorrectly until November 27-28 (depending on timezones etc); the grader was set such that $\tan(\theta/2)$ was correct if you specified θ as the number given ***in degrees***, despite the calculator using radians. As such, the accepted μ_s was about 2.24(!) in my case, rather than the actually correct 0.36 or so that is now accepted.

(58) I will begin by assuming that there is no friction. That means that forces D, C and B are purely horizontal, and that force A is purely vertical. It also means that the middle ball must provide both an upwards and a rightwards force on the top ball.

Drawing this out (anything else might just be insanity; see partial drawing below), it's clear that $A = 3mg$, or there cannot be equilibrium, if it A is the only upwards force.

The distance between the center of the bottom ball and the center of the middle ball is exactly $2R$ (same for the middle and top balls).
The distance from the right side to the center of the bottom ball is R; the distance from the left side to the center of the middle ball is also R. Therefore, since the entire tube is $3R$, the horizontal distance between the two centers must also be R.

Using the Pythagorean theorem, the vertical distance between the centers must then be $\sqrt{3}$ times R (for both the top-middle and the middle-bottom balls).

So, forces... forces...
Consider the forces on the top ball. There is a force to the left, which cannot cause a torque relative to its center, since the angle between the position vector and the force vector would be 180 degrees.
Likewise, mg due to gravity cannot cause a torque, as it acts on the center.
This means that only the contact force due to the middle ball remains, which must therefore ***create no torque***, or the top ball would have a net torque! There is no other force that could possibly create an opposing torque and cancel it out.
The only way this can happen is if the net normal force is pointing straight towards the center of the top ball!

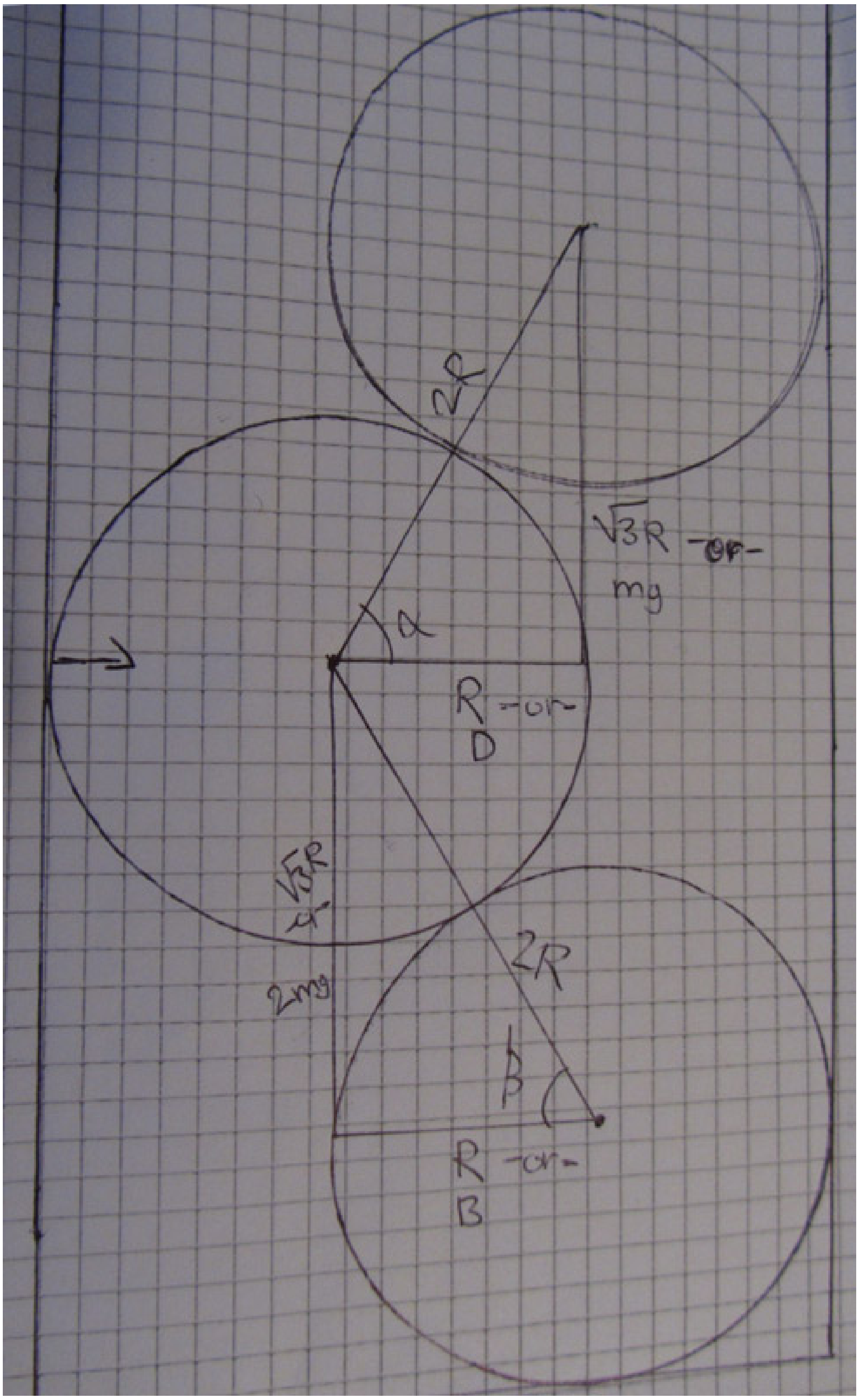
This then puts another constraint on the normal force, so we now know: it must be D in magnitude to the right (or there is a net horizontal force on the top ball), mg up (or there is a net downwards force on the top ball), ***and*** be at the correct angle, or there is a net torque.

We can draw a triangle showing the angle; as mentioned, it is R wide and $\sqrt{3}R$ high, with a $2R$ hypotenuse (between the two balls' centers). Drawing the angle, we find

$$\tan\alpha = \frac{\sqrt{3}R}{R} = \sqrt{3}$$

We then draw a vector triangle for the forces; the angle must be the same, or the net force won't point towards the center of the top ball! For the same α, clearly $\tan\alpha$ must also be the same. Relating the forces instead, we have D horizontally and mg on the vertical side, so

$$\tan\alpha = \frac{mg}{D}$$



I didn't label the forces here, since it make it very difficult to get it at all readable. Doing so is practically mandatory to solve this though, in my opinion; this was my second, simplified drawing.

(This is perhaps the cleanest thing I've drawn in years, which is why I don't post hand-drawn stuff often. It's usually much harder to read, which says something!)

α must then be the same for the net force vector, or that force will create a torque on the top ball. We can set the two tangents equal and find D:

$$\frac{mg}{D} = \sqrt{3}$$
$$D = \frac{mg}{\sqrt{3}}$$

Nice! What about the bottom ball? We have a very similar situation there! There is an upwards force $2mg$ to the middle ball instead of mg, since the bottom ball supports both of those above it.

For the sides, we again find:

$$\tan\beta = \frac{\sqrt{3}R}{R} = \sqrt{3}$$

The forces again need the same angle, so we can find the tangent for the forces, and set the two equal again:

$$\tan\beta = \frac{2mg}{B}$$
$$\sqrt{3} = \frac{2mg}{B}$$
$$B\sqrt{3} = 2mg$$
$$B = \frac{2mg}{\sqrt{3}}$$

Finally, for the middle ball, we can simply sum the horizontal forces; the one to the right needs to be equal to the sum of those to the right, or there is a net force. C to the right must cancel with $B + D$ to the left, and we know those two.

$$C = B + D$$
$$C = \frac{2mg}{\sqrt{3}} + \frac{mg}{\sqrt{3}} = \frac{3mg}{\sqrt{3}} = \frac{3\sqrt{3}mg}{3} = \sqrt{3}mg$$

And that's it! Easy once I found the trick, but I have to admit it took a while. If I hadn't drawn it out, it would have been way harder.

(59) The equations look like capstan equations, which is not entirely unexpected: we have differing tensions in something wound around a cylinder (or two).

Indeed, the recommended reading is the book's derivation of the capstan equation.

Let's start by looking at part one. I will look at the rightmost wheel, and basically assume the other one doesn't exist.

$T_2 > T_1$, or the torque would be in the opposite direction of the rotation, and so it wouldn't be in any kind of equilibrium. Therefore, the frictional force is counterclockwise along the wheel, "helping" T_1, so that there can be equilibrium.

We therefore have the same situation as the book, and don't need to think of the opposite case (reversing directions or such).

Since the derivation is fairly complex, and the book derivation applies to this situation, I will use some results from there, to get started. There is a sign difference that we can ignore if we only keep track of directions/which tension is the larger one.

$$\frac{dT}{T} = \mu_s d\theta$$

for one wheel, which answers part (a) as-is.
Part (b) is not as straightforward, with or without the book's help. First, we have one useful relationship given to us in the question:

$$\tau = R(T_2 - T_1) \Rightarrow \frac{\tau}{R} = T_2 - T_1$$

We'll need that later.
If we integrate the previous equation, from T_1 to T_2 on the left-hand side, and from 0 to π on the right, we find

$$\ln \frac{T_2}{T_1} = \mu_s \pi \Rightarrow \frac{T_2}{T_1} = e^{\mu_s \pi}$$

And so, indeed, T_2 will be larger than T_1. Solved for T_2, we have, of course

$$T_2 = T_1 e^{\mu_s \pi}$$

We now have two equations and two unknowns, so we can solve the rest from here.

$$\begin{aligned}\frac{\tau}{R} &= T_2 - T_1 \\ T_2 &= T_1 e^{\mu_s \pi}\end{aligned}$$

We can find T_1 by substitution; we stick $T_1 e^{\cdots}$ in for T_2 in the first equation and solve:

$$\begin{aligned}T_1 e^{\mu_s \pi} - T_1 &= \frac{\tau}{R} \\ T_1 \left(e^{\mu_s \pi} - 1\right) &= \frac{\tau}{R} \\ T_1 &= \frac{\tau}{R} \frac{1}{e^{\mu_s \pi} - 1}\end{aligned}$$

We have a simple relationship between T_2 and T_2 above, so finding T_2 is trivial now – at least getting it mathematically equivalent. To get it to look like one of the answer options (as this was the week's only multiple choice question), we need to divide through by the exponential, and use $1/e^x = e^{-x}$:

$$
\begin{aligned}
T_2 = T_1 e^{\mu_s \pi} &= \frac{\tau}{R} \frac{e^{\mu_s \pi}}{e^{\mu_s \pi} - 1} \\
&= \frac{\tau}{R} \frac{1}{1 - \frac{1}{e^{\mu_s \pi}}} \\
&= \frac{\tau}{R} \frac{1}{1 - e^{-\mu_s \pi}}
\end{aligned}
$$

(60) Hmm, I wonder if this can be solved in the very naive way. If we consider it attached at the very top, then essentially 100% of the weight is below that point. Therefore, we only need to find the stress due to the weight of the entire bar, $mg = (AL\rho)g$.

The ultimate tensile stress is given as a pressure, force per unit area; $P_{ult} = F/A$. We need to multiply it by the cross-sectional area A to find a force (comparable to a weight, since both are in newtons):

$$
AL\rho g = P_{ult} A
$$

A cancels:

$$
\begin{aligned}
L\rho g &= P_{ult} \\
L &= \frac{P_{ult}}{\rho g}
\end{aligned}
$$

And indeed, plugging in the values, this is correct! The answers are 4574 m (4.6 km) for iron, and 10194 m (10.2 km) for titanium.

(61) The recommended reading gives us a not-so-small hint that this is a simple harmonic oscillation.

With the condition given, there will always be slipping, and therefore always kinetic friction. We know nothing about the speed of the rotation, but since the frictional force is given by $\mu_k N$, that shouldn't matter, as long as there is always slipping.

Newton's second law in the horizontal direction (with rightwards as positive) gives us

$$
ma_x = \mu_k N_L - \mu_k N_R = \mu_k (N_L - N_R)
$$

Rewritten,

$$
\ddot{x} = \frac{\mu_k}{m}(N_L - N_R)
$$

Vertically (with upwards as positive):

$$
0 = N_L + N_R - mg
$$

Two equations, three unknowns. Now, if the center of the bar is at $x > 0$, it's clear that $N_R > N_L$, and vice versa if $x < 0$. The above equations doesn't account for that. The net torque on the bar (about the center, say) must also be zero, or it won't remain horizontal. We can capture that as

$$0 = (x + D/2)N_L - (D/2 - x)N_R$$

since gravity acting at the center of mass can cause no torque relative to the center of mass. It's unfortunate that we need to find N_L and N_R too, or there would certainly be less algebra involved. We begin by finding N_L and N_R; for that, we only need the last two equations. After that, we have one (differential) equation and one unknown left.

The vertical force equation easily gives us

$$N_L = mg - N_R$$

Solving the torque equation for N_R gives us

$$\frac{(x + D/2)}{(D/2 - x)}N_L = N_R$$

Substitute that back:

$$\begin{aligned}
N_L &= mg - \frac{(x + D/2)}{(D/2 - x)}N_L \\
N_L\left(1 + \frac{(x + D/2)}{(D/2 - x)}\right) &= mg \\
N_L &= \frac{mg}{1 + \frac{(x+D/2)}{(D/2-x)}} \\
N_L &= \frac{mg(D - 2x)}{2D}
\end{aligned}$$

And, substitute that into the equation for N_R, below:

$$\begin{aligned}
N_R &= mg - N_L \\
N_R &= mg - \frac{mg(D - 2x)}{2D}
\end{aligned}$$

For part (b), we substitute this back into the $\ddot{x}$ equation:

$$\begin{aligned}
\ddot{x} &= \frac{mu_k}{m}\left(\frac{mg(D - 2x)}{2D} - mg + \frac{mg(D - 2x)}{2D}\right) \\
\ddot{x} &= \mu_k g\left(\frac{D}{D} - \frac{2x}{D} - 1\right) \\
\ddot{x} &= -\frac{2\mu_k gx}{D}
\end{aligned}$$

The sign changes in step 1, since we get a double negative on the fraction when calculating $N_L - N_R$. Finally, for part (c), we notice that this is a simple harmonic motion, and solve it accordingly.

$$\begin{aligned}\ddot{x} + \mu_k g \frac{2}{D} x = 0 \\ x = x_0 \cos(\omega t) \\ \omega = \sqrt{\frac{2\mu_k g}{D}}\end{aligned}$$

So, all in all,

$$x = x_0 \cos\left(\sqrt{\frac{2\mu_k g}{D}} t\right)$$

If we write x as $x = \cos(\omega t + \varphi)$ and set $t = 0$, we find

$$x_0 = x_0 \cos(\varphi)$$

and so $\cos(\varphi) = 1 \Rightarrow \varphi = 0$, which is why I didn't include it above. (I figured as much since it was released from rest, not to mention they didn't ask for it.)

(62) First, let's identify the forces present. There's the spring force of magnitude kx, and the frictional force F_f.
When the spring is stretched, the spring force is towards the right, in the direction of the acceleration. The frictional force is opposite that, and will provide a torque that causes the cylinder to roll.
If we use rightwards as positive (since the acceleration will begin in that direction), kx will begin negative, since the initial position is $x = -x_0$. As usual, then, we must write $-kx$ for the spring force. The frictional force also has a negative, since it's towards the left when the acceleration is positive:

$$m\ddot{x} = -kx - F_f$$

Next, since there is pure roll, we can use $a = \ddot{x} = \alpha R$. We also have that $\tau = I\alpha$, which leads us to (via $\tau = RF_f$ and $I = \frac{1}{2}MR^2$):

$$\begin{aligned}RF_f &= (\frac{1}{2}MR^2)(\ddot{x}/R) \\ F_f &= \frac{1}{2}M\ddot{x}\end{aligned}$$

We could also write an equation relating vertical forces, but it turns out we don't need to.
If we substitute the value of F_f into the previous equation,

$$
\begin{aligned}
M\ddot{x} &= -kx - \frac{1}{2}M\ddot{x} \\
\frac{3}{2}M\ddot{x} &= -kx \\
\ddot{x} + \frac{2k}{3M}x &= 0
\end{aligned}
$$

A simple harmonic oscillation, as we would expect. The solution is then

$$
\begin{aligned}
x &= x_0 \cos(\omega t + \pi) \\
\omega &= \sqrt{\frac{2k}{3M}} \\
T &= \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{3M}{2k}}
\end{aligned}
$$

where I wrote the phase as π since at $t = 0$, we need $x = -x_0$. I could also have written the entire right-hand side as negative.

(63) The liquid has a velocity that is the same everywhere (under these conditions), $\dot{x}$. Therefore, the liquid as a whole has a kinetic energy of

$$
\frac{1}{2}M\dot{x}^2 = \frac{1}{2}AL\rho\dot{x}^2
$$

There is also gravitational potential energy. We define $U = 0$ at the equilibrium point. The change is then that a height of fluid x of mass $m = Ax\rho$ is moved upwards a distance x. (It's essentially taken from the left side and moved upwards on the right side, gaining potential energy.)
The sum of these two energies must be a constant:

$$
\frac{1}{2}AL\rho\dot{x}^2 + Ax\rho gx = \text{constant}
$$

using $mgh = (Ax\rho)gx$.

We take the time derivative of this; the rate of change in the energy must be zero if it's constant, which the differentiation takes of for us.

$$
\begin{aligned}
\frac{1}{2}AL\rho\dot{x}^2 + A\rho gx^2 &= \text{constant} \\
\frac{1}{2}AL\rho 2\dot{x}\ddot{x} + A\rho g 2x\dot{x} &= 0 \\
L\ddot{x} + 2gx &= 0 \\
\ddot{x} + \frac{2g}{L}x &= 0
\end{aligned}
$$

$\dot{x}$, A and ρ cancel, and we end up with a simple harmonic oscillation, as expected (and as usual, at this point!). The solution is

$$
\begin{aligned}
x &= x_0 \cos(\omega t) \\
\omega &= \sqrt{\frac{2g}{L}} \\
f &= \frac{1}{2\pi}\sqrt{\frac{2g}{L}}
\end{aligned}
$$

... though in reality there will be losses which cause damping, so T will be longer, and the amplitude will decrease rather rapidly, rather than stay constant forever as this solution predicts.

(64) The total volume of the hydrometer is

$$V_{sphere} + V_{cylinder} = \frac{4}{3}\pi R^3 + \pi r^2 \ell$$

while the submerged part is

$$\frac{4}{3}\pi R^3 + \pi r^2(\ell - h)$$

Since it floats, the upwards buoyant force must be equal to the downwards gravitational force Mg.
The buoyant force is equal to the weight of the displaced water, which is the submerged volume times ρ (which is its mass) times g. That is,

$$Mg = \rho g \left(\frac{4}{3}\pi R^3 + \pi r^2(\ell - h)\right)$$

$$\rho = \frac{M}{\frac{4}{3}\pi R^3 + \pi r^2(\ell - h)}$$

(65) The volumetric flow rate must be the same both the thick part at d_1 and the thinner at d_2, since water is practically incompressible.
Therefore, the velocity must be greater at point 2 than at point 1.
I will, for consistency, use v_1 for the velocity at point 1; $v_1 = v_m$.

$$
\begin{aligned}
Q &= v_1 A_1 = v_2 A_2 \\
Q &= v_1 \pi \left(\frac{d_1}{2}\right)^2 = v_2 \pi \left(\frac{d_2}{2}\right)^2
\end{aligned}
$$

This gives us

$$v_1 d_1^2 - v_2 d_2^2 = 0$$

We can also relate the energies at the two points via Bernoulli's equation. We have kinetic energy (per unit volume), gravitational potential energy (per unit volume), and pressure. The GPE is equal at the two points, as they are at equal height with equal ρ, so if we wrote it down it would simply cancel.

$$\frac{1}{2}\rho v_1^2 + P_1 = \frac{1}{2}\rho v_2^2 + P_2$$

We don't know v_1, v_2, P_1 or P_2, so we have four unknowns. We can rewrite this a bit, though.

$$P_1 - P_2 = \frac{1}{2}\rho\left(v_2^2 - v_1^2\right) \tag{A.1}$$

We can use the height of the water columns to figure out the pressure difference.

The air at the top of the water columns are at atmospheric pressure, call it $P_0 = 1\,\text{atm}$.

The height of the left column, measured from the horizontal center line, depends on $P_1 - P_0$, via Pascal's law:

$$P_1 - P_0 = \rho g h_1$$

The right column is similar.

$$P_2 - P_0 = \rho g h_2$$

We don't know h_1 or h_2, but we know $h_1 - h_2 = \Delta h$. If we subtract the two equations,

$$\begin{aligned}(P_1 - P_0) - (P_2 - P_0) &= \rho g h_1 - \rho g h_2\\ P_1 - P_2 &= \rho g \Delta h\end{aligned}$$

We use this in equation (A.1). That gives us these two equations (after ρ cancels):

$$\begin{aligned}g\Delta h &= \frac{1}{2}\left(v_2^2 - v_1^2\right)\\ v_1 d_1^2 - v_2 d_2^2 &= 0\end{aligned}$$

Since we don't care about v_2, we can solve the second equation for it, substitute that into the first, and then just forget about v_2 altogether.

$$v_2 = v_1\frac{d_1^2}{d_2^2}$$

$$
\begin{aligned}
2g\Delta h &= \left(v_1 \frac{d_1^2}{d_2^2}\right)^2 - v_1^2 \\
2g\Delta h &= v_1^2 \left(\frac{d_1^4}{d_2^4} - 1\right) \\
\sqrt{\frac{2g\Delta h}{\frac{d_1^4}{d_2^4} - 1}} &= v_1 \\
\sqrt{\frac{2g\Delta h\ d_2^4}{d_1^4 - d_2^4}} &= v_1
\end{aligned}
$$

For the number we were given, this gives us $v_1 = v_m = 1.6174\,\mathrm{m/s}$.
Using the simple relationship $Q = v_1 A_1 = v_1 \left(\frac{d_1}{2}\right)^2$ we find a flow rate of $Q = 0.203\,\mathrm{m^3/s}$.

(66) We can solve this in multiple ways:
Solution 1

The pressure at that depth is $P_1 = 1\,\mathrm{atm} + \rho g h$. The pressure difference between inside and outside the bucket is then simply $\rho g h$.
We can apply Bernoulli's equation here, again while ignoring the term related to gravitational potential energy, as there is no height difference involved (if we consider a point at that depth, but at the container's left side, as being inside). Using P_1 for the pressure inside the bucket at depth h, and P_2 for the pressure outside:

$$
\begin{aligned}
\frac{1}{2}\rho v_{inside}^2 + P_1 &= \frac{1}{2}\rho v^2 + P_2 \\
\frac{1}{2}\rho v_{inside}^2 + 1\,\mathrm{atm} + \rho g h &= \frac{1}{2}\rho v^2 + 1\,\mathrm{atm} \\
\frac{1}{2} v_{inside}^2 + g h &= \frac{1}{2} v^2 \\
h &= \frac{v^2}{2g}
\end{aligned}
$$

Here, I consider v_{inside} to be negligible compared to v, so I ignore it. It we consider v_{inside} to be the velocity ***just*** inside the hole, that is clearly not correct. However, the rest of the equation is equally valid at the leftmost edge of the container.
Solution 2

I feel a bit funny about the assumption $v_{inside} = 0$ while considering a point at depth h in the liquid, as the equation doesn't specify where that point is: near the hole, or far from it.
We can solve this in a slightly different way. We again begin with Bernoulli's equation, but this time, we consider a point at the surface of the liquid (above the hole), and a point just outside the hole. Both are exposed to the atmosphere, so $P_1 = P_2 = 1\,\mathrm{atm}$ and we don't need to specify that in the equation, as it will simply cancel.

Instead, we have the gravitational potential energy per unit volume, $\rho g y$, in the equation. On the left side, we have at the top of the container, where it is $\rho g h$; I define the zero level to be at the hole, so the term only exists on the left-hand side.

$$
\begin{aligned}
\frac{1}{2}\rho v_{surface}^2 + \rho g h &= \frac{1}{2}\rho v^2 \\
2gh &= v^2 \\
h &= \frac{v^2}{2g}
\end{aligned}
$$

As before, we approximate the other velocity, this time at the surface, to be zero. We find exactly the same result using this method.

(67) The buoyant force is given by the weight of the displaced fluid – air in this case – so this should be very simple. Weight is given by mass times g, while mass is ρV, so $F_B = V\rho_{air} g$:

$$F_B = (540\,000.0\,\mathrm{m}^3)(4.3\times10^{-3}\,\mathrm{kg/m}^3)(10\,\mathrm{m/s}^2) \approx 23220N$$

Very simple indeed.

Appendix B

GNU Free Documentation License

Preamble

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "**Document**", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "**you**". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "**Modified Version**" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "**Secondary Section**" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "**Invariant Sections**" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "**Cover Texts**" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "**Transparent**" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "**Opaque**".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "**Title Page**" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

The "**publisher**" means any person or entity that distributes copies of the Document to the public.

A section “**Entitled XYZ**” means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as “**Acknowledgements**”, “**Dedications**”, “**Endorsements**”, or “**History**”.) To “**Preserve the Title**” of such a section when you modify the Document means that it remains a section “Entitled XYZ” according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document’s license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

C. State on the Title page the name of the publisher of the Modified Version, as the publisher.

D. Preserve all the copyright notices of the Document.

E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document’s license notice.

H. Include an unaltered copy of this License.

I. Preserve the section Entitled “History”, Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled “History” in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the “History” section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

K. For any section Entitled “Acknowledgements” or “Dedications”, Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

M. Delete any section Entitled “Endorsements”. Such a section may not be included in the Modified Version.

N. Do not retitle any existing section to be Entitled “Endorsements” or to conflict in title with any Invariant Section.

O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

11. RELICENSING

"Massive Multiauthor Collaboration Site" (or "MMC Site") means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A "Massive Multiauthor Collaboration" (or "MMC") contained in the site means any set of copyrightable works thus published on the MMC site.

"CC-BY-SA" means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

"Incorporate" means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is "eligible for relicensing" if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.