

CHAPTER

10

Conservation Laws, Collisions and Gravity

Some things never change. Knowing which ones is half of physics.

IN THIS CHAPTER

10.1	Newton's Cradle	STEP II 1990
10.2	The Stacked-Ball Effect	STEP II 2010
10.3	Rockets and Guns	STEP III 2017
10.4	A Molecule and a Piston	STEP I 2002
10.5	Infinitely Many Bounces in Finite Time	STEP III 2002
10.6	Two Bodies Become One	STEP III 2004
10.7	Gravity Inside a Shell	STEP III 1999
10.8	Three Stars in a Triangle	STEP I 1999

What Stays the Same

There is a fact, or if you wish, a law, governing all natural phenomena that are known to date. There is no known exception to this law.

— Richard Feynman, on the conservation of energy, *The Feynman Lectures on Physics* (1963)

Mechanics at school is usually a matter of forces: draw a diagram, resolve, apply $F = ma$. That works for a block on a slope. It does not work for a collision lasting a thousandth of a second, during which the forces are huge and unknown, or for a rocket that changes its mass as it goes, or for three stars pulling on each other at once. In all of these, the way forward is to stop asking about forces and ask instead what *cannot change*.

Momentum is conserved whenever no outside force acts. Energy is conserved in elastic collisions, and its loss can be measured in the others. Mass, angular momentum and the motion of the centre of mass follow similar rules. These laws do not tell you everything, but they are always true, they apply to systems far too complicated to analyse in detail, and they often rule out most of the possibilities before any calculation begins.

Why conservation laws matter

The conservation laws are the most reliable tools in physics. In the radioactive process called beta decay, the energies of the emerging particles did not add up; rather than abandon conservation of energy, Wolfgang Pauli predicted in 1930 an invisible particle carrying the missing energy, and the neutrino was detected twenty-six years later. Engineers design crash barriers, airbags and rockets by momentum and energy budgets. Astronomers weigh stars and find planets by watching centres of mass. And in 1918 Emmy Noether explained why the laws hold: each conservation law is the shadow of a symmetry of nature. Momentum is conserved because the laws of physics are the same everywhere; energy, because they are the same at all times.

Gravity is the second theme of this chapter. It is the force Newton used to unify the heavens and the Earth, and it is a natural partner of the conservation laws: orbits, slingshots and collapsing clouds are all understood through energy and momentum, together with one remarkable fact about spheres.

What you will meet at university

A first-year dynamics course starts here. You will meet systems of particles, the centre of mass and the reduced mass, collisions in different frames, variable-mass systems such as rockets, and the motion of bodies under gravity: Kepler's laws, orbits, escape velocity.

Later courses reformulate mechanics in the languages of Lagrange and Hamilton, where Noether’s theorem becomes a precise statement and conservation laws can be read off from the symmetries of a single function. Beyond that lie celestial mechanics and the three-body problem, kinetic theory and statistical mechanics (where the piston of Section 10.4 becomes the theory of gases), and Einstein’s theory of gravity.

The questions that run through this chapter

Three big questions

1. **What can conservation laws decide on their own?** When do momentum and energy determine an outcome, and when do they only rule things out?
2. **How do small steps add up?** What happens when a simple event, a collision or a firing, is repeated many times, or infinitely often?
3. **How do we tame many bodies?** Which choices of coordinates, and which symmetries, turn a problem about several interacting bodies into one we can solve?

The chapter begins with collisions: what Newton’s cradle can and cannot do (Section 10.1), and how a chain of balls multiplies speed (Section 10.2). It then follows repeated processes: guns fired from a truck and the rocket equation, a molecule speeding up against a piston, and a puck that bounces infinitely often in a finite time. The last three sections turn to bodies that attract or repel: the centre of mass and reduced mass, the gravity of a spherical shell, and Lagrange’s rotating triangle of stars.

Ideas in this chapter

§	Topic	Idea
10.1	Newton’s cradle	Conservation laws as constraints
10.2	Stacked balls	Iterating one collision; slingshots
10.3	Rockets	Variable mass; the rocket equation
10.4	Kinetic theory	An adiabatic invariant
10.5	Inelastic collisions	Infinitely many bounces in finite time
10.6	The two-body problem	Centre of mass and reduced mass
10.7	Newtonian gravity	The shell theorem; free-fall collapse
10.8	The three-body problem	Lagrange’s equilateral solution

How to read this chapter

Before calculating, list what is conserved, and in which frame. Then ask whether the problem looks simpler from somewhere else: from the centre of mass, from the moving piston, from the bouncing ball. Many of the questions in this chapter become almost trivial in the right frame, and most of the work in mechanics is finding it.

10.1 Newton's Cradle

The question

The identical uniform smooth spherical marbles $A_1, A_2, \dots, A_n$, where $n \geq 3$, each of mass m , lie in that order in a smooth straight trough, with each marble touching the next. The marble A_{n+1} , which is similar to A_n but has mass λm , is placed in the trough so that it touches A_n . Another marble A_0 , identical to A_n , slides along the trough with speed u and hits A_1 . It is given that kinetic energy is conserved throughout.

- (i) Show that if $\lambda < 1$, there is a possible subsequent motion in which only A_n and A_{n+1} move (and A_0 is reduced to rest), but that if $\lambda > 1$, such a motion is not possible.
- (ii) If $\lambda > 1$, show that a subsequent motion in which only A_{n-1} , A_n and A_{n+1} move is not possible.
- (iii) If $\lambda > 1$, find a possible subsequent motion in which only two marbles move.

The hook

Pull back one ball of a Newton's cradle and let it go: one ball flies out at the other end. Pull back two, and two fly out. Why? Why not one ball at twice the speed, or three balls at two-thirds of the speed? Both would carry the right momentum.

The answer begins with the two great conservation laws of mechanics. Momentum and energy together are much more restrictive than either alone, and they rule out most of the outcomes you might imagine. This question uses them as constraints: not to calculate what happens, but to decide what *can* happen. That way of thinking, with conservation laws as rules that every possible outcome must obey, runs through the whole of physics.

Before you start



Ask yourself A marble of mass m moving at speed u hits a stationary marble of mass M , and kinetic energy is conserved. What are the velocities afterwards? What if $M = m$?

Let the velocities afterwards be p and q . Momentum gives $mp + Mq = mu$ and energy gives $mp^2 + Mq^2 = mu^2$. Writing these as $m(u - p) = Mq$ and $m(u - p)(u + p) = Mq^2$

and dividing, $u + p = q$. Then

$$p = \frac{m - M}{m + M} u, \quad q = \frac{2m}{m + M} u.$$

When $M = m$, the first marble stops and the second moves off at speed u : the two simply exchange velocities. That is why the marbles of equal mass pass the motion along the line.

Working it through

STAGE 1 Only the last two move



Ask yourself Suppose only A_n (velocity p) and A_{n+1} (velocity q) move afterwards. Write down the conservation laws. Which marble's motion constrains the sign of p ?

Momentum and energy give

$$mp + \lambda mq = mu, \quad \frac{1}{2}mp^2 + \frac{1}{2}\lambda mq^2 = \frac{1}{2}mu^2.$$

These are the equations for a marble of mass m hitting one of mass λm , so (apart from the solution $p = u, q = 0$, which would need A_n to pass through A_{n+1})

$$p = \frac{1 - \lambda}{1 + \lambda} u, \quad q = \frac{2}{1 + \lambda} u.$$

If $\lambda < 1$, both are positive and $p < q$: A_n follows A_{n+1} but more slowly, so they separate, and everything is consistent. If $\lambda > 1$, then $p < 0$: A_n moves backwards. But A_{n-1} is at rest directly behind it, touching it, and A_n cannot pass through it. So this motion is impossible. ■

STAGE 2 Three marbles



Ask yourself Now suppose only A_{n-1} , A_n and A_{n+1} move, with velocities a, b, c . What inequalities must a, b, c satisfy, given which marbles are at rest and which are behind which?

The marble A_{n-2} is at rest behind A_{n-1} (it exists, since $n \geq 3$), so $a \geq 0$. And no marble can overtake the one in front, so $a \leq b \leq c$. Altogether $0 \leq a \leq b \leq c$. The conservation laws are

$$a + b + \lambda c = u, \quad a^2 + b^2 + \lambda c^2 = u^2.$$



Ask yourself Square the momentum equation and compare it with the energy equation. What is left?

Squaring, $u^2 = a^2 + b^2 + \lambda^2 c^2 + 2ab + 2\lambda ac + 2\lambda bc$. Subtracting the energy equation,

$$0 = \lambda(\lambda - 1)c^2 + 2ab + 2\lambda c(a + b).$$

With $\lambda > 1$ and $a, b, c \geq 0$, every term on the right is non-negative, so every term is zero. In particular $c = 0$, and then $0 \leq a \leq b \leq c$ forces $a = b = 0$. But then the total momentum is 0, not mu . Contradiction: this motion is impossible. ■



Pause and try Notice the structure of the argument. Squaring the momentum equation produces cross terms like $2ab$, and the energy equation has none. The difference measures how “spread out” the motion is. Can you see why the same argument rules out *any* outcome in which only $A_k, \dots, A_{n+1}$ move with non-negative velocities, when $\lambda > 1$?

STAGE 3 Two marbles that can move



Ask yourself If the marbles moving forwards must all be at the far end, and that fails, perhaps one marble moves *backwards*. Which marble is free to move backwards?

Only A_0 has nothing behind it. Try: A_0 moves with velocity p (possibly negative), A_{n+1} with velocity q , and all the others are at rest. The conservation laws are again those of a collision between masses m and λm , so

$$p = -\frac{\lambda - 1}{\lambda + 1}u, \quad q = \frac{2}{\lambda + 1}u.$$

A_0 bounces back, A_{n+1} moves forwards, and the marbles in between stay at rest, all consistent. It is exactly as if A_0 had hit A_{n+1} directly, with the line of equal marbles in between acting as a perfect messenger. ■

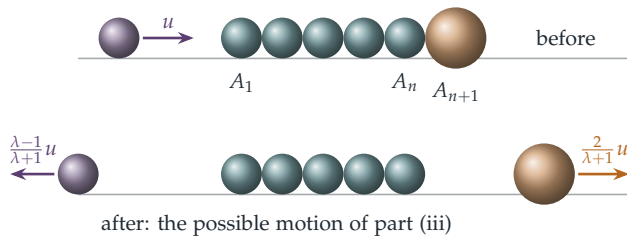


Figure 10.1.1 A row of equal marbles (teal) ending in a heavier one (amber, $\lambda > 1$). After the impact, A_0 bounces back, the heavy marble moves off, and every marble in between is left at rest, exactly as if A_0 had hit A_{n+1} directly.

Looking back

- **Conservation laws as constraints.** We never followed the collisions in detail. We asked which outcomes are allowed by momentum, energy and geometry.

- **Physical inequalities matter.** “Marbles cannot pass through each other” gave the inequalities $0 \leq a \leq b \leq c$, and these, not the conservation laws, ruled out parts (i) and (ii).
- **Square and subtract.** Comparing the square of the momentum with the energy gives a sum of cross terms that must vanish.

What’s really going on

Why one ball comes out

In a Newton’s cradle of equal balls, suppose k balls leave together at speed w . Momentum gives $kw = u$ and energy gives $kw^2 = u^2$. Dividing, $w = u$ and $k = 1$. So among outcomes where the departing balls move together, only “one ball at the original speed” obeys both laws. Two balls at half speed carry the right momentum but only half the energy; one ball at twice the speed would have the right momentum only if it were half a ball.

But this does not settle the matter. With n balls there are n unknown velocities and only two conservation laws, so there are infinitely many outcomes that obey both (try finding one with three balls moving at different speeds). Which one happens depends on how the balls deform during contact. In the early 1980s Frank Herrmann and his colleagues showed that the familiar behaviour depends on the shock passing along the line *without spreading out*, which happens for hard steel balls with tiny gaps between them. With soft balls, or balls pressed tightly together, the cradle behaves less cleanly, and careful measurements show that the other balls always move a little.

Conservation laws and symmetry

Where do the conservation laws come from? In 1918 Emmy Noether proved a theorem that is one of the foundations of modern physics: every continuous symmetry of a physical system gives a conserved quantity. If the laws of physics are the same here as one metre to the left (symmetry under translation in space), momentum is conserved. If they are the same today as tomorrow (symmetry under translation in time), energy is conserved. If they are the same in every direction (rotational symmetry), angular momentum is conserved. You will meet Noether’s theorem in a second-year course on Lagrangian mechanics; this chapter uses its consequences throughout.

Huygens, Wren and Wallis

In 1668 the Royal Society invited solutions to the problem of colliding bodies. John Wallis treated inelastic collisions, and Christopher Wren and Christiaan Huygens independently solved the elastic case. Huygens’s key insight was to change frame: view the collision from a boat moving with the centre of mass, where two equal balls simply bounce back with their speeds reversed, and then transform back to the shore. He also noticed that a certain quantity, the sum of the masses times the squares of the

speeds, was the same before and after, though it was another century before kinetic energy was recognised as a conserved quantity in its own right.

The thinking skill

Use conserved quantities to rule things out

When a process is too complicated to follow, list the quantities that cannot change, and see which outcomes they allow.

This is how physicists reason about collisions in particle accelerators, where the detailed interactions are unknown but momentum, energy and charge must balance. Combined with an inequality (here, that marbles cannot pass through each other), it can prove that an outcome is impossible.

Try this next

- 1 The same problem again.** (From STEP I, 2006.) Particles $A_1, \dots, A_n$ lie at rest in a trough; A_{n-1} has mass m and A_n has mass λm with $\lambda > 1$. A particle A_0 of mass m hits A_1 with speed u . Show that it is impossible for exactly one particle to be moving after all collisions, or for only A_{n-1} and A_n , or only A_{n-2}, A_{n-1}, A_n . Given that exactly two particles move, find their speeds.
- 2 Following the collisions.** Model the impact in part (iii) as a sequence of separate two-marble collisions, starting with A_0 hitting A_1 . Follow the sequence through and show that it ends in the motion found in part (iii).
- 3 Two balls in.** In a cradle of five equal balls, two balls are released together and hit the other three with speed u . Show that “two balls leave with speed u ” is consistent with both conservation laws, and find another outcome that is.
- 4 Light end.** If $\lambda < 1$, follow the sequence of two-marble collisions. Do any marbles other than A_n and A_{n+1} end up moving?

Further reading

- **David Morin, *Introduction to Classical Mechanics*** (Cambridge University Press), Chapter 5. Conservation of momentum and energy, with many problems on collisions.
- **Frank Herrmann & Peter Schmälzle, “Simple explanation of a well-known collision experiment”, *American Journal of Physics* 49 (1981).** Why Newton’s cradle works.
- **Dwight Neuenschwander, *Emmy Noether’s Wonderful Theorem*** (Johns Hopkins University Press). An accessible account of the link between symmetry and conservation.

10.2 The Stacked-Ball Effect

The question

- (i) In an experiment, a particle A of mass m is at rest on a smooth horizontal table. A particle B of mass bm , where $b > 1$, is projected along the table directly towards A with speed u . The collision is perfectly elastic.

Find an expression for the speed of A after the collision in terms of b and u , and show that, irrespective of the relative masses of the particles, A cannot be made to move at twice the initial speed of B .

- (ii) In a second experiment, a particle B_1 is projected along the table directly towards A with speed u . This time, particles $B_2, B_3, \dots, B_n$ are at rest in order on the line between B_1 and A . The mass of B_i ($i = 1, 2, \dots, n$) is $\lambda^{n+1-i}m$, where $\lambda > 1$. All collisions are perfectly elastic. Show that, by choosing n sufficiently large, there is no upper limit on the speed at which A can be made to move.

In the case $\lambda = 4$, determine the least value of n for which A moves at more than $20u$. You may use the approximation $\log_{10} 2 \approx 0.30103$.

The hook

Hold a tennis ball on top of a basketball and drop them together. The basketball bounces off the floor, and the tennis ball shoots up to several times the height you dropped it from, sometimes high enough to hit the ceiling. Try it outdoors.

Part (i) of the question shows that a single collision has a speed limit: however heavy the ball that hits it, a stationary ball cannot be sent off at more than twice the speed of the incoming one. Part (ii) shows how to break the limit: pass the motion down a chain of ever-lighter balls, and the speed multiplies at every step. The same idea sends space probes to the outer planets, and has been suggested as part of the mechanism of a supernova.

Before you start



Ask yourself From Section 10.1: if a mass M moving at speed u hits a stationary mass m elastically, what is the speed of m afterwards?

The result there (with the roles of the masses swapped) gives the speed of the struck particle as $\frac{2M}{M+m}u$. Another way to remember it: in an elastic collision the *relative* velocity is reversed. If the heavy ball barely slows down, the light ball must leave at about u more than the heavy ball's speed, that is about $2u$.

Working it through

STAGE 1 The speed limit

With $M = bm$, the speed of A after the collision is

$$v_A = \frac{2bm}{bm + m} u = \frac{2b}{b + 1} u.$$



Ask yourself How does $\frac{2b}{b+1}$ compare with 2? What happens as $b \rightarrow \infty$?

Since $\frac{2b}{b+1} = 2 - \frac{2}{b+1} < 2$ for every b , the speed of A is always less than $2u$, however heavy B is. It approaches $2u$ as $b \rightarrow \infty$: an infinitely heavy ball would bounce A off as if from a moving wall. ■

STAGE 2 Passing it along a chain



Ask yourself In part (ii), each particle is λ times as heavy as the next one. What does part (i) say about each collision in the chain?

B_1 (mass $\lambda^n m$) hits B_2 (mass $\lambda^{n-1} m$), which is λ times lighter. By part (i) with $b = \lambda$, B_2 moves off at $\frac{2\lambda}{\lambda+1}u$. Then B_2 hits B_3 , which is again λ times lighter, and B_3 moves off at $\left(\frac{2\lambda}{\lambda+1}\right)^2 u$. After n such collisions, the last one being B_n hitting A ,

$$v_A = \left(\frac{2\lambda}{\lambda+1}\right)^n u.$$

(Later collisions between the B s can only push A further; they cannot slow it, since nothing ever hits A from in front.)



Ask yourself Is $\frac{2\lambda}{\lambda+1}$ bigger or smaller than 1?

Since $\lambda > 1$, $2\lambda > \lambda + 1$, so the ratio $r = \frac{2\lambda}{\lambda+1}$ exceeds 1, and $r^n \rightarrow \infty$ as $n \rightarrow \infty$. So by taking n large enough, A can be made to move as fast as we like. ■

STAGE 3 Twenty times as fast

With $\lambda = 4$, $r = \frac{8}{5} = 1.6$, and we need $1.6^n > 20$. Taking logarithms to base 10,

$$\log_{10} 1.6 = \log_{10} 16 - 1 = 4\log_{10} 2 - 1 \approx 0.20412, \quad \log_{10} 20 = 1 + \log_{10} 2 \approx 1.30103,$$

so we need $n > \frac{1.30103}{0.20412} \approx 6.37$. The least value is $n = 7$. (Indeed $1.6^6 \approx 16.8$ and $1.6^7 \approx 26.8$.) ■

Looking back

- **Iterate a single result.** Part (i) was used n times in part (ii), turning a bounded gain into an unbounded one.
- **A limit that is never reached.** The bound $2u$ is approached but never attained, so the right statement is a strict inequality.
- **Logarithms turn powers into products.** The question of how many steps are needed became a division.

What's really going on**The basketball and the tennis ball**

Let the two balls fall together from height H , reaching the floor at speed $U = \sqrt{2gH}$. Suppose everything is perfectly elastic and the basketball is much heavier. It bounces off the floor with speed U upwards, and immediately meets the tennis ball, still moving down at speed U . In the basketball's frame, the tennis ball approaches at $2U$ and, bouncing off a very heavy object, leaves at $2U$. Back in the ground frame, that is $2U + U = 3U$ upwards (Figure 10.2.1). Since height is proportional to the square of the launch speed, the tennis ball rises to $9H$.

Real balls are not perfectly elastic. With coefficient of restitution e and mass ratio M/m , the tennis ball leaves at speed $\frac{Me(2+e)-m}{M+m}U$. STEP I 2014 Q10, in the exercises, uses a measured height to work backwards to the mass ratio.

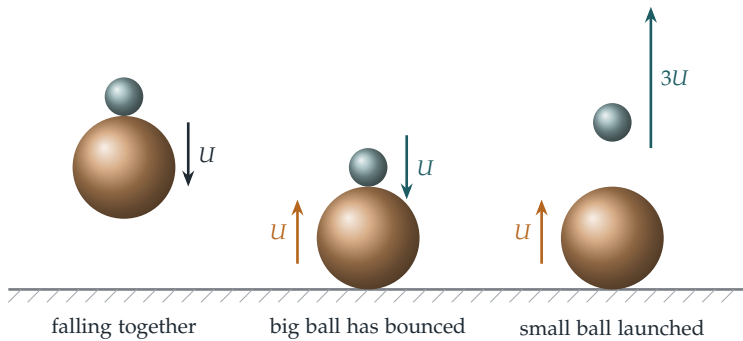


Figure 10.2.1 The stacked-ball effect, for perfectly elastic balls with the lower one much heavier. The big ball bounces up at speed U and meets the small ball coming down at U . Relative to the big ball, the small one arrives at $2U$ and leaves at $2U$, so it leaves the ground frame at $3U$, reaching nine times the original height.

Gravitational slingshots

A space probe flying past a planet does not touch it, but gravity acts like a very soft elastic collision with a very heavy object. In the planet's frame, the probe swings round and leaves with the same speed it arrived with. If the planet is moving towards the probe at speed V , then in the Sun's frame the probe can gain up to $2V$, exactly as the tennis ball does. The two *Voyager* probes, launched in 1977, used a rare alignment of the outer planets to pass from Jupiter to Saturn (and, for *Voyager 2*, on to Uranus and Neptune), gaining speed at each encounter. Without these gravity assists, the journey to Neptune would have taken about thirty years instead of twelve. The energy comes from the planet, which slows down by an immeasurably small amount.

Superballs and supernovae

In 1968 W. R. Mellen described in the *American Journal of Physics* a “superball rebound projectile” made by stacking several balls of decreasing size, and noted that the top ball can be launched at many times the impact speed. Astrophysicists have pointed out a loose analogy with a core-collapse supernova: the infalling outer layers of a star rebound off the stiff, dense core, and a shock wave carries momentum outwards into ever less dense material, accelerating it to enormous speeds. Demonstrations of the stacked-ball effect are used to explain the idea to students, though real supernovae involve much more physics, including neutrinos, than bouncing balls.

The thinking skill

Iterate a single result

When a process is a repetition of one simple step, solve the step once and compose it.

A gain factor less than 2 per step becomes unbounded after many steps, because the factors multiply. At university this is the idea behind transfer matrices, multi-stage amplifiers and multi-stage rockets (Section 10.3): the gain of a chain is the product of the gains of its links.

Try this next

- 1 The real thing.** (From STEP I, 2014.) A ball of mass M and radius R is released from rest with its centre at height $H + R$, with a ball of mass m and radius r on top. The coefficient of restitution for every impact is e . Show that, bouncing alone, the large ball would rise until its centre is at height $R + He^2$. Given $R = 0.2$, $r = 0.05$, $H = 1.8$, $e = \frac{2}{3}$ and that the small ball's centre reaches height 4.5, show that $M/m = 9$.
- 2 An inelastic chain.** (From STEP I, 2017.) Particles $P_1, P_2, \dots$ of masses $\lambda m, \lambda^2 m, \dots$ lie at $x = 1, 2, \dots$, and a particle of mass m hits P_1 at speed u ; the coefficient of restitution is e . Show that P_1 moves off at $\frac{1+e}{1+\lambda}u$. Find the speed of P_n after its first collision, and show that if $e > \lambda$ each particle is involved in exactly two collisions. What happens if $\lambda e = 1$?
- 3 Three balls.** Three perfectly elastic balls are stacked, each much lighter than the one below. Show that the top ball leaves at $7U$ and rises to $49H$. With k balls, show that the top one leaves at $(2^k - 1)U$.
- 4 Where does the energy come from?** In part (ii), show that the total kinetic energy is conserved even though A 's speed is unbounded. What happens to the heavy particles?

Further reading

- **David Morin**, *Introduction to Classical Mechanics* (Cambridge University Press), Chapter 5, which includes the stacked-ball problem.
- **W. R. Mellen**, “Superball rebound projectiles”, *American Journal of Physics* 36 (1968). Short and fun.
- **Gary Flandro**, “Fast reconnaissance missions to the outer solar system utilizing energy derived from the gravitational field of Jupiter”, *Astronautica Acta* 12 (1966). The paper that proposed the Voyager “grand tour”.

10.3 Rockets and Guns

The question

A railway truck, initially at rest, can move forwards without friction on a long straight horizontal track. On the truck, n guns are mounted parallel to the track and facing backwards, where $n > 1$. Each of the guns is loaded with a single projectile of mass m . The mass of the truck and guns (but not including the projectiles) is M .

When a gun is fired, the projectile leaves its muzzle horizontally with a speed $v - V$ relative to the ground, where V is the speed of the truck immediately before the gun is fired.

- (i) All n guns are fired simultaneously. Find the speed, u , with which the truck moves, and show that the kinetic energy, K , which is gained by the system (truck, guns and projectiles) is given by

$$K = \frac{1}{2}nmv^2\left(1 + \frac{nm}{M}\right).$$

- (ii) Instead, the guns are fired one at a time. Let u_r be the speed of the truck when r guns have been fired, so that $u_0 = 0$. Show that, for $1 \leq r \leq n$,

$$u_r - u_{r-1} = \frac{mv}{M + (n - r)m} \quad (*)$$

and hence that $u_n < u$.

- (iii) Let K_r be the total kinetic energy of the system when r guns have been fired (one at a time), so that $K_0 = 0$. Using $(*)$, show that, for $1 \leq r \leq n$,

$$K_r - K_{r-1} = \frac{1}{2}mv^2 + \frac{1}{2}mv(u_r - u_{r-1})$$

and hence show that

$$K_n = \frac{1}{2}nmv^2 + \frac{1}{2}mvu_n.$$

Deduce that $K_n < K$.

The hook

A rocket in space has nothing to push against. It moves forward by throwing mass backwards, and conservation of momentum does the rest. A truck firing guns off its back is a rocket broken into pieces: each projectile is a lump of exhaust.

So should the guns fire all at once, or one after another? This question gives a definite answer, but look carefully at what it compares. Firing all at once gives the truck more speed, but it also uses more energy, so the guns are not doing the same job in the two cases. Change what is held fixed, and the answer reverses. Sorting this out leads to the most important equation in spaceflight, Tsiolkovsky's *rocket equation*, and to the reason why real rockets have several stages.

Before you start



Ask yourself Before a gun fires, the truck moves at V . Afterwards the projectile's velocity is $V - v$ (that is, speed $v - V$ backwards). What is the projectile's velocity *relative to the truck before firing*?

It is $(V - v) - V = -v$. So each gun throws its projectile backwards at speed v relative to the truck's velocity just before firing. That is the fixed rule of this question; keep it in mind, because it is what decides the answer.

Working it through

STAGE 1 All at once

The system starts at rest, with zero momentum. After firing, the truck (mass M) moves forward at u , and the n projectiles move backwards at speed v (since $V = 0$ before firing). So

$$Mu - nmv = 0, \quad u = \frac{nmv}{M}.$$

The kinetic energy gained is

$$K = \frac{1}{2}Mu^2 + \frac{1}{2}nmv^2 = \frac{1}{2}\frac{n^2m^2v^2}{M} + \frac{1}{2}nmv^2 = \frac{1}{2}nmv^2\left(1 + \frac{nm}{M}\right). \quad \blacksquare$$

STAGE 2 One at a time



Ask yourself Just before the r th gun fires, what is the mass of the truck with its remaining projectiles, and what is its speed? Write down conservation of momentum for the r th firing.

Before the r th firing, the truck carries $n - r + 1$ projectiles, so it has mass $M + (n - r + 1)m$ and speed u_{r-1} . Afterwards, the truck with $n - r$ projectiles has speed u_r , and the fired projectile has velocity $u_{r-1} - v$. Momentum is conserved:

$$(M + (n - r + 1)m)u_{r-1} = (M + (n - r)m)u_r + m(u_{r-1} - v).$$

The mu_{r-1} terms rearrange to give $(M + (n - r)m)(u_r - u_{r-1}) = mv$, which is (*). ■

Adding up the increments,

$$u_n = \sum_{r=1}^n \frac{mv}{M + (n-r)m} = \frac{mv}{M} + \frac{mv}{M+m} + \cdots + \frac{mv}{M + (n-1)m}.$$

Every term after the first is smaller than $\frac{mv}{M}$, and $n > 1$, so $u_n < \frac{nmv}{M} = u$. ■



Pause and try Why does firing one at a time do worse here? When the first gun fires, it has to accelerate the truck *and* the $n - 1$ unfired projectiles, so it produces less of a speed increase than it would on the bare truck.

STAGE 3 The energy bill



Ask yourself Write down the change in kinetic energy at the r th firing. How can (*) help you simplify the term involving $u_r^2 - u_{r-1}^2$?

The kinetic energy changes by

$$K_r - K_{r-1} = \frac{1}{2}(M + (n-r)m)u_r^2 + \frac{1}{2}m(u_{r-1} - v)^2 - \frac{1}{2}(M + (n-r+1)m)u_{r-1}^2.$$

The last two terms combine to $\frac{1}{2}m(v^2 - 2vu_{r-1}) - \frac{1}{2}(M + (n-r)m)u_{r-1}^2$. Then, factorising a difference of squares and using (*),

$$\frac{1}{2}(M + (n-r)m)(u_r^2 - u_{r-1}^2) = \frac{1}{2}mv(u_r + u_{r-1}).$$

So

$$K_r - K_{r-1} = \frac{1}{2}mv(u_r + u_{r-1}) + \frac{1}{2}mv^2 - mvu_{r-1} = \frac{1}{2}mv^2 + \frac{1}{2}mv(u_r - u_{r-1}). \quad \blacksquare$$

Summing from $r = 1$ to n , the increments $u_r - u_{r-1}$ telescope to u_n :

$$K_n = \frac{1}{2}nmv^2 + \frac{1}{2}mvu_n. \quad \blacksquare$$

Finally, $K = \frac{1}{2}nmv^2 + \frac{1}{2}nmv^2 \cdot \frac{nm}{M} = \frac{1}{2}nmv^2 + \frac{1}{2}mvu$, so $K_n < K$ because $u_n < u$. ■

Looking back

- **Momentum for each step.** Each firing was a small two-body problem: the truck with its load, and one projectile.
- **Telescoping sums.** The increments (*) and the energy increments added up neatly.
- **Read the conclusion carefully.** Firing all at once gives more speed, but costs more energy. The question compares two processes that do different amounts of work.

What's really going on

What should be held fixed?

A real gun, or a rocket engine, throws its projectile at a fixed speed relative to *itself*, measured just after the projectile leaves. In this question the reference is the truck's speed *before* firing. For a small projectile the difference is tiny, but when all n projectiles are fired at once it matters a great deal: in part (i), a heavy load of projectiles is thrown back at speed v as if the truck did not recoil at all, which needs a lot of energy.

If instead every projectile leaves at speed v relative to the truck *after* firing, the conclusion reverses. All at once gives $u = \frac{nmv}{M+nm}$; one at a time gives $u_n = \sum_{k=1}^n \frac{mv}{M+km}$, which is larger, because each term is at least $\frac{mv}{M+nm}$. Now firing one at a time wins: each projectile pushes off from a truck that is already moving, and each later gun has less mass to accelerate. (Check this in the exercises.)

The rocket equation

Now let the projectiles become very small and very numerous, with total mass $\mu = nm$ fixed: this is a rocket burning fuel continuously. The sum for u_n becomes an integral (both conventions give the same limit):

$$u_n = \sum_{k=0}^{n-1} \frac{mv}{M+km} \rightarrow \int_0^\mu \frac{v ds}{M+s} = v \ln \frac{M+\mu}{M}.$$

This is *Tsiolkovsky's rocket equation*: a rocket whose exhaust leaves at speed v relative to the rocket gains speed

$$\Delta u = v \ln \frac{m_{\text{initial}}}{m_{\text{final}}}.$$

Figure 10.3.1 compares the three answers. The logarithm is the problem: to reach the orbital speed of about 7.8 km/s with an exhaust speed of about 4.5 km/s (good chemical fuel), the mass ratio must be $e^{7.8/4.5} \approx 5.7$, and once air resistance and gravity are included it is nearer 10 or more. Almost all of a rocket is fuel.

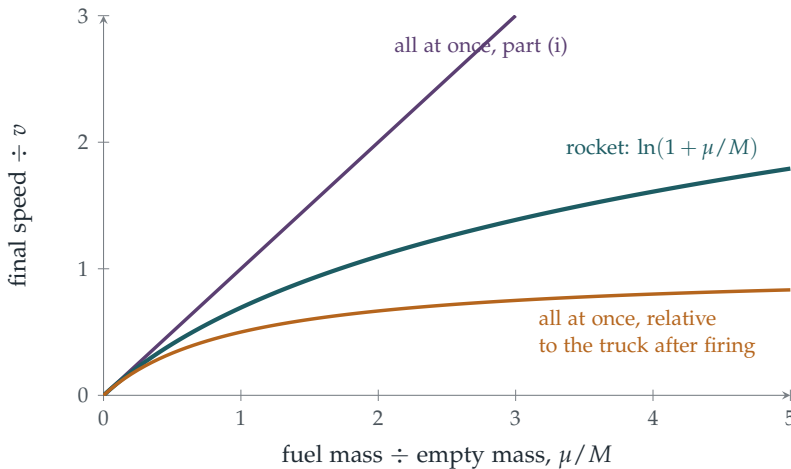


Figure 10.3.1 Final speed against the ratio of fuel to empty mass. Firing everything at once gives μ/M (plum) with the question's convention, but only $\frac{\mu/M}{1+\mu/M}$ (amber) if the exhaust speed is measured relative to the recoiling truck. The rocket, firing continuously, lies in between at $\ln(1 + \mu/M)$ (teal). The picture is the inequality $\frac{x}{1+x} \leq \ln(1+x) \leq x$.

Why rockets have stages

The logarithm also explains staging. A rocket that drops its empty fuel tanks part way up no longer has to accelerate their dead mass. Since the gains of successive stages *add* (like the guns), while each stage's gain depends on the *ratio* of its masses, splitting a rocket into stages lets it reach speeds that a single stage of the same total mass cannot. The Saturn V that took astronauts to the Moon had three stages; its first stage burned about 2000 tonnes of fuel in under three minutes.

Tsiolkovsky

Konstantin Tsiolkovsky, a deaf schoolteacher in the provincial Russian town of Kaluga, published the rocket equation in 1903, in a paper entitled "Exploration of outer space by means of reaction devices". He worked out multi-stage rockets, liquid hydrogen and oxygen fuel, and space stations, decades before any of them could be built. The equation had appeared earlier in less prominent places (William Moore derived a version in 1810), but Tsiolkovsky was the first to see what it meant for spaceflight. "The Earth is the cradle of humanity," he wrote, "but one cannot live in the cradle forever."

The thinking skill

From a discrete process to a limit

A continuous process can often be understood as the limit of many small discrete steps. Solve the discrete problem exactly, then let the step size shrink: sums become integrals.

Here n guns became a rocket and a sum became a logarithm. At university you will see the same move whenever a physical law is derived: a small element is analysed, and the results are added up into an integral or a differential equation.

Try this next

- 1 **The other convention.** Redo parts (i) and (ii) supposing each projectile leaves at speed v relative to the truck *after* firing. Show that now firing one at a time gives the larger final speed.
- 2 **A comet in a dust cloud.** (From STEP III, 2009.) A comet of mass M enters a stationary dust cloud at speed V ; after travelling a distance x its mass is $M(1 + bx)$. If no other force acts, show that $(1 + bx)v = V$ and find x in terms of t . If a constant force Mf acts in the direction of motion, use force = rate of change of momentum to show that $v = \frac{ft+V}{1+bx}$. Show that the comet can move at constant speed if $bV^2 = f$, and that otherwise its speed tends to $\sqrt{f/b}$.
- 3 **Two stages.** A rocket has exhaust speed v , empty mass M and fuel mass $9M$. Find its final speed. Now split it into two stages, each with empty mass $\frac{1}{2}M$ and fuel $4.5M$, the first stage being dropped when empty. Compare the final speeds.
- 4 **The inequality.** Prove that $\frac{x}{1+x} \leq \ln(1+x) \leq x$ for $x \geq 0$, by comparing $\frac{1}{1+s}$ with 1 and with $\frac{1}{1+x}$ on $[0, x]$.

Further reading

- **Daniel Kleppner & Robert Kolenkow, *An Introduction to Mechanics*** (Cambridge University Press), Chapter 4. Momentum, variable mass and the rocket equation.
- **George Sutton & Oscar Biblarz, *Rocket Propulsion Elements*** (Wiley). The standard engineering text; the opening chapters are readable after this section.
- **Hermann Oberth, *Ways to Spaceflight*** (NASA translation, free online). A pioneer's own account.

10.4 A Molecule and a Piston

The question

A bicycle pump consists of a cylinder and a piston. The piston is pushed in with steady speed u . A particle of air moves to and fro between the piston and the end of the cylinder, colliding perfectly elastically with the piston and the end of the cylinder, and always moving parallel with the axis of the cylinder. Initially, the particle is moving towards the piston at speed v . Show that the speed, v_n , of the particle just after the n th collision with the piston is given by $v_n = v + 2nu$.

Let d_n be the distance between the piston and the end of the cylinder at the n th collision, and let t_n be the time between the n th and $(n + 1)$ th collisions. Express $d_n - d_{n+1}$ in terms of u and t_n , and show that

$$d_{n+1} = \frac{v + (2n - 1)u}{v + (2n + 1)u} d_n.$$

Express d_n in terms of d_1 , u , v and n .

In the case $v = u$, show that $ut_n = \frac{d_1}{n(n + 1)}$.

The hook

Pump up a bicycle tyre and feel the pump: it gets hot. Squeezing a gas warms it, and this is the principle behind diesel engines (which ignite their fuel by compression alone), refrigerators and the formation of stars. But why does compression heat a gas? Temperature is the average kinetic energy of the molecules, so the question is: why do the molecules speed up?

This question answers it with a single molecule. Each time it bounces off the advancing piston, it picks up a little extra speed, just as a tennis ball does off a moving racket. Following the bounces exactly reveals a hidden law: the molecule's speed times the length of the cylinder stays almost constant. That is an *adiabatic invariant*, and it is the microscopic origin of the gas laws you meet in physics.

Before you start



Ask yourself A ball moving at speed w hits a wall that is moving towards it at speed u , and bounces elastically. What is its speed afterwards? (Think in the wall's frame.)

In the wall's frame, the ball approaches at $w + u$ and, bouncing elastically off a very heavy object, leaves at $w + u$. Back in the ground frame, the wall is moving at u in the same direction as the rebounding ball, so the ball's speed is $w + 2u$. Each bounce off the advancing piston adds $2u$. (Bouncing off the stationary end of the cylinder changes nothing.)

Working it through

STAGE 1 The speed after n bounces

The particle starts at speed v . Each collision with the piston adds $2u$, and collisions with the end do not change the speed. So after n collisions with the piston,

$$v_n = v + 2nu. \quad \blacksquare$$

STAGE 2 A recurrence for the length



Ask yourself Between the n th and $(n + 1)$ th collisions, the particle goes from the piston to the end and back. How far does it travel? How far does the piston move?

In time t_n the piston moves ut_n , so $d_n - d_{n+1} = ut_n$. Meanwhile the particle, at speed v_n , travels from the piston to the end (distance d_n) and back to the piston (distance d_{n+1}): so $d_n + d_{n+1} = v_n t_n$. Dividing one equation by the other eliminates t_n :

$$\frac{d_n + d_{n+1}}{d_n - d_{n+1}} = \frac{v_n}{u}, \quad \text{so} \quad d_{n+1}(v_n + u) = d_n(v_n - u).$$

With $v_n = v + 2nu$,

$$d_{n+1} = \frac{v + (2n - 1)u}{v + (2n + 1)u} d_n. \quad \blacksquare$$

STAGE 3 Telescoping



Ask yourself Write d_n as d_1 times a product of $n - 1$ such fractions. What cancels?

$$d_n = d_1 \cdot \frac{v + u}{v + 3u} \cdot \frac{v + 3u}{v + 5u} \cdots \frac{v + (2n - 3)u}{v + (2n - 1)u} = \frac{v + u}{v + (2n - 1)u} d_1. \quad \blacksquare$$

Each numerator cancels the previous denominator, leaving only the first numerator and the last denominator.

STAGE 4 The case $v = u$

Then $d_n = \frac{2u}{2nu}d_1 = \frac{d_1}{n}$, and

$$ut_n = d_n - d_{n+1} = d_1 \left(\frac{1}{n} - \frac{1}{n+1} \right) = \frac{d_1}{n(n+1)}. \quad \blacksquare$$

Adding these up, $u(t_1 + t_2 + \dots) = d_1(1 - \frac{1}{2} + \frac{1}{2} - \frac{1}{3} + \dots) = d_1$: the piston reaches the end of the cylinder at time d_1/u after the first collision, as it must. Before it gets there, the particle bounces infinitely often (Figure 10.4.1).

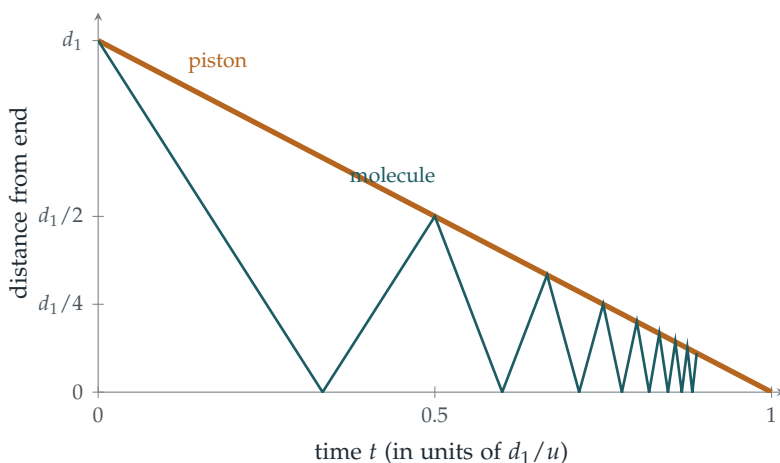


Figure 10.4.1 Space-time diagram for the case $v = u$: distance from the end of the cylinder against time. The piston (amber) advances steadily; the molecule (teal) bounces between it and the end. Its zigzags get steeper (faster) and shorter. The n th collision with the piston happens at distance d_1/n , so the molecule's speed is roughly proportional to $1/(\text{length})$.

Looking back

- **Change frame for a collision.** In the piston's frame the bounce is trivial.
- **Two equations, eliminate the unknown.** The distances travelled by the particle and the piston in the same time gave a recurrence without t_n .
- **Telescoping.** The product collapsed, just as the sum of $\frac{1}{n(n+1)}$ did.

What's really going on

An adiabatic invariant

Multiply the speed by the length:

$$v_n d_n = (v + 2nu) \cdot \frac{v + u}{v + (2n - 1)u} d_1 \approx (v + u) d_1$$

when n is large, or when u is much smaller than v . So

$$\text{speed} \times \text{length} \approx \text{constant}.$$

If the piston moves slowly compared with the molecule, the product vd hardly changes, however far the piston moves. Quantities like this, which stay nearly constant when a system is changed slowly, are called *adiabatic invariants*. In general, the adiabatic invariant of a back-and-forth motion is $\oint p dq$, the momentum integrated round one cycle; here that is $mv \times 2d$.

From one molecule to a gas

Since $v \propto 1/d$, the kinetic energy $\frac{1}{2}mv^2$ is proportional to $1/d^2$. For a gas of many molecules moving along one direction, temperature is proportional to the average kinetic energy, so $T \propto 1/d^2$: halving the length quadruples the temperature. In three dimensions the energy is shared between three directions of motion, and the same argument (which Feynman gives in his *Lectures*) leads to the law of adiabatic compression of a monatomic gas,

$$TV^{2/3} = \text{constant}, \quad \text{or equivalently} \quad pV^{5/3} = \text{constant}.$$

The exponent $\frac{5}{3}$ is the famous ratio γ of specific heats for a monatomic gas. Compressing air in a bicycle pump to a fifth of its volume would, if done quickly and with no heat escaping, raise its absolute temperature by a factor of $5^{0.4} \approx 1.9$ (air molecules have two atoms, giving $\gamma = 1.4$ instead): from 20°C to about 280°C . The pump does not get that hot only because the compression is slow and heat leaks away.

Fermi's cosmic rays

In 1949 Enrico Fermi proposed that cosmic rays, particles arriving from space with enormous energies, are accelerated exactly as the molecule in this question is: by bouncing repeatedly off moving magnetic fields in interstellar gas clouds, gaining a little energy each time. “Fermi acceleration” is now thought to operate at the shock waves of supernova remnants, where particles cross the shock back and forth like a molecule between a piston and a wall. Adiabatic invariants also played a role in the birth of quantum mechanics: Paul Ehrenfest argued in 1916 that the quantities that are quantised must be adiabatic invariants, since a slow change cannot make a system jump between quantum states.

The thinking skill

Build a recurrence from a physical process

When a process repeats, describe one cycle exactly and write down how the state changes from one cycle to the next.

Here the state was the pair (speed, length) at each collision. The recurrence could be solved exactly, and its solution revealed an approximately conserved quantity. At university, the same approach (one cycle, then a recurrence or a “return map”) is used for orbits, oscillators and the iterated maps of Section 8.7.

Try this next

- 1 The invariant exactly.** Show that $v_n d_n - (v + u) d_1 = \frac{u(v+u)d_1}{v+(2n-1)u}$, and deduce that $v_n d_n \rightarrow (v + u) d_1$ as $n \rightarrow \infty$.
- 2 Pulling the piston out.** Repeat the question with the piston withdrawn at speed $u < v$. Show that the particle slows down, and find $v_n d_n$. (Expanding gases cool.)
- 3 Energy.** In the case $v = u$, the molecule’s speed grows without limit as the piston approaches the end. Find the average force on the piston during the n th cycle (change of momentum divided by t_n) and show that it grows like n^3 . What does this say about the work needed to push the piston all the way in?
- 4 A bouncing ball.** A ball bounces elastically on a floor under gravity, and the floor is raised slowly. Guess an adiabatic invariant for this motion. (*Hint:* the analogue of vd is speed times the time between bounces, or the height to the power $\frac{3}{2}$.)

Further reading

- **Richard Feynman, Robert Leighton & Matthew Sands, *The Feynman Lectures on Physics*, Vol. I** (free online), Chapter 39, “The kinetic theory of gases”. The same piston argument, extended to three dimensions.
- **Stephen Blundell & Katherine Blundell, *Concepts in Thermal Physics*** (Oxford University Press). Adiabatic processes and kinetic theory at first-year level.
- **Lev Landau & Evgeny Lifshitz, *Mechanics*** (Butterworth-Heinemann), §49. A concise treatment of adiabatic invariants, for later.

10.5 Infinitely Many Bounces in Finite Time

The question

A particle moves on a smooth triangular horizontal surface AOB with angle $AOB = 30^\circ$. The surface is bounded by two vertical walls OA and OB and the coefficient of restitution between the particle and the walls is e , where $e < 1$. The particle, which is initially at point P on the surface and moving with velocity u_1 , strikes the wall OA at M_1 , with angle $PM_1A = \theta$, and rebounds, with velocity v_1 , to strike the wall OB at N_1 , with angle $M_1N_1B = \theta$. Find e and $\frac{v_1}{u_1}$ in terms of θ .

The motion continues, with the particle striking side OA at $M_2, M_3, \dots$ and striking side OB at $N_2, N_3, \dots$. Show that, if $\theta < 60^\circ$, the particle reaches O in a finite time.

The hook

Zeno of Elea argued that you can never reach a wall: first you must cover half the distance, then half of what remains, and so on for ever. The resolution, that infinitely many steps can take a finite total time, is an infinite series that converges. This question is Zeno's paradox made real: a puck sliding into the corner of a wedge bounces off the walls infinitely many times, yet arrives at the corner after a finite time.

Nature does this too. A ball bouncing on the floor with a coefficient of restitution less than 1 bounces infinitely often in a finite time and then comes to rest. In computer simulations of granular materials (sand, grain, snow) this "inelastic collapse" is a real difficulty: the computer is asked to handle infinitely many collisions in a finite time.

Before you start

Ask yourself When a particle hits a smooth wall with coefficient of restitution e , what happens to the component of velocity along the wall? And the component perpendicular to it?

A smooth wall exerts no force along itself, so the component of velocity parallel to the wall is unchanged. The perpendicular component is reversed and multiplied by e . So if the particle arrives at angle α to the wall and leaves at angle β , then

$$\tan \beta = e \tan \alpha,$$

and since the parallel component is unchanged, $(\text{speed in})\cos \alpha = (\text{speed out})\cos \beta$.

Working it through

STAGE 1 The angle of departure



Ask yourself The particle arrives at M_1 at angle θ to the wall OA (measured on the side away from O). At what angle φ to OA does it leave? Use the triangle OM_1N_1 to find φ in terms of θ .

The particle leaves M_1 heading towards O 's side, at angle φ to the wall, so the angle OM_1N_1 is φ . In triangle OM_1N_1 , the angle at O is 30° . At N_1 , the angle M_1N_1B is θ and O, N_1, B lie on a line, so the angle ON_1M_1 is $180^\circ - \theta$. The angles of the triangle add to 180° :

$$30^\circ + \varphi + (180^\circ - \theta) = 180^\circ, \quad \text{so} \quad \varphi = \theta - 30^\circ.$$

(In particular, $\theta > 30^\circ$.) The restitution law $\tan \varphi = e \tan \theta$ then gives

$$e = \frac{\tan(\theta - 30^\circ)}{\tan \theta},$$

and conservation of the parallel component, $u_1 \cos \theta = v_1 \cos \varphi$, gives

$$\frac{v_1}{u_1} = \frac{\cos \theta}{\cos(\theta - 30^\circ)}. \quad \blacksquare$$

STAGE 2 The pattern repeats



Ask yourself At N_1 , the particle arrives at angle θ to the wall OB , measured on the side away from O . Compare this with its arrival at M_1 . What happens next?

The situation at N_1 is the mirror image of the situation at M_1 : the particle meets a wall at angle θ , on the side away from O . So it leaves at angle $\theta - 30^\circ$ and meets OA again at angle θ , and so on for ever. Every bounce is a scaled copy of the first. Each triangle $OM_kN_k, ON_kM_{k+1}, \dots$ is similar to the one before (Figure 10.5.1), and the speed is multiplied by the same factor at every bounce.

STAGE 3 Two ratios



Ask yourself By what factor ρ does the distance from O shrink from one bounce to the next? By what factor k does the speed shrink? So by what factor does the time for each leg shrink?

The sine rule in triangle OM_1N_1 gives

$$\rho = \frac{ON_1}{OM_1} = \frac{\sin \varphi}{\sin(180^\circ - \theta)} = \frac{\sin(\theta - 30^\circ)}{\sin \theta}.$$

All lengths in the next triangle are ρ times those in this one, so each leg of the path is ρ times as long as the previous leg. The speed is multiplied by $k = \frac{\cos \theta}{\cos(\theta-30^\circ)}$ at each bounce. Time is length divided by speed, so each leg takes

$$\frac{\rho}{k} = \frac{\sin(\theta-30^\circ) \cos(\theta-30^\circ)}{\sin \theta \cos \theta} = \frac{\sin(2\theta-60^\circ)}{\sin 2\theta}$$

times as long as the previous one.

STAGE 4 A geometric series

The total time is $T_1(1 + r + r^2 + \dots)$, where T_1 is the time for the first leg and $r = \frac{\sin(2\theta-60^\circ)}{\sin 2\theta}$. This is finite if $0 < r < 1$.



Ask yourself When is $\sin(2\theta - 60^\circ) < \sin 2\theta$? Use the identity $\sin X - \sin Y = 2 \cos \frac{1}{2}(X + Y) \sin \frac{1}{2}(X - Y)$.

$$\sin 2\theta - \sin(2\theta - 60^\circ) = 2 \cos(2\theta - 30^\circ) \sin 30^\circ = \cos(2\theta - 30^\circ),$$

which is positive when $2\theta - 30^\circ < 90^\circ$, that is, when $\theta < 60^\circ$. And $r > 0$ since $\theta > 30^\circ$. So for $30^\circ < \theta < 60^\circ$ the series converges: the particle makes infinitely many bounces, its distance from O shrinks by the factor $\rho < 1$ each time, and it reaches O after the finite time $\frac{T_1}{1-r}$. ■

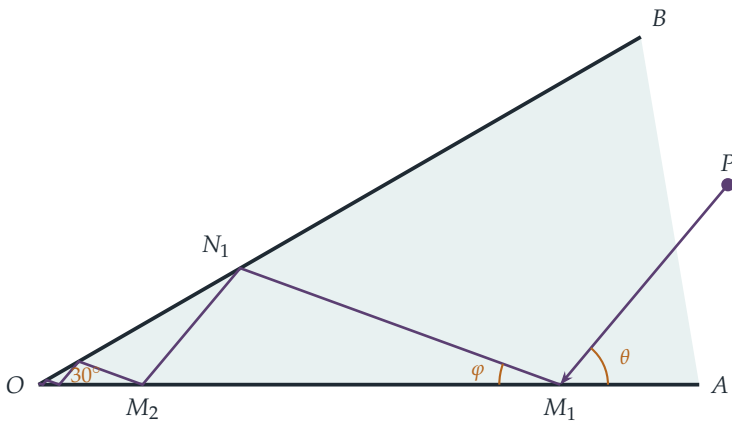


Figure 10.5.1 The path in the wedge with $\theta = 50^\circ$, so $\varphi = 20^\circ$ and $e = \tan 20^\circ / \tan 50^\circ \approx 0.31$. Each bounce is a copy of the previous one, scaled by $\rho \approx 0.45$, and the time for each leg is about 0.65 times that of the leg before. The particle reaches O after infinitely many bounces, in a finite time.

Looking back

- **Self-similarity.** Once one bounce was understood, every later bounce was a scaled copy, so the whole infinite process reduced to two ratios.

- **Combine ratios correctly.** Distance shrinks by ρ and speed by k , so time shrinks by ρ/k .
- **Infinitely many events, finite total.** A geometric series with ratio less than 1 converges.

What's really going on

When $\theta \geq 60^\circ$

At $\theta = 60^\circ$, the ratio is exactly 1: every leg takes the same time, and the particle never reaches O , although it gets ever closer. (The speed and the leg lengths both shrink by the same factor.) For $\theta > 60^\circ$, the legs take longer and longer. In every case the particle approaches O and makes infinitely many bounces; only for $\theta < 60^\circ$ does it get there in finite time. The value $e = \tan 30^\circ / \tan 60^\circ = \frac{1}{3}$ is the dividing line.

Bouncing balls and inelastic collapse

The simplest example of the same phenomenon: a ball dropped onto a floor rises to height e^2H , then e^4H , and so on. The time between bounces is proportional to the speed at each bounce, which shrinks by the factor e , so the total time is a geometric series: with $T_0 = \sqrt{2H/g}$ the time of the first fall, the ball stops bouncing after

$$T_0 + 2eT_0 + 2e^2T_0 + \cdots = T_0 \frac{1+e}{1-e}.$$

Infinitely many bounces, and then rest. In 1992 Sean McNamara and William Young showed that the same thing happens in simulations of many inelastic particles in a line: a group of them can undergo infinitely many collisions in a finite time, an “inelastic collapse”. Simulations of sand and granular flows must include special rules to avoid it, and the phenomenon helps explain why granular gases spontaneously form clusters.

Zeno and convergence

Zeno's paradoxes, from around 450 BC, troubled philosophers for two thousand years. The mathematical resolution, that an infinite sum can have a finite value, was made precise only in the nineteenth century, when Cauchy, Weierstrass and others defined convergence by means of limits (the subject of Chapter 2). The surprising feature of this question is that the process is *physically* infinite: there really are infinitely many collisions, in a finite time. Mathematicians studying billiards in polygons and the motion of colliding particles call such behaviour a *Zeno phenomenon*, and it is one of the subtleties in proving that systems of colliding hard spheres have well-defined motions for all time.

The thinking skill

Sum an infinite process

When a process repeats with a constant scale factor, the total of any additive quantity (time, distance, energy) is a geometric series. Decide convergence from the ratio.

Here the key was to find the right ratio: not the ratio of distances, nor of speeds, but of times. At university, the same reasoning governs the convergence of iterative methods, the theory of fractals, and the analysis of algorithms that repeatedly halve a problem.

Try this next

- 1 The total distance.** Show that the total distance travelled from M_1 to O is finite for every θ between 30° and 90° , and find it in terms of OM_1 and θ .
- 2 Other wedges.** Repeat the question for a wedge of angle α instead of 30° . Show that $\varphi = \theta - \alpha$, and find the condition on θ for the particle to reach O in a finite time.
- 3 Elastic walls.** If $e = 1$, can the particle ever reach O ? (*Hint:* reflect the wedge repeatedly in its walls, so that the path becomes a straight line. How many copies of a 30° wedge fit round O ?)
- 4 The bouncing ball.** A ball is dropped from height 1 m with $e = 0.8$. How long does it bounce for, and how far does it travel in total? (Take $g = 9.8 \text{ m s}^{-2}$.)

Further reading

- **David Morin, *Introduction to Classical Mechanics*** (Cambridge University Press), Chapter 5. Includes several problems on bouncing balls and infinite sequences of collisions.
- **Sean McNamara & William Young, “Inelastic collapse and clumping in a one-dimensional granular medium”, *Physics of Fluids A* 4 (1992).**
- **Serge Tabachnikov, *Geometry and Billiards*** (American Mathematical Society; free online). The mathematics of billiards, including the unfolding trick in the exercises.

10.6 Two Bodies Become One

The question

Particles P , of mass 2, and Q , of mass 1, move along a line. Their distances from a fixed point are x_1 and x_2 , respectively where $x_2 > x_1$. Each particle is subject to a repulsive force from the other of magnitude $\frac{2}{z^3}$, where $z = x_2 - x_1$.

Initially, $x_1 = 0$, $x_2 = 1$, Q is at rest and P moves towards Q with speed 1. Show that z obeys the equation $\frac{d^2z}{dt^2} = \frac{3}{z^3}$.

By first writing $\frac{d^2z}{dt^2} = v \frac{dv}{dz}$, where $v = \frac{dz}{dt}$, show that $z = \sqrt{4t^2 - 2t + 1}$.

By considering the equation satisfied by $2x_1 + x_2$, find x_1 and x_2 in terms of t .

The hook

Two bodies attract or repel each other: the Earth and the Moon, two stars, a proton and an electron. Each body's motion depends on where the other one is, so the equations are coupled, and it looks as if both motions must be found at once. Newton's great simplification was to change variables. Instead of the two positions, use the *centre of mass*, which moves as if no forces acted at all, and the *separation*, which moves like a single particle with a modified mass. One hard problem becomes two easy ones.

This question carries out that separation for two particles on a line. Along the way there is a small miracle: the separation turns out to be the square root of a quadratic in t , for a reason that has nothing to do with the question as set and everything to do with motion in a plane.

Before you start



Ask yourself Write down Newton's second law for each particle. In which direction does each force act?

The force is repulsive, so it pushes P (on the left) to the left and Q (on the right) to the right:

$$2\ddot{x}_1 = -\frac{2}{z^3}, \quad 1 \cdot \ddot{x}_2 = +\frac{2}{z^3}.$$

These are coupled, because each involves $z = x_2 - x_1$.

Working it through

STAGE 1 The separation

? **Ask yourself** Which combination of the two equations involves only z ?

Subtract: $\ddot{z} = \ddot{x}_2 - \ddot{x}_1 = \frac{2}{z^3} + \frac{1}{z^3} = \frac{3}{z^3}$. ■

The separation behaves like a single particle of mass $\frac{2}{3}$ pushed by the force $\frac{2}{z^3}$. The number $\frac{2}{3} = \frac{2 \times 1}{2+1}$ is the *reduced mass* $\frac{m_1 m_2}{m_1 + m_2}$.

STAGE 2 Energy for the separation

? **Ask yourself** With $\dot{z} = v \frac{dv}{dz}$, separate the variables and integrate. What are the initial values of z and $v = \dot{z}$?

From $v \frac{dv}{dz} = \frac{3}{z^3}$, integrating gives $\frac{1}{2}v^2 = -\frac{3}{2z^2} + C$. Initially $z = 1$ and $\dot{z} = \dot{x}_2 - \dot{x}_1 = 0 - 1 = -1$, so $\frac{1}{2} = -\frac{3}{2} + C$, giving $C = 2$:

$$v^2 = 4 - \frac{3}{z^2}.$$

(This is conservation of energy for the separation: kinetic energy plus potential energy $\frac{3}{2z^2}$ per unit reduced mass is constant.)

STAGE 3 A surprisingly simple solution

? **Ask yourself** The answer to show is $z^2 = 4t^2 - 2t + 1$. So look at $w = z^2$ instead of z . What is $\dot{w}^2$ in terms of w ?

With $w = z^2$, $\dot{w} = 2z\dot{z}$, so

$$\dot{w}^2 = 4z^2\dot{z}^2 = 4z^2\left(4 - \frac{3}{z^2}\right) = 16w - 12.$$

Differentiating, $2\dot{w}\ddot{w} = 16\dot{w}$, so (whenever $\dot{w} \neq 0$) $\ddot{w} = 8$. Hence w is a quadratic in t : $w = 4t^2 + \alpha t + \beta$. Initially $w = 1$ and $\dot{w} = 2z\dot{z} = -2$, so $\beta = 1$ and $\alpha = -2$:

$$z = \sqrt{4t^2 - 2t + 1}. \quad \blacksquare$$

(You can check directly: $\dot{w}^2 = (8t - 2)^2 = 64t^2 - 32t + 4 = 16(4t^2 - 2t + 1) - 12$.) The particles come closest at $t = \frac{1}{4}$, when $z = \frac{\sqrt{3}}{2}$, and then separate for ever.

STAGE 4 The centre of mass

Ask yourself What equation does $2x_1 + x_2$ satisfy? Why is it the natural combination to look at?

Adding the two equations of motion, $2\ddot{x}_1 + \ddot{x}_2 = -\frac{2}{z^3} + \frac{2}{z^3} = 0$. The forces are equal and opposite (Newton's third law), so the total momentum $2\dot{x}_1 + \dot{x}_2$ is constant. It starts at $2 \times 1 + 0 = 2$, and $2x_1 + x_2$ starts at 1, so

$$2x_1 + x_2 = 1 + 2t.$$

The centre of mass, at $\frac{2x_1 + x_2}{3}$, moves steadily at speed $\frac{2}{3}$. Combining with $x_2 - x_1 = z$,

$$x_1 = \frac{1 + 2t - \sqrt{4t^2 - 2t + 1}}{3}, \quad x_2 = \frac{1 + 2t + 2\sqrt{4t^2 - 2t + 1}}{3}. \quad \blacksquare$$

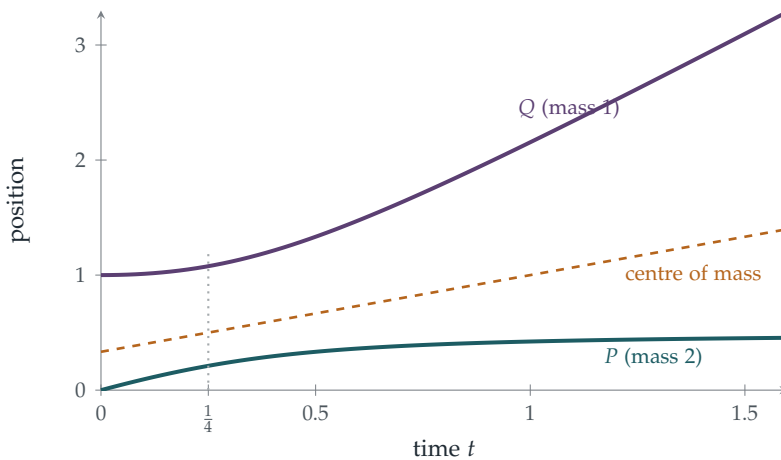


Figure 10.6.1 Positions of P (teal) and Q (plum) against time. P moves towards Q , is slowed and turned back by the repulsion, and approaches a constant position, while Q is pushed away. Their closest approach is at $t = \frac{1}{4}$. The centre of mass (amber, dashed) moves at constant speed throughout.

Looking back

- **Decouple by changing variables.** The difference of the equations involved only z ; the weighted sum involved only the centre of mass.
- **Energy integral.** Writing $\dot{z} = v \frac{dv}{dz}$ turned a second-order equation into a first-order one: conservation of energy.
- **Change the unknown.** The equation for z was awkward; the equation for z^2 was trivial.

What's really going on

Why z^2 is quadratic

For any force law of the form $\ddot{z} = k/z^3$, the same calculation gives $\frac{d^2}{dt^2}(z^2) = 2\dot{z}^2 + 2z\ddot{z} = 2\dot{z}^2 + \frac{2k}{z} = \text{constant}$, since $\frac{1}{2}\dot{z}^2 + \frac{k}{2z^2}$ is the conserved energy. So z^2 is always a quadratic in t . Why?

Here is the hidden reason. Think of a particle moving freely, in a straight line at constant speed, in a *plane*. Its distance r from the origin satisfies $r^2 = (\text{a quadratic in } t)$, since $r^2 = (x_0 + at)^2 + (y_0 + bt)^2$. And in polar coordinates its radial equation of motion is $\ddot{r} - r\dot{\theta}^2 = 0$; with the angular momentum $h = r^2\dot{\theta}$ constant, that becomes

$$\ddot{r} = \frac{h^2}{r^3}.$$

The inverse-cube repulsion is exactly the *centrifugal* effect of the hidden angular motion. In our problem, $z^2 = 4t^2 - 2t + 1 = (2t - \frac{1}{2})^2 + \frac{3}{4}$ is the squared distance from the origin of a point moving at speed 2 along a straight line that passes the origin at distance $\frac{\sqrt{3}}{2}$. The one-dimensional repelling particles are the shadow of free motion in two dimensions.

The two-body problem

The method of this question works for any force between two bodies that depends only on their separation and acts along the line between them, in three dimensions as well as one. The centre of mass moves uniformly, and the separation vector $\mathbf{r}$ obeys $\mu\ddot{\mathbf{r}} = \mathbf{F}(\mathbf{r})$, with $\mu = \frac{m_1 m_2}{m_1 + m_2}$. For gravity this is the *Kepler problem*, solved by Newton in the *Principia*: the separation traces an ellipse, parabola or hyperbola. The Earth and the Sun both orbit their common centre of mass, which lies inside the Sun; the Earth and the Moon orbit a point about 4700 km from the Earth's centre, inside the Earth but well away from its centre, which is why the Earth wobbles monthly. Astronomers find planets round other stars by detecting exactly this wobble in the star.

Integrable systems

In 1969 Francesco Calogero studied many particles on a line repelling each other with inverse-square potentials (that is, inverse-cube forces), and found that, remarkably, the system can be solved exactly for any number of particles. Jürgen Moser explained the reason in 1975, using a matrix whose eigenvalues stay constant as it evolves: the *Lax pairs* of Section 7.7. The *Calogero–Moser system* is now a central example of an *integrable system*, a system with enough conserved quantities to be solved completely, and it connects mechanics with random matrices, quantum field theory and the theory of solitons.

The thinking skill

Choose coordinates that decouple

When two quantities are coupled, look for combinations that are not. Sums and differences, weighted by the right constants, often separate a problem into independent pieces.

Centre of mass and relative coordinates are the first example. At university the same idea appears as diagonalising a matrix (normal modes, Section 11.2), as separation of variables for partial differential equations, and in quantum mechanics.

Try this next

- 1 Two masses on a string.** (From STEP III, 2023.) Particles A (mass m) and B (mass M) are joined by a light string of length r on a smooth table. Initially A is at the origin and B at $(r, 0)$; B is at rest and A is given velocity u in the y direction. Writing θ for the angle of $\overrightarrow{AB}$ clockwise from the x -axis, show that $\theta = ut/r$, and find $my + MY$ in terms of t . Show that $y = \frac{1}{m+M}(mut + Mr \sin \frac{ut}{r})$, and that if $M > m$ the y -velocity of A is sometimes negative.
- 2 Attraction.** Repeat the question with an attractive force $\frac{2}{z^3}$. Show that z^2 is still a quadratic in t , and find when the particles collide.
- 3 The hidden plane.** For the motion in the question, find the straight line in the plane whose distance from the origin is $z(t)$, and the angle $\theta(t)$ swept out. Check that $z^2\dot{\theta}$ is constant and that $(z^2\dot{\theta})^2 = 3$.
- 4 Kepler's third law.** Two stars of masses m_1 and m_2 move in circles about their centre of mass, a distance a apart, under gravity. Show that the period T satisfies
$$T^2 = \frac{4\pi^2 a^3}{G(m_1 + m_2)}.$$

Further reading

- **John Taylor, *Classical Mechanics*** (University Science Books), Chapter 8. The two-body problem, reduced mass and orbits.
- **Daniel Kleppner & Robert Kolenkow, *An Introduction to Mechanics*** (Cambridge University Press), Chapter 10, on central forces.
- **Francesco Calogero, *Classical Many-Body Problems Amenable to Exact Treatments*** (Springer). For the ambitious: the world of exactly solvable systems.

10.7 Gravity Inside a Shell

The question

The gravitational force between two point particles of masses m and m' is mutually attractive and has magnitude

$$\frac{Gmm'}{r^2},$$

where G is a constant and r is the distance between them.

A particle of unit mass lies on the axis of a thin uniform circular ring of radius r and mass m , at a distance x from its centre. Explain why the net force on the particle is directed towards the centre of the ring and show that its magnitude is

$$\frac{Gmx}{(x^2 + r^2)^{3/2}}.$$

The particle now lies inside a thin hollow spherical shell of uniform density, mass M and radius a , at a distance b from its centre. Show that the particle experiences no gravitational force due to the shell.

The hook

Imagine a hollow planet: a thin spherical shell of rock with nothing inside. Standing inside it, near the wall, you might expect to be pulled towards the nearby rock. In fact you would float: the gravitational pull of the shell is exactly zero everywhere inside it. The near side of the shell is closer, but the far side is bigger, and the two effects cancel perfectly.

This is *Newton's shell theorem*. It took Newton years to prove, and it was essential to his theory of gravity: together with its partner (outside a sphere, the pull is as if all the mass were at the centre) it justifies treating the Earth and the Sun as points. In the exercises and the What's Really Going On section, the theorem is used to compute how long a cloud of gas takes to collapse into a star.

Before you start



Ask yourself Consider two small, diametrically opposite pieces of the ring. Draw the forces they exert on the particle. Which components cancel, and which add?

The two pieces are at the same distance from the particle. Their pulls have equal components along the axis, which add, and equal and opposite components perpendicular to the axis, which cancel. Pairing up the whole ring in this way, the net force is along the axis, towards the centre of the ring.

Working it through

STAGE 1 The ring

Every piece of the ring is at distance $\sqrt{x^2 + r^2}$ from the particle, and the line to it makes angle ψ with the axis, where $\cos \psi = \frac{x}{\sqrt{x^2 + r^2}}$. A piece of mass δm pulls with force $\frac{G \delta m}{x^2 + r^2}$, of which the component along the axis is $\frac{G \delta m}{x^2 + r^2} \cdot \frac{x}{\sqrt{x^2 + r^2}}$. Every piece has the same distance and the same angle, so the components simply add up to

$$\frac{Gmx}{(x^2 + r^2)^{3/2}}. \quad \blacksquare$$

STAGE 2 Slicing the shell into rings



Ask yourself Put the centre of the shell at O and the particle on the axis at distance b from O . Slice the shell into thin rings perpendicular to the axis, labelled by the angle ϕ that a point of the ring makes with the axis at O . What are the radius, the position and the mass of the ring?

The ring at angle ϕ has radius $a \sin \phi$ and its centre is at distance $a \cos \phi$ from O along the axis, so at distance $a \cos \phi - b$ from the particle. Its width is $a \delta \phi$, so its area is $2\pi a \sin \phi \cdot a \delta \phi$, a fraction $\frac{1}{2} \sin \phi \delta \phi$ of the shell's area $4\pi a^2$. Its mass is $\frac{1}{2} M \sin \phi \delta \phi$. By Stage 1, it pulls the particle along the axis, towards the $\phi = 0$ end, with force

$$\frac{G \cdot \frac{1}{2} M \sin \phi \delta \phi (a \cos \phi - b)}{((a \cos \phi - b)^2 + a^2 \sin^2 \phi)^{3/2}} = \frac{GM}{2} \cdot \frac{(a \cos \phi - b) \sin \phi \delta \phi}{(a^2 - 2ab \cos \phi + b^2)^{3/2}}.$$

(The sign takes care of itself: rings beyond the particle, with $a \cos \phi > b$, pull one way, and rings behind it pull the other.)

STAGE 3 The right variable



Ask yourself The denominator involves $s^2 = a^2 - 2ab \cos \phi + b^2$. What does s represent geometrically? What is ds in terms of $d\phi$?

By the cosine rule, s is the distance from the particle to the ring. Differentiating, $2s \, ds = 2ab \sin \phi \, d\phi$, and solving $s^2 = a^2 - 2ab \cos \phi + b^2$ for $a \cos \phi$ gives $a \cos \phi - b = \frac{a^2 - b^2 - s^2}{2b}$. So the total force is

$$\frac{GM}{2} \int \frac{a^2 - b^2 - s^2}{2bs^3} \cdot \frac{s \, ds}{ab} = \frac{GM}{4ab^2} \int_{a-b}^{a+b} \left(\frac{a^2 - b^2}{s^2} - 1 \right) ds,$$

where s runs from $a - b$ (the nearest point, $\phi = 0$) to $a + b$ (the farthest, $\phi = \pi$).

STAGE 4 It vanishes

$$\int_{a-b}^{a+b} \left(\frac{a^2 - b^2}{s^2} - 1 \right) ds = (a^2 - b^2) \left(\frac{1}{a-b} - \frac{1}{a+b} \right) - 2b = (a^2 - b^2) \cdot \frac{2b}{a^2 - b^2} - 2b = 0.$$

The shell exerts no force on the particle. ■



Pause and try What changes if the particle is *outside* the shell, with $b > a$? The limits of s become $b - a$ and $b + a$. Redo the last line and show that the force is $\frac{GM}{b^2}$ towards the centre, as if all the mass were concentrated there.

Looking back

- **Symmetry first.** Pairing opposite pieces showed the direction of the force before any calculation.
- **Build up from simpler shapes.** Points made a ring; rings made a shell.
- **Choose the variable that matches the physics.** Changing from the angle ϕ to the distance s turned an awkward integral into a trivial one.

What's really going on

Newton's cone argument

Newton's own proof (*Principia*, Book I, Proposition 70) uses no integration at all. Draw a narrow double cone with its vertex at the particle, cutting the shell in two small patches (Figure 10.7.1). The areas of the patches are proportional to the squares of their distances from the particle, r_1^2 and r_2^2 (the patches make equal angles with the cone, since a chord meets a circle at equal angles at its two ends). So their masses are proportional to r_1^2 and r_2^2 , and their pulls, proportional to mass divided by distance squared, are *equal* and opposite. Every double cone through the particle cancels, so the

total force is zero. The inverse-square law is exactly what makes the cancellation work: with any other power of distance, the shell would push or pull.

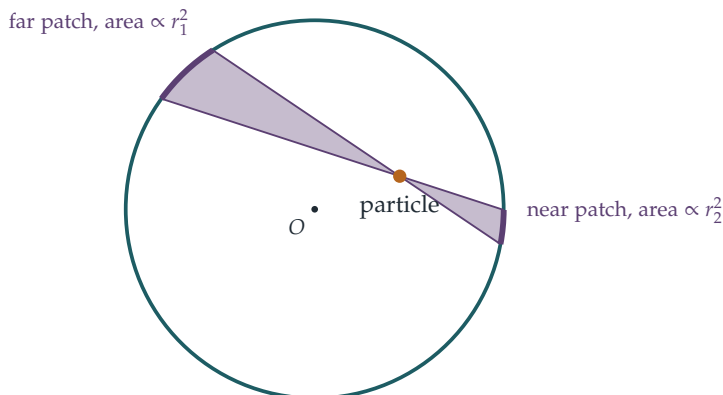


Figure 10.7.1 Newton's argument. A thin double cone from the particle cuts the shell in two patches. The far patch is bigger, by the ratio of the squared distances, but its pull is weaker by exactly the same ratio. The two pulls cancel, and so does every other pair: inside the shell, gravity vanishes.

Collapsing clouds

STEP III 1987 Q14 uses the theorem the other way round. Outside a uniform sphere, the pull on a particle at the surface is GMm/r^2 , as if the whole mass were at the centre. Now take a uniform spherical cloud of dust, at rest, of radius R and mass M , and let it collapse under its own gravity. A particle on the surface feels only the mass inside it, which stays M as everything falls inwards, so its inward speed $v = -\dot{r}$ satisfies

$$v \frac{dv}{dr} = -\frac{GM}{r^2}, \quad \text{so} \quad v^2 = 2GM \left(\frac{1}{r} - \frac{1}{R} \right).$$

Substituting $r = R \cos^2 \alpha$ (one of the exercises) gives the time to reach radius r as $\sqrt{\frac{R^3}{2GM}} \left(\alpha + \frac{1}{2} \sin 2\alpha \right)$. The whole cloud collapses to a point when $\alpha = \frac{\pi}{2}$, after the *free-fall time*

$$t_{\text{ff}} = \frac{\pi}{2} \sqrt{\frac{R^3}{2GM}} = \sqrt{\frac{3\pi}{32G\rho}},$$

where ρ is the density. Remarkably, every shell of the cloud reaches the centre at the same moment, since the time depends only on the density (Figure 10.7.2). For a typical interstellar cloud, with about 10^{-17} kg per cubic metre, t_{ff} is around 700 000 years: this sets the timescale on which stars form.

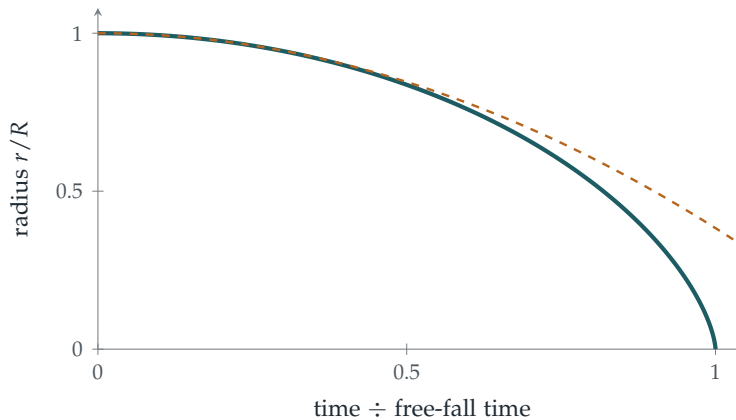


Figure 10.7.2 Radius of a collapsing uniform dust cloud against time (teal). It starts slowly, like a dropped stone (the dashed amber parabola is the constant-acceleration approximation $r = R - \frac{1}{2} \frac{GM}{R^2} t^2$), then accelerates sharply and reaches zero at $t = t_{\text{ff}}$. The curve is half of a cycloid.

Newton's delay and Gauss's law

Newton had the inverse-square law by the 1660s, but did not publish his theory of gravity until 1687. One reason, according to some historians, was that he could not justify treating the Earth as a point mass until he had proved the shell theorem; once he had, the rest of the *Principia* followed. Today the theorem is a one-line consequence of *Gauss's law*, which says that the total gravitational field flowing out through any closed surface is proportional to the mass inside. Applied to a sphere inside the shell, it encloses no mass, so the field is zero. The same law governs electric fields, which is why there is no electric field inside a charged metal shell, and why your car protects you from lightning.

The thinking skill

Integration as summation

To find the effect of a continuous body, cut it into pieces whose effects you know, and add them up. Use symmetry to choose the pieces.

Points made a ring, rings made a shell, shells make a solid sphere. At university this is how every field is computed from its sources (gravitational, electric, magnetic), and it leads to the vector calculus of Gauss's and Stokes's theorems.

Try this next

- 1 Outside the shell.** Complete the Pause box: show that outside the shell the force is $\frac{GM}{b^2}$. Deduce that a solid uniform sphere of mass M pulls an outside particle as if its mass were at the centre.
- 2 The collapse time.** (From STEP III, 1987.) Using $v^2 = 2GM(\frac{1}{r} - \frac{1}{R})$ and the substitution $r = R \cos^2 \alpha$, show that the time to reach radius r is $\sqrt{\frac{R^3}{2GM}}(\alpha + \frac{1}{2} \sin 2\alpha)$.
- 3 A tunnel through the Earth.** Inside a uniform solid sphere, show that the force at distance b from the centre is $\frac{GMb}{a^3}$, proportional to b . Deduce that a ball dropped down a tunnel through the centre of the Earth oscillates with period $2\pi\sqrt{a^3/(GM)}$, about 84 minutes, the same as the period of a satellite skimming the surface.
- 4 Another force law.** If gravity followed an inverse-cube law, would the force inside a shell still be zero? (Repeat Stage 3 with $\frac{1}{s^3}$ in place of $\frac{1}{s^2}$.)

Further reading

- **Daniel Kleppner & Robert Kolenkow, *An Introduction to Mechanics*** (Cambridge University Press), the chapters on gravitation, including the shell theorem.
- **Subrahmanyan Chandrasekhar, *Newton's Principia for the Common Reader*** (Oxford University Press). Newton's proofs, including Propositions 70 and 71, explained by a Nobel laureate.
- **Andrew Pontzen, *The Universe in a Box*** (Riverhead). How gravity, collapse and simulation explain the structure of the universe.

10.8 Three Stars in a Triangle

The question

The force of attraction between two stars of masses m_1 and m_2 a distance r apart is $\gamma m_1 m_2 / r^2$. The Starmakers of Kryton place three stars of equal mass m at the corners of an equilateral triangle of side a . Show that it is possible for each star to revolve round the centre of mass of the system with angular velocity $(3\gamma m / a^3)^{1/2}$.

Find a corresponding result if the Starmakers place a fourth star, of mass λm , at the centre of mass of the system.

The hook

Two bodies attracting each other by gravity can be solved completely, as Newton showed and as Section 10.6 explained: they move in ellipses about their centre of mass. Add a third body and everything changes. The *three-body problem* has no general solution in terms of familiar functions, and in 1890 Henri Poincaré discovered that its solutions can be chaotic, the first appearance of chaos in mathematics.

Yet some special solutions are perfectly simple. In 1772 Joseph-Louis Lagrange found that three bodies at the corners of an equilateral triangle can rotate rigidly, the triangle turning like a wheel. This question finds that solution for equal masses. Lagrange's triangle is not just a curiosity: the Sun, Jupiter and thousands of "Trojan" asteroids form exactly this configuration, and space telescopes are parked at similar points today.

Before you start



Ask yourself For a star to move in a circle of radius R with angular velocity ω , what must the total force on it be?

It must have magnitude $m\omega^2 R$ and point towards the centre of the circle. So we need to check that the total gravitational force on each star points at the centre of the triangle, and find the ω that makes its size right.

Working it through

STAGE 1 The geometry

Ask yourself How far is each corner of an equilateral triangle of side a from its centre?

The centre lies on each median, two-thirds of the way from the corner. A median has length $a \sin 60^\circ = \frac{\sqrt{3}}{2}a$, so the distance from each corner to the centre is

$$R = \frac{2}{3} \cdot \frac{\sqrt{3}}{2}a = \frac{a}{\sqrt{3}}.$$

By symmetry the centre of the triangle is the centre of mass of three equal stars.

STAGE 2 The force on one star

Ask yourself Each of the other two stars pulls with force $\frac{\gamma m^2}{a^2}$. What is the angle between the two pulls? What is their resultant, and in which direction does it point?

The two pulls point along the two sides of the triangle that meet at the star, so the angle between them is 60° . By symmetry their resultant lies along the bisector of that angle, which passes through the centre. Each pull makes 30° with the bisector, so the resultant has magnitude

$$2 \cdot \frac{\gamma m^2}{a^2} \cos 30^\circ = \frac{\sqrt{3} \gamma m^2}{a^2},$$

pointing at the centre (Figure 10.8.1).

STAGE 3 The angular velocity

Setting the resultant equal to $m\omega^2 R$:

$$m\omega^2 \frac{a}{\sqrt{3}} = \frac{\sqrt{3} \gamma m^2}{a^2}, \quad \text{so} \quad \omega^2 = \frac{3\gamma m}{a^3}.$$

With this angular velocity, the same calculation applies to each star (by symmetry), so all three move round the centre in circles, keeping the triangle rigid. ■

STAGE 4 A fourth star at the centre

Ask yourself A star of mass λm at the centre is at distance $\frac{a}{\sqrt{3}}$ from each corner. What extra force does it exert on each corner star? Does the centre of mass move? Does the central star feel any net force?

The central star does not change the centre of mass, and by symmetry the three pulls on it cancel, so it stays at rest. It pulls each corner star towards the centre with force $\frac{\gamma \lambda m^2}{(a/\sqrt{3})^2} = \frac{3\gamma \lambda m^2}{a^2}$. The total inward force on a corner star is now $(\sqrt{3} + 3\lambda) \frac{\gamma m^2}{a^2}$, so

$$m\omega^2 \frac{a}{\sqrt{3}} = (\sqrt{3} + 3\lambda) \frac{\gamma m^2}{a^2}, \quad \omega^2 = \frac{(3 + 3\sqrt{3}\lambda) \gamma m}{a^3} = \frac{3\gamma m}{a^3} (1 + \sqrt{3}\lambda). \quad \blacksquare$$

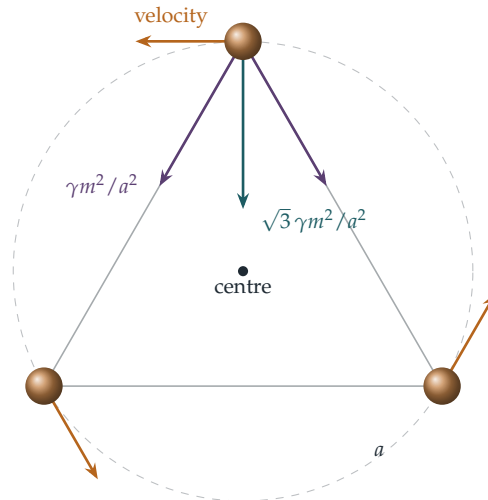


Figure 10.8.1 Three equal stars at the corners of an equilateral triangle. The two pulls on the top star (plum) combine into a single force towards the centre (teal). If each star moves on the dashed circle with the right speed (amber), the whole triangle rotates rigidly. This is Lagrange's equilateral solution of the three-body problem.

Looking back

- **Guess a symmetric solution.** We did not solve the three-body problem; we guessed a highly symmetric motion and checked that Newton's laws allow it.
- **Symmetry fixes directions.** The resultant force had to lie along the bisector, so only its size needed calculating.
- **Check every body.** By symmetry, verifying one star verified all three; the central star needed a separate (easy) check.

What's really going on

Kepler's law in disguise

Write the answer as $\omega^2 a^3 = \gamma(3m)$. For two bodies in circular orbit a distance a apart, Kepler's third law (an exercise in Section 10.6) says $\omega^2 a^3 = \gamma(m_1 + m_2)$, the total mass. The three stars obey the same law, with the total mass $3m$. In fact Lagrange showed that the equilateral solution exists for *any* three masses, not just equal ones: the triangle rotates about the centre of mass with $\omega^2 a^3 = \gamma(m_1 + m_2 + m_3)$. Euler had found a second family a few years earlier, with the three bodies on a rotating straight line. These five configurations (three collinear, two triangular, up to labelling) are the only ones in which three bodies keep the same shape.

Lagrange points

Take m_1 to be the Sun, m_2 a planet, and m_3 a tiny asteroid. The equilateral solution says the asteroid can sit at the third corner of an equilateral triangle with the Sun and the planet, ahead of or behind the planet in its orbit. These are the *Lagrange points* L_4 and L_5 ; Euler's collinear solutions give L_1 , L_2 and L_3 . Jupiter has over ten thousand known "Trojan" asteroids clustered around its L_4 and L_5 points. The James Webb Space Telescope orbits near the Sun–Earth L_2 point, 1.5 million km beyond the Earth, where it can keep the Sun, Earth and Moon all behind its sunshield.

Stability

Would the Starmakers' triangle last? For equal masses, no: the slightest disturbance grows, and the stars fly apart or collide. Edward Routh showed in 1875 that Lagrange's triangle is stable only if one mass dominates, precisely when

$$27(m_1m_2 + m_2m_3 + m_3m_1) < (m_1 + m_2 + m_3)^2.$$

For the Sun and Jupiter this holds comfortably, which is why the Trojans survive. For three equal masses, $27 \times 3m^2 > 9m^2$, so the configuration is unstable.

Poincaré's mistake and the figure eight

In 1887 King Oscar II of Sweden offered a prize for a solution of the many-body problem. Henri Poincaré won it in 1889 with a long memoir on the three-body problem, but while the memoir was being printed an editor found a gap. In correcting it, Poincaré discovered that some solutions behave with "a complexity so astonishing that I cannot even attempt to draw it": the first glimpse of chaos. The printed copies were recalled and destroyed, at Poincaré's expense. New special solutions are still being found: in 1993 Cristopher Moore discovered numerically, and in 2000 Alain Chenciner and Richard Montgomery proved, that three equal masses can chase each other round a figure-eight curve. Thousands more periodic orbits have since been found by computer.

The thinking skill

Guess a symmetric solution and verify it

When a problem is too hard to solve in general, look for solutions with as much symmetry as possible. A symmetric guess reduces many unknowns to one or two, and checking it is often easy.

This is a powerful strategy in mathematics and physics: symmetric equilibria, steady states, travelling waves and rigidly rotating configurations are often the first solutions found, and the starting points for studying the rest. The next

question is always whether the symmetric solution is stable.

Try this next

- 1 Four stars in a square.** Four stars of mass m are placed at the corners of a square of side a . Show that they can rotate rigidly about the centre with $\omega^2 = \frac{(4+\sqrt{2})\gamma m}{2a^3}$.
- 2 Two equal stars and a planet.** Two stars of mass M orbit their centre of mass, a distance a apart. A planet of negligible mass sits at the centre. Is it in equilibrium? Is it stable to small displacements along the line joining the stars? Perpendicular to it?
- 3 Unequal masses.** Three stars of masses m_1, m_2, m_3 are at the corners of an equilateral triangle of side a , rotating about their centre of mass G . Show that the gravitational force on m_1 is $\frac{\gamma m_1(m_1+m_2+m_3)}{a^3} \overrightarrow{P_1 G}$, where P_1 is its position. (*Hint:* write the forces as vectors and use the definition of the centre of mass.) Deduce Lagrange's result.
- 4 n stars.** n equal stars sit at the corners of a regular n -gon inscribed in a circle of radius R . Find ω^2 for rigid rotation as a sum over k of $\frac{1}{\sin(k\pi/n)}$.

Further reading

- **June Barrow-Green, *Poincaré and the Three Body Problem*** (American Mathematical Society). The story of the prize and Poincaré's discovery of chaos.
- **Richard Montgomery, "A new solution to the three-body problem"**, *Notices of the American Mathematical Society* 48 (2001). The figure eight, explained by one of its discoverers; free online.
- **Florin Diacu & Philip Holmes, *Celestial Encounters*** (Princeton University Press). A popular history of the three-body problem and chaos.

What We Have Learned

We began by saying that some things never change, and that knowing which ones is half of physics. We asked three questions: what conservation laws can decide on their own, how small steps add up, and how to tame many bodies. Here is what the chapter's eight questions showed.

1. What can conservation laws decide on their own?

For two bodies colliding elastically in a line, momentum and energy determine the outcome completely, and that single result was used again and again. For more bodies, they do not. In Section 10.1, a row of marbles had many outcomes consistent with both laws, and it took a further physical fact (marbles cannot pass through each other) to rule some out. The square of the momentum minus the energy gave a sum of cross terms that had to vanish, an inequality argument that proved certain motions impossible. And the familiar behaviour of Newton's cradle turned out to depend on details beyond the conservation laws: how the shock travels through the balls.

Conservation laws also exposed hidden assumptions. In Section 10.3, firing guns all at once gave more speed than firing them one at a time, but only because it used more energy; holding the right quantity fixed reversed the conclusion.

2. How do small steps add up?

Several sections repeated one simple step. In Section 10.2, a collision that can at most double a speed became, through a chain of lighter and lighter balls, unlimited gain: the principle behind the basketball and tennis ball and behind gravitational slingshots. In Section 10.3, many small firings became an integral, and Tsiolkovsky's logarithm, which explains why rockets are mostly fuel and why they have stages. In Section 10.4, each bounce off a moving piston added $2u$ to a molecule's speed, and an exact recurrence revealed that speed times length stays almost constant: an adiabatic invariant, and the reason compressed gases heat up. And in Section 10.5, infinitely many bounces took a finite time, because the times formed a convergent geometric series.

3. How do we tame many bodies?

By choosing coordinates and exploiting symmetry. In Section 10.6, the centre of mass moved uniformly and the separation behaved like a single particle with the reduced mass; a hidden plane even explained why the separation squared was a quadratic in time. In Section 10.7, symmetry and integration showed that a spherical shell exerts no gravity inside and acts as a point mass outside, which let a collapsing cloud of dust be treated as a single equation, with a free-fall time that sets the pace of star formation. And

in Section 10.8, a symmetric guess solved a special case of the unsolvable three-body problem: Lagrange’s rotating triangle, realised in the Trojan asteroids.

4. A toolkit for mechanics

Method	Where you used it
Conservation laws as constraints	Newton’s cradle (10.1)
Change of frame	Collisions (10.2); piston (10.4)
Iterating one result	Stacked balls (10.2)
Momentum for variable mass	Rockets (10.3)
Sum to integral	Rocket equation (10.3); shell (10.7)
Recurrence and telescoping	Piston (10.4)
Geometric series	Wedge (10.5)
Centre of mass and reduced mass	Two bodies (10.6)
Energy integral $v dv/dz$	Two bodies (10.6); collapse (10.7)
Symmetric guess	Three stars (10.8)

Where next

- Chapter 11 turns from conservation to oscillation and rotation: resonance, normal modes, the real pendulum, and motion in accelerating and rotating frames.
- The energy method of Section 8.2 and the limit cycles of Section 8.4 are the differential-equation side of the same ideas.
- The chaos Poincaré found in the three-body problem is the subject of Section 8.7.

If you want to go further before university

- **Daniel Kleppner & Robert Kolenkow, *An Introduction to Mechanics*:** the classic first-year text; demanding but rewarding.
- **David Morin, *Introduction to Classical Mechanics*:** hundreds of problems, many of them as surprising as the ones in this chapter.
- **Richard Feynman, *The Character of Physical Law*:** lectures for a general audience on conservation laws, symmetry and gravity.
- **Dwight Neuenschwander, *Emmy Noether’s Wonderful Theorem*:** why symmetry implies conservation.

Feynman said that there is no known exception to the conservation of energy. This chapter has tried to show how much they can do: rule out impossible outcomes, reveal hidden invariants, and turn hopelessly complicated systems into ones we can understand. But it has also shown their limits. They constrain, but they do not always decide; and where they stop, the details of forces, symmetries and frames must take over.