

CHAPTER

11

Oscillations, Rotation and Moving Frames

Things that swing, things that spin, and things seen from somewhere that moves.

IN THIS CHAPTER

11.1	Resonance	STEP II 1991
11.2	Normal Modes	STEP III 2014
11.3	The Real Pendulum	STEP III 1998
11.4	The Tautochrone	STEP III 1994
11.5	Gravity in an Accelerating Train	STEP II 2021
11.6	Backspin	STEP III 1997
11.7	Simpson's Rule in a Rod	STEP II 1989
11.8	Tilted Equilibria	STEP III 2025
11.9	The Safety Parabola	STEP II 2014
11.10	A Ball That Comes Back Towards You	STEP II 2019
11.11	How Fielders Catch	STEP II 1998
11.12	The Ant on the Rubber Rope	STEP II 2018
11.13	The Snowplough Problem	STEP III Specimen paper
11.14	Parabola, Not Catenary	STEP I 1991

Motion, Seen Clearly

The career of a young theoretical physicist consists of treating the harmonic oscillator in ever-increasing levels of abstraction.

— Attributed to Sidney Coleman

The previous chapter used conservation laws to understand collisions and gravity. This one turns to the other great themes of a first course in dynamics: things that oscillate, things that rotate, and the art of describing motion from the most useful point of view.

Oscillation is everywhere. Pendulums and springs are the textbook examples, but the same mathematics describes the vibrations of molecules, the swaying of buildings, the circuits in a radio and, in quantum mechanics, the behaviour of light itself. The reason is that almost anything near a stable equilibrium behaves like a spring: pull it a little way and the restoring force is proportional to the displacement. So the simple harmonic oscillator is the first approximation to almost every physical system, and much of physics consists of understanding what happens when the approximation is not quite right.

Why this matters

Engineers must know the natural frequencies of every structure they build, to keep them away from the frequencies of wind, traffic and earthquakes. Chemists identify molecules by their vibrations. Timekeeping, from Huygens's pendulum clocks to quartz watches and atomic clocks, is the art of making a perfect oscillator. Rotation governs everything from the spin of a snooker ball to the stability of satellites. And the question of which frames of reference are "natural" led, through the equivalence of gravity and acceleration, to Einstein's general relativity.

The chapter ends with six questions grouped as *surprising trajectories*: projectiles, ants, fielders and snowploughs whose behaviour is not what intuition predicts. They are short, elegant problems, and each has a proper mathematical idea at its centre: envelopes, monotonicity, feedback, divergence, linear differential equations and the difference between a parabola and a catenary.

What you will meet at university

A first-year dynamics course covers simple, damped and forced harmonic motion, resonance, and coupled oscillators and their normal modes (where linear algebra and mechanics meet). It treats rigid bodies, moments of inertia and rolling, and motion in accelerating and rotating frames, with the centrifugal and Coriolis forces. Later courses develop all of this into Lagrangian and Hamiltonian mechanics, where the

choice of coordinates becomes a central theme, and into the theory of waves, which are oscillations spread out in space. Nonlinear oscillations, like the large-amplitude pendulum, lead to elliptic functions and to the dynamical systems of Chapter 8.

The questions that run through this chapter

Three big questions

1. **When is motion simple?** When does a system oscillate with a single, fixed frequency, and what happens when it does not?
2. **What is the right point of view?** Which coordinates, which variables and which frame of reference make a problem easy?
3. **Can intuition be trusted?** Where does everyday intuition about motion go wrong, and how does mathematics correct it?

The chapter begins with oscillations: resonance (Section 11.1), normal modes, the real pendulum and the tautochrone. It then turns to frames and rigid bodies: an accelerating train, a spinning marble, a rod that behaves like three particles, and a solid that can rest at three different tilts. The last six sections are the surprising trajectories.

Ideas in this chapter

§	Topic	Idea
11.1	Forced oscillations	Resonance and beats
11.2	Coupled oscillators	Normal modes are eigenvectors
11.3	The real pendulum	Elliptic integrals; the AGM
11.4	The tautochrone	Simple harmonic in arc length
11.5	Moving frames	Fictitious forces; equivalence principle
11.6	Rolling and sliding	A conserved angular momentum
11.7	Moments of inertia	Dynamical equivalence; Simpson's rule
11.8	Stability	Normals through the centre of mass
11.9	Projectiles	The envelope of trajectories
11.10	Projectiles	Distance need not increase
11.11	Catching	A feedback rule
11.12	Stretching ropes	Fractions and the harmonic series
11.13	Snowploughs	Time as a function of distance
11.14	Suspension bridges	Parabola, not catenary

Two sections, Section 11.6 and Section 11.7, use rotational dynamics and moments of inertia, which were removed from the STEP specification in 2019. They are included because they are excellent questions; each gives the facts it needs.

How to read this chapter

Before you start any calculation, ask which variable the problem wants. Arc length for the tautochrone, $x \pm y$ for coupled springs, the fraction of the rope for the ant, time as a function of distance for the snowplough: in each case, the right choice turns a difficult problem into a familiar one.

11.1 Resonance

The question

A particle is attached to one end B of a light elastic string of unstretched length a . Initially the other end A is at rest and the particle hangs at rest at a distance $a + c$ vertically below A . At time $t = 0$, the end A is forced to oscillate vertically, its downwards displacement at time t being $b \sin pt$. Let $x(t)$ be the downwards displacement of the particle at time t from its initial equilibrium position. Show that, while the string remains taut, $x(t)$ satisfies

$$\frac{d^2x}{dt^2} = -n^2(x - b \sin pt),$$

where $n^2 = g/c$, and that if $0 < p < n$, $x(t)$ is given by

$$x(t) = \frac{bn}{n^2 - p^2}(n \sin pt - p \sin nt).$$

Write down a necessary and sufficient condition that the string remains taut throughout the subsequent motion, and show that it is satisfied if $pb < (n - p)c$.

The hook

Push a child on a swing at just the right moments and small pushes build into large swings. Push at the wrong rhythm and nothing much happens. Every oscillating system has a *natural frequency*, and driving it at that frequency produces a response out of all proportion to the effort. This is *resonance*. It is why a singer can shatter a wine glass, why soldiers break step when crossing a bridge, and why a radio can pick out one station from thousands.

This question shakes the top of an elastic string and asks when the string stays taut. The answer contains the factor $\frac{1}{n-p}$, which blows up as the driving frequency p approaches the natural frequency n : a gentle shake near resonance is enough to make the string go slack.

Before you start



Ask yourself The particle hangs in equilibrium with extension c . If the string has modulus λ , what is the relation between λ , a , c , m and g ? What is the extension when the top has moved down by $b \sin pt$ and the particle by x ?

In equilibrium, tension equals weight: $\frac{\lambda c}{a} = mg$. When the top is displaced down by $b \sin pt$ and the particle by x , the string's length increases by $x - b \sin pt$, so the extension is $c + x - b \sin pt$.

Working it through

STAGE 1 The equation of motion

Taking downwards as positive, the forces on the particle are its weight mg and the tension $\frac{\lambda}{a}(c + x - b \sin pt)$ upwards:

$$m\ddot{x} = mg - \frac{\lambda}{a}(c + x - b \sin pt) = -\frac{\lambda}{a}(x - b \sin pt),$$

using $\frac{\lambda c}{a} = mg$. Since $\frac{\lambda}{ma} = \frac{g}{c} = n^2$, this is $\ddot{x} = -n^2(x - b \sin pt)$. ■

The particle is pulled towards the moving point $b \sin pt$ with a spring force. Without the driving, it would oscillate with angular frequency n , its natural frequency.

STAGE 2 Solving



Ask yourself The equation is $\ddot{x} + n^2x = n^2b \sin pt$. Try a particular solution of the form $x = C \sin pt$. Then add the general solution of $\ddot{x} + n^2x = 0$. Which initial conditions does the particle satisfy?

Substituting $x = C \sin pt$ gives $(-p^2 + n^2)C = n^2b$, so $C = \frac{n^2b}{n^2 - p^2}$ (we need $p \neq n$). The complementary function is $D \cos nt + E \sin nt$. The particle starts at rest in equilibrium: $x(0) = 0$ forces $D = 0$, and $\dot{x}(0) = 0$ gives $Cp + En = 0$, so $E = -\frac{nbp}{n^2 - p^2}$. Altogether,

$$x(t) = \frac{bn}{n^2 - p^2}(n \sin pt - p \sin nt). \quad \blacksquare$$

STAGE 3 Staying taut



Ask yourself The string is taut exactly when its extension is positive. Write the extension $c + x - b \sin pt$ in terms of t .

Subtracting $b \sin pt$ from $x(t)$,

$$\begin{aligned} x - b \sin pt &= \frac{b}{n^2 - p^2} (n^2 \sin pt - np \sin nt - (n^2 - p^2) \sin pt) \\ &= \frac{bp}{n^2 - p^2} (p \sin pt - n \sin nt). \end{aligned}$$

So the string remains taut throughout if and only if

$$c + \frac{bp}{n^2 - p^2} (p \sin pt - n \sin nt) \geq 0 \quad \text{for all } t \geq 0.$$

 **Ask yourself** How negative can $p \sin pt - n \sin nt$ possibly be?

Since $|\sin| \leq 1$, we have $p \sin pt - n \sin nt \geq -(p + n)$. So the extension is at least

$$c - \frac{bp(n + p)}{n^2 - p^2} = c - \frac{bp}{n - p},$$

which is positive if $pb < (n - p)c$. ■

 **Pause and try** What happens to this condition as p approaches n ? How small must the shaking amplitude b be when $p = 0.9n$? When $p = 0.99n$?

Looking back

- **Measure from equilibrium.** Using the equilibrium extension c removed gravity from the equation.
- **Particular plus complementary.** The forced response at frequency p plus the free oscillation at frequency n , with constants fixed by the initial conditions.
- **Bound, do not solve.** We did not find the minimum extension exactly; a crude bound on the sines was enough for a sufficient condition.

What's really going on

The resonance limit

Let $p \rightarrow n$ in the solution. Both the numerator and the denominator tend to zero, and a limit calculation (or l'Hôpital's rule, differentiating with respect to p) gives

$$x(t) \rightarrow \frac{b}{2} (\sin nt - nt \cos nt).$$

The term $nt \cos nt$ grows without limit: driving exactly at the natural frequency makes the amplitude increase linearly in time, each push adding to the motion in step with

it. Just below resonance, the two frequencies drift in and out of step, producing *beats*: the amplitude swells and dies away with the slow frequency $n - p$ (Figure 11.1.1). The same beats are heard when two guitar strings are almost in tune.

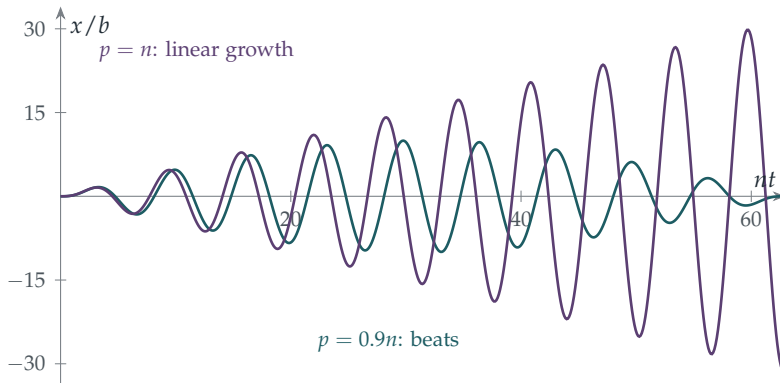


Figure 11.1.1 The response of the particle (ignoring slackness), in units of the shaking amplitude b . With $p = 0.9n$ (teal) the response swells to about ten times b and dies away again, with a beat period of $\frac{2\pi}{n-p}$. At exact resonance, $p = n$ (plum), the amplitude grows without limit.

Damping and the resonance curve

Real systems lose energy, through friction or air resistance, and that limits the growth. With damping, the steady response to a driving force at frequency p has an amplitude that peaks near the natural frequency; the sharper the peak, the less damping there is. Engineers describe this by the *quality factor* Q : roughly, the number of oscillations a system makes before losing most of its energy. A wine glass has Q in the thousands, which is why a sustained note at exactly its natural frequency can break it. A car's suspension is deliberately given a low Q , so that bumps do not set it bouncing.

Bridges that move

In 1831 a column of soldiers marching in step over the Broughton Suspension Bridge near Manchester set it vibrating until it collapsed; since then, troops have been ordered to break step on bridges. The London Millennium Bridge, opened in June 2000, began to sway sideways as soon as crowds crossed it: pedestrians unconsciously adjusted their steps to the sway, feeding it energy in step, and the bridge closed for two years while dampers were fitted. The famous collapse of the Tacoma Narrows Bridge in 1940 is often described as resonance, but it was a different phenomenon, *aeroelastic flutter*, in which the wind's force depends on the bridge's own motion. Its analysis needs the ideas of self-excited oscillation and limit cycles (Section 8.4).

The thinking skill

Superpose a forced response and a free response

For a linear equation driven by a periodic force, the solution is the response at the driving frequency plus a free oscillation at the natural frequency, chosen to fit the starting conditions.

Linearity makes this work: solutions can be added. At university, the same principle, applied to sums of many driving frequencies, leads to Fourier analysis, transfer functions in engineering, and Green's functions in physics.

Try this next

- 1 **A falling lift.** (From STEP II, 1996.) A particle hangs from the ceiling of a lift by an elastic string, extended by a . The lift starts to descend with constant acceleration $f < \frac{g}{2}$. Show that the extension y of the string satisfies $\ddot{y} + \frac{g}{a}y = g - f$, that the string never becomes slack, and that the oscillation has amplitude $\frac{af}{g}$. After time T the lift stops accelerating. Show that the amplitude becomes $\frac{2af}{g} \left| \sin \frac{1}{2}\omega T \right|$, where $\omega^2 = \frac{g}{a}$.
- 2 **At resonance.** Solve $\ddot{x} + n^2x = n^2b \sin nt$ with $x(0) = \dot{x}(0) = 0$ directly, by trying a particular solution $x = Ct \cos nt$. Check that it agrees with the limit above.
- 3 **Beats.** Show that $n \sin pt - p \sin nt$ can be written as a sum of two terms, one of which is approximately $2n \cos \frac{(n+p)t}{2} \sin \frac{(p-n)t}{2}$ when p is close to n . Hence explain the beat period $\frac{2\pi}{n-p}$.
- 4 **Damped resonance.** For $\ddot{x} + 2k\dot{x} + n^2x = F \cos pt$ with small $k > 0$, find the steady-state amplitude, and show that it is greatest when $p^2 = n^2 - 2k^2$.

Further reading

- **A. P. French, *Vibrations and Waves*** (MIT Introductory Physics Series). Chapter 4 on forced oscillations and resonance is the classic account.
- **Steven Strogatz, *Sync*** (Penguin). How things fall into step, from fireflies to the Millennium Bridge.
- **Henry Petroski, *To Engineer Is Human*** (Vintage). Engineering failures, including the bridges in the history box.

11.2 Normal Modes

The question

Two particles X and Y , of equal mass m , lie on a smooth horizontal table and are connected by a light elastic spring of natural length a and modulus of elasticity λ . Two more springs, identical to the first, connect X to a point P on the table and Y to a point Q on the table. The distance between P and Q is $3a$.

Initially, the particles are held so that $XP = a$, $YQ = \frac{1}{2}a$, and $PXYQ$ is a straight line. The particles are then released.

At time t , the particle X is a distance $a + x$ from P and the particle Y is a distance $a + y$ from Q . Show that

$$m \frac{d^2x}{dt^2} = -\frac{\lambda}{a}(2x + y)$$

and find a similar expression involving $\frac{d^2y}{dt^2}$. Deduce that

$$x - y = A \cos \omega t + B \sin \omega t$$

where A and B are constants to be determined and $ma\omega^2 = \lambda$. Find a similar expression for $x + y$.

Show that Y will never return to its initial position.

The hook

One mass on a spring bounces up and down in a simple, regular way. Join two masses with springs and the motion looks complicated: each mass jerks about irregularly, energy sloshing from one to the other. But hidden inside the complication are two perfectly simple motions, the *normal modes*, in which the masses move together or exactly opposite, each mode oscillating at its own frequency. Every motion of the system is a combination of the two.

This question finds the normal modes of two masses between three springs, and then proves something striking: started from rest with one mass displaced, the system *never* returns to its starting position. The reason is that the ratio of the two frequencies is $\sqrt{3}$, and $\sqrt{3}$ is irrational.

Before you start



Ask yourself With X at distance $a + x$ from P and Y at distance $a + y$ from Q , how long is the middle spring XY ? Which springs are stretched and which compressed when x and y are positive?

The three lengths add up to $3a$, so $XY = 3a - (a + x) - (a + y) = a - x - y$. When $x, y > 0$, the outer springs are stretched (by x and y) and the middle one is compressed (by $x + y$).

Working it through

STAGE 1 The equations of motion

Measure x in the direction from P towards Q . The spring PX has extension x , so its tension $\frac{\lambda x}{a}$ pulls X back towards P . The spring XY has extension $-(x + y)$, so its tension $-\frac{\lambda(x+y)}{a}$ pulls X towards Y (a negative tension is a push). So

$$m\ddot{x} = -\frac{\lambda x}{a} - \frac{\lambda(x+y)}{a} = -\frac{\lambda}{a}(2x + y). \quad \blacksquare$$

By symmetry (swap the roles of X and Y , P and Q),

$$m\ddot{y} = -\frac{\lambda}{a}(2y + x).$$

STAGE 2 Decoupling



Ask yourself Each equation involves both x and y . Which combinations of the two equations involve only one combination of x and y ?

Subtracting and adding,

$$m(\ddot{x} - \ddot{y}) = -\frac{\lambda}{a}(x - y), \quad m(\ddot{x} + \ddot{y}) = -\frac{3\lambda}{a}(x + y).$$

Each is the equation of simple harmonic motion: $x - y$ oscillates with $\omega^2 = \frac{\lambda}{ma}$, and $x + y$ with $3\omega^2$.

STAGE 3 The initial conditions

Initially $XP = a$ and $YQ = \frac{1}{2}a$, so $x = 0$ and $y = -\frac{1}{2}a$, and both particles are at rest. So $x - y = \frac{1}{2}a$ and $x + y = -\frac{1}{2}a$ at $t = 0$, with zero rates of change, giving

$$x - y = \frac{1}{2}a \cos \omega t, \quad x + y = -\frac{1}{2}a \cos \sqrt{3}\omega t.$$

So $A = \frac{1}{2}a$ and $B = 0$. ■

STAGE 4 Never returning



Ask yourself Write y in terms of t . For Y to be back at its initial position, $y = -\frac{1}{2}a$. What must both cosines be?

Subtracting, $y = \frac{1}{2}((x+y) - (x-y)) = -\frac{1}{4}a(\cos \sqrt{3}\omega t + \cos \omega t)$. For $y = -\frac{1}{2}a$ we need $\cos \sqrt{3}\omega t + \cos \omega t = 2$, and since each cosine is at most 1, both must equal 1:

$$\omega t = 2\pi k, \quad \sqrt{3}\omega t = 2\pi j$$

for integers k and j . For $t > 0$, $k \geq 1$, and dividing, $\sqrt{3} = \frac{j}{k}$, which is impossible because $\sqrt{3}$ is irrational. So Y never returns to its initial position. ■

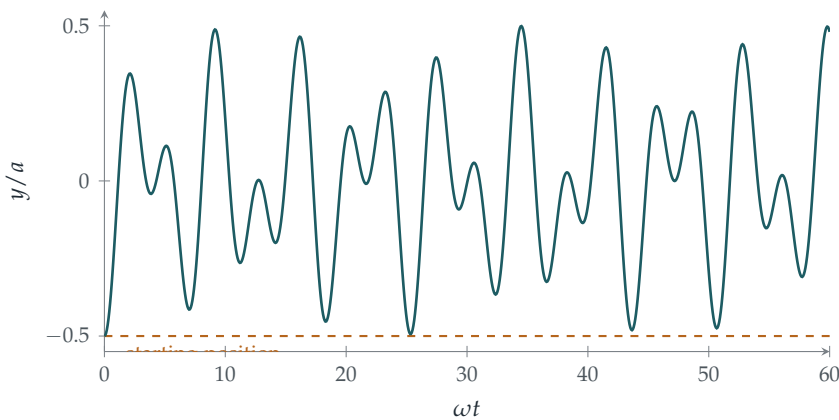


Figure 11.2.1 The displacement y of the particle Y . It is the sum of two oscillations whose frequencies are in the ratio $\sqrt{3} : 1$, so the pattern never repeats. Y comes close to its starting position (dashed) again and again, for instance near $\omega t \approx 25$ (where $\omega t \approx 8\pi$ and $\sqrt{3}\omega t \approx 14\pi$), but never reaches it exactly.

Looking back

- **Find the combinations that decouple.** $x - y$ and $x + y$ each satisfied an equation of their own.
- **Symmetry suggests them.** The system is symmetric under swapping the two sides, so symmetric and antisymmetric combinations are natural.
- **Irrationality has physical consequences.** The motion is not periodic because a frequency ratio is irrational.

What's really going on

Normal modes are eigenvectors

Write the equations as a matrix equation:

$$m \begin{pmatrix} \ddot{x} \\ \ddot{y} \end{pmatrix} = -\frac{\lambda}{a} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

The combinations that decouple are the eigenvectors of the matrix: $(1, -1)$ with eigenvalue 1, and $(1, 1)$ with eigenvalue 3 (Section 7.2). In the first mode, $x = -y$: both particles move the same way along the line (remember that x and y are measured from opposite ends), and the middle spring keeps its length. In the second, $x = y$: they move in opposite directions, squeezing and stretching the middle spring, which makes the frequency higher. The squares of the frequencies are the eigenvalues, multiplied by $\frac{\lambda}{ma}$.

This works for any number of masses and springs. A system of N coupled oscillators has N normal modes, given by the eigenvectors of a symmetric matrix, and the spectral theorem guarantees that they always exist and are perpendicular. Molecules vibrate in normal modes, which is how chemists identify them by the infrared light they absorb. Buildings and bridges have normal modes, which engineers must keep away from the frequencies of wind and earthquakes. And as the number of masses tends to infinity, the normal modes become the harmonics of a vibrating string, the starting point of Fourier analysis.

Coming close without arriving

Although Y never returns exactly, it comes as close as you like. The reason is a theorem of Kronecker: if α is irrational, the multiples $\alpha, 2\alpha, 3\alpha, \dots$, taken modulo 1, come arbitrarily close to every number in $[0, 1]$. So we can find times when ωt is a multiple of 2π and $\sqrt{3}\omega t$ is within any tolerance of one. The motion is *quasi-periodic*: it winds round a torus (one angle for each mode) without ever closing up. The same distinction, between rational and irrational frequency ratios, is at the heart of the KAM theorem on the stability of the solar system.

The thinking skill

Change variables to decouple equations

When equations are coupled, look for combinations of the variables that evolve independently. In linear problems, these are the eigenvectors.

This is the same idea as the centre of mass and relative coordinates of Section 10.6. At university it becomes diagonalisation: choose a basis in which the matrix is diagonal, and a system of coupled equations falls apart into independent ones.

Try this next

- 1 **Energy.** Show that the total energy $\frac{1}{2}m(\dot{x}^2 + \dot{y}^2) + \frac{\lambda}{2a}(x^2 + y^2 + (x + y)^2)$ is constant, and that it splits into a sum of two terms, one for each normal mode, each of which is separately constant.
- 2 **Pure modes.** How should the particles be started so that the motion is periodic, with period $\frac{2\pi}{\omega}$? With period $\frac{2\pi}{\sqrt{3}\omega}$?
- 3 **A stiffer middle spring.** If the middle spring has modulus μ instead of λ , find the two normal-mode frequencies. For which values of μ/λ is every motion periodic?
- 4 **Three masses.** Three equal masses are connected in a line by four identical springs between two fixed points. Find the matrix and its eigenvalues $(2 - \sqrt{2}, 2, 2 + \sqrt{2})$, and describe the three normal modes.

Further reading

- **A. P. French, *Vibrations and Waves*** (MIT Introductory Physics Series), Chapter 5, on coupled oscillators and normal modes.
- **Gilbert Strang, *Introduction to Linear Algebra*** (Wellesley-Cambridge), the chapter on eigenvalues, which includes springs and masses as a main example.
- **Steven Strogatz, *Nonlinear Dynamics and Chaos*** (CRC Press), Chapter 8, on quasi-periodic motion on a torus.

11.3 The Real Pendulum

The question

Consider a simple pendulum of length l and angular displacement θ , which is **not** assumed to be small. Show that

$$\frac{1}{2}l\left(\frac{d\theta}{dt}\right)^2 = g(\cos\theta - \cos\gamma),$$

where γ is the maximum value of θ . Show also that the period P is given by

$$P = 2\sqrt{\frac{l}{g}} \int_0^\gamma (\sin^2(\gamma/2) - \sin^2(\theta/2))^{-\frac{1}{2}} d\theta.$$

By using the substitution $\sin(\theta/2) = \sin(\gamma/2) \sin\phi$, and then finding an approximate expression for the integrand using the binomial expansion, show that for small values of γ the period is approximately

$$2\pi\sqrt{\frac{l}{g}}\left(1 + \frac{\gamma^2}{16}\right).$$

The hook

Every physics student learns that a pendulum's period is $2\pi\sqrt{l/g}$, independent of the size of the swing. Galileo believed this, supposedly after timing a swinging lamp in Pisa cathedral against his pulse, and it is the reason pendulum clocks work. But it is not quite true. The formula comes from the approximation $\sin\theta \approx \theta$, and a pendulum swinging through a wide arc takes longer. At 90° the period is 18% longer; a pendulum released from almost upside down can take as long as you like.

This question finds the exact period as an integral, and then the first correction for moderate swings. The integral cannot be evaluated in elementary functions: it is an *elliptic integral*, and the story of how such integrals were tamed runs from Gauss's arithmetic-geometric mean to the theory of elliptic curves.

Before you start



Ask yourself The bob moves on a circle of radius l . What are its speed and its height (relative to the lowest point) in terms of θ and $\dot{\theta}$?

The speed is $l\dot{\theta}$, and the height above the lowest point is $l(1 - \cos \theta)$. Since the string does no work (its tension is always perpendicular to the motion), energy is conserved.

Working it through

STAGE 1 Energy

At the extreme angle γ the bob is momentarily at rest. Conservation of energy gives

$$\frac{1}{2}ml^2\dot{\theta}^2 + mgl(1 - \cos \theta) = mgl(1 - \cos \gamma),$$

and dividing by ml , $\frac{1}{2}l\dot{\theta}^2 = g(\cos \theta - \cos \gamma)$. ■

STAGE 2 The period as an integral



Ask yourself How can you write $\cos \theta - \cos \gamma$ in terms of half-angles? And how much of the period is spent going from $\theta = 0$ to $\theta = \gamma$?

Using $\cos \theta = 1 - 2\sin^2 \frac{\theta}{2}$, we have $\cos \theta - \cos \gamma = 2(\sin^2 \frac{\gamma}{2} - \sin^2 \frac{\theta}{2})$, so

$$\dot{\theta} = 2\sqrt{\frac{g}{l}} \left(\sin^2 \frac{\gamma}{2} - \sin^2 \frac{\theta}{2} \right)^{1/2}.$$

By symmetry the swing from 0 to γ takes a quarter of the period, so

$$P = 4 \int_0^\gamma \frac{d\theta}{\dot{\theta}} = 2\sqrt{\frac{l}{g}} \int_0^\gamma \left(\sin^2 \frac{\gamma}{2} - \sin^2 \frac{\theta}{2} \right)^{-1/2} d\theta. \quad \blacksquare$$

STAGE 3 Removing the singularity



Ask yourself The integrand becomes infinite at $\theta = \gamma$. Write $k = \sin \frac{\gamma}{2}$ and substitute $\sin \frac{\theta}{2} = k \sin \phi$. What range of ϕ corresponds to $0 \leq \theta \leq \gamma$? What happens to the integrand?

As θ goes from 0 to γ , ϕ goes from 0 to $\frac{\pi}{2}$. Differentiating, $\frac{1}{2} \cos \frac{\theta}{2} d\theta = k \cos \phi d\phi$, so

$$d\theta = \frac{2k \cos \phi d\phi}{\cos \frac{\theta}{2}} = \frac{2k \cos \phi d\phi}{\sqrt{1 - k^2 \sin^2 \phi}}.$$

Meanwhile $\sin^2 \frac{\gamma}{2} - \sin^2 \frac{\theta}{2} = k^2(1 - \sin^2 \phi) = k^2 \cos^2 \phi$. The awkward factors cancel:

$$P = 2\sqrt{\frac{l}{g}} \int_0^{\pi/2} \frac{1}{k \cos \phi} \cdot \frac{2k \cos \phi}{\sqrt{1 - k^2 \sin^2 \phi}} d\phi = 4\sqrt{\frac{l}{g}} \int_0^{\pi/2} \frac{d\phi}{\sqrt{1 - k^2 \sin^2 \phi}}.$$

The new integrand is perfectly well behaved, since $k < 1$.

STAGE 4 The first correction



Ask yourself For small k , expand $(1 - k^2 \sin^2 \phi)^{-1/2}$ by the binomial theorem. What is $\int_0^{\pi/2} \sin^2 \phi d\phi$?

To first order, $(1 - k^2 \sin^2 \phi)^{-1/2} \approx 1 + \frac{1}{2}k^2 \sin^2 \phi$, and $\int_0^{\pi/2} \sin^2 \phi d\phi = \frac{\pi}{4}$. So

$$P \approx 4\sqrt{\frac{l}{g}} \left(\frac{\pi}{2} + \frac{k^2}{2} \cdot \frac{\pi}{4} \right) = 2\pi\sqrt{\frac{l}{g}} \left(1 + \frac{k^2}{4} \right).$$

For small γ , $k = \sin \frac{\gamma}{2} \approx \frac{\gamma}{2}$, so $k^2 \approx \frac{\gamma^2}{4}$ and

$$P \approx 2\pi\sqrt{\frac{l}{g}} \left(1 + \frac{\gamma^2}{16} \right). \quad \blacksquare$$

For a swing of 20° ($\gamma \approx 0.35$) the correction is less than 1%; at 60° it is about 7% (Figure 11.3.1).

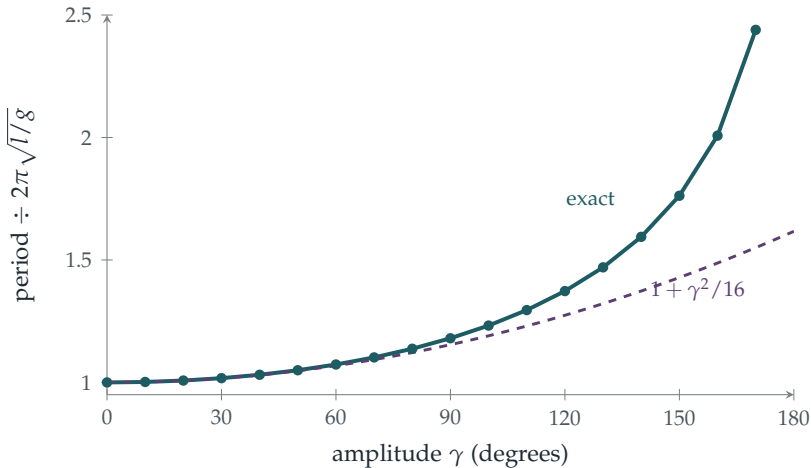


Figure 11.3.1 The exact period of a pendulum (teal), relative to the small-angle value, against its amplitude, compared with the approximation $1 + \gamma^2/16$ (plum, dashed). The approximation is excellent up to about 60° . As $\gamma \rightarrow 180^\circ$ the true period tends to infinity: a pendulum balanced upside down takes for ever to fall.

Looking back

- **Energy gives a first integral.** A second-order equation became a first-order one, and the period became an integral.
- **A substitution tames a singularity.** The integrand blew up at the turning point; the substitution $\sin \frac{\theta}{2} = k \sin \phi$ removed the problem.
- **Exact first, then approximate.** We found an exact formula and expanded it, which also tells us how good the approximation is.

What's really going on

Elliptic integrals

The integral

$$K(k) = \int_0^{\pi/2} \frac{d\phi}{\sqrt{1 - k^2 \sin^2 \phi}}$$

is the *complete elliptic integral of the first kind*, so called because a close relative gives the perimeter of an ellipse. The exact period is $P = 4\sqrt{l/g} K(\sin \frac{\gamma}{2})$. Expanding the integrand fully and integrating term by term using the Wallis integrals $\int_0^{\pi/2} \sin^{2n} \phi d\phi$ of Section 4.4, one finds

$$P = 2\pi\sqrt{\frac{l}{g}} \left(1 + \left(\frac{1}{2}\right)^2 k^2 + \left(\frac{1 \cdot 3}{2 \cdot 4}\right)^2 k^4 + \left(\frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6}\right)^2 k^6 + \dots \right).$$

No finite combination of familiar functions gives $K(k)$; it is a genuinely new function.

Gauss's remarkable formula

There is a far better way to compute K . In Section 2.6 you met the *arithmetic–geometric mean*: start with a and b , and repeatedly replace them by their arithmetic and geometric means. Gauss discovered in 1799 that

$$K(k) = \frac{\pi}{2 \operatorname{AGM}(1, \sqrt{1 - k^2})},$$

so the period of a pendulum with amplitude γ is exactly

$$P = \frac{2\pi\sqrt{l/g}}{\operatorname{AGM}(1, \cos \frac{\gamma}{2})}.$$

The AGM converges so fast that three or four steps give the period to ten significant figures; the values in Figure 11.3.1 were computed this way. Gauss wrote in his diary that this discovery “opens an entirely new field of analysis”, and he was right: inverting elliptic integrals led Abel and Jacobi to *elliptic functions*, and those led to *elliptic curves*, which are now central to number theory (they were the key to the proof of Fermat’s Last Theorem) and to the cryptography that secures the internet.

Galileo's lamp and Huygens's clock

Galileo's claim that a pendulum's period does not depend on its amplitude, made around 1602, was a good approximation for the small swings he studied. Christiaan Huygens realised that it was not exact, and in 1659 found the curve along which the period really is independent of amplitude: the cycloid, the subject of Section 11.4. His pendulum clocks, made from 1656 onwards, improved timekeeping from errors of about a quarter of an hour a day to about fifteen seconds. Marine chronometers, which solved the problem of finding longitude at sea, later abandoned the pendulum altogether, but pendulum clocks remained the most accurate timekeepers in the world until the 1920s.

The thinking skill

From an exact integral to a series approximation

When a quantity is given by an integral that cannot be evaluated, transform it into a well-behaved form, then expand the integrand and integrate term by term.

This is how physicists find corrections to idealised results, and it is the start of *perturbation theory*. At university you will learn when term-by-term integration is justified (uniform convergence), and how to estimate the error made by stopping.

Try this next

- 1 A tethered particle.** (From STEP III, 2015.) A particle of mass m on a smooth horizontal rail is attached to a peg at distance a from the rail by an elastic string of natural length a and modulus λ . Starting with the string just taut, it is given speed v along the rail. Show that $\dot{x}^2 = v^2 - k^2(\sqrt{x^2 + a^2} - a)^2$, where $k^2 = \frac{\lambda}{ma}$, find the greatest displacement, and show that for $v \ll ka$ the period is approximately $\sqrt{\frac{32a}{kv}} \int_0^1 \frac{du}{\sqrt{1-u^4}}$. Why does the period now *depend* on v , even for small oscillations?
- 2 The next term.** Carry the binomial expansion one term further and show that $P \approx 2\pi\sqrt{\frac{l}{g}}\left(1 + \frac{\gamma^2}{16} + \frac{11\gamma^4}{3072}\right)$.
- 3 The AGM by hand.** Use three steps of the AGM to estimate the period of a pendulum with amplitude 90° , and compare with the value 1.1803 in the figure.
- 4 Clocks.** A pendulum clock is regulated to keep perfect time with an amplitude of 4° . If its amplitude increases to 5° , how many seconds a day does it lose?

Further reading

- ▶ **David Morin**, *Introduction to Classical Mechanics* (Cambridge University Press), Chapter 4, which treats the large-amplitude pendulum.
- ▶ **Jonathan Borwein & Peter Borwein**, *Pi and the AGM* (Wiley). Gauss's discovery and its uses, including the fastest ways to compute π .
- ▶ **Dava Sobel**, *Longitude* (Fourth Estate). The story of timekeeping at sea, beyond the pendulum.

11.4 The Tautochrone

The question

A smooth, axially symmetric bowl has its vertical cross-sections determined by $s = 2\sqrt{ky}$, where s is the arc-length measured from its lowest point V , and y is the height above V . A particle is released from rest at a point on the surface at a height h above V . Explain why

$$\left(\frac{ds}{dt}\right)^2 + 2gy$$

is constant.

Show that the time for the particle to reach V is

$$\pi\sqrt{\frac{k}{2g}}.$$

Two elastic particles of mass m and αm , where $\alpha < 1$, are released simultaneously from opposite sides of the bowl at heights $\alpha^2 h$ and h respectively. If the coefficient of restitution between the particles is α , describe the subsequent motion.

The hook

Release two marbles at the same moment inside a bowl, one from near the rim and one from just above the bottom. Which reaches the bottom first? In an ordinary bowl, the one starting lower usually wins. But there is a bowl in which they always arrive together, wherever they start. It is the *tautochrone* (Greek for “same time”), and its cross-section is a *cycloid*, the curve traced by a point on the rim of a rolling wheel.

Huygens discovered this in 1659 and used it to design a pendulum clock whose period would not depend on its swing, correcting the flaw found in Section 11.3. This question proves the tautochrone property in a few lines, by choosing the right variable: distance along the curve, rather than horizontal position.

Before you start



Ask yourself The bowl is smooth. Which forces act on the particle, and which of them does work?

Gravity and the normal reaction. The normal reaction is perpendicular to the surface, and so to the motion, so it does no work. Only gravity does work, and energy is conserved.

Working it through

STAGE 1 Energy along the curve

The speed of the particle along the surface is $\left|\frac{ds}{dt}\right|$, since s is arc length. Conservation of energy gives $\frac{1}{2}m\dot{s}^2 + mgy = \text{constant}$, so $\dot{s}^2 + 2gy$ is constant. ■

STAGE 2 Simple harmonic motion in s



Ask yourself Use $s = 2\sqrt{ky}$ to replace y by a function of s . What kind of motion does the resulting equation describe?

Since $y = \frac{s^2}{4k}$, the energy equation becomes

$$\dot{s}^2 + \frac{g}{2k}s^2 = \text{constant} = \frac{g}{2k}s_0^2,$$

where $s_0 = 2\sqrt{kh}$ is the starting arc length (the particle starts at rest). Differentiating, $2\dot{s}\ddot{s} + \frac{g}{k}s\dot{s} = 0$, so

$$\ddot{s} = -\frac{g}{2k}s.$$

This is simple harmonic motion in the variable s , with $\omega^2 = \frac{g}{2k}$, whatever the amplitude: $s = s_0 \cos \omega t$.

STAGE 3 The time to the bottom

The particle reaches V (where $s = 0$) after a quarter of a period:

$$\frac{\pi}{2\omega} = \frac{\pi}{2} \sqrt{\frac{2k}{g}} = \pi \sqrt{\frac{k}{2g}},$$

which does not depend on h . ■



Pause and try Why does this work? The “restoring force” along the curve is g times the slope, $g \frac{dy}{ds} = \frac{g}{2k}s$: exactly proportional to the distance from the



bottom, measured along the curve. That is the defining property of simple harmonic motion. A particle starting higher has further to go, but is pushed proportionally harder.

STAGE 4 Two particles



Ask yourself When do the particles reach V , and how fast are they moving? Who has more momentum?

Both reach V at the same time $T = \pi\sqrt{k/(2g)}$, from opposite sides. The speed at V is ω times the amplitude in s , which is proportional to $\sqrt{\text{height}}$. Let the heavier particle, released from h , have speed U at V ; the lighter one, released from $\alpha^2 h$, has speed αU . Their momenta are $\alpha m \cdot U$ and $m \cdot \alpha U$ in opposite directions: equal and opposite. The total momentum at the collision is zero.



Ask yourself With total momentum zero and coefficient of restitution α , what are the velocities just after the collision?

The momentum stays zero, so after the collision the particles again move in opposite directions with momenta of equal size, and their relative speed is α times the approach speed $\alpha U + U$. If the lighter particle (mass αm) now moves at speed w , the heavier (mass m) moves at speed αw , and $w + \alpha w = \alpha(1 + \alpha)U$, so $w = \alpha U$. Each particle bounces back with its speed multiplied by α : the particle of mass αm leaves at αU , and the one of mass m at $\alpha^2 U$.

STAGE 5 The whole motion

Each particle now climbs back up its own side, to a height reduced by the factor α^2 (since height is proportional to speed squared): the particle of mass αm to $\alpha^2 h$ and the particle of mass m to $\alpha^4 h$. By the tautochrone property, both return to V at exactly the same moment, a time $2T$ later, with momenta again equal and opposite. The same collision happens again, with all speeds scaled by α . So:

- ▶ the particles always collide at V , at times $T, 3T, 5T, \dots$;
- ▶ after the n th collision they rise to heights $\alpha^{2n} h$ and $\alpha^{2n+2} h$;
- ▶ the motion dies away geometrically but never stops: there are infinitely many collisions, and they continue for ever. ■

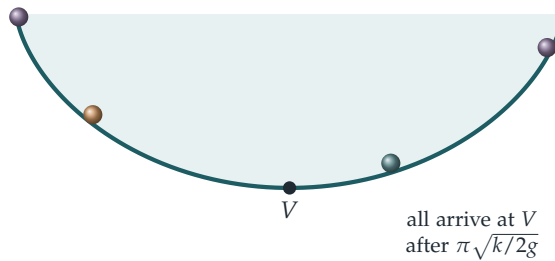


Figure 11.4.1 The tautochrone: an upside-down cycloid, traced by a point on the rim of a wheel rolling along a line. Marbles released from rest anywhere on it reach the lowest point V at the same moment. For the cycloid generated by a wheel of radius R , arc length and height satisfy $s^2 = 8Ry$, so $k = 2R$.

Looking back

- **Choose the natural variable.** In terms of arc length s , the motion is exactly simple harmonic. In terms of x or y it is a mess.
- **Energy gives the equation.** Differentiating the energy equation produced the equation of motion directly.
- **Symmetry in time.** Because every journey to the bottom takes the same time, the two particles were synchronised for ever, and the collisions repeated exactly.

What's really going on

Why the cycloid?

A wheel of radius R rolling under a line traces the cycloid $x = R(t + \sin t)$, $y = R(1 - \cos t)$. Then $\frac{dx}{dt} = R(1 + \cos t)$ and $\frac{dy}{dt} = R \sin t$, so the speed along the curve is $R\sqrt{2 + 2\cos t} = 2R \cos \frac{t}{2}$. Integrating, the arc length from the bottom is $s = 4R \sin \frac{t}{2}$, while $y = 2R \sin^2 \frac{t}{2}$. Hence $s^2 = 8Ry$: the cycloid is exactly the curve with $s = 2\sqrt{ky}$, with $k = 2R$. Conversely, the requirement that the restoring force be proportional to s forces the curve to be a cycloid.

The brachistochrone

In 1696 Johann Bernoulli challenged the mathematicians of Europe: along what curve does a bead slide from one point to a lower point (not directly below it) in the *shortest* time? The answer, the *brachistochrone*, is again a cycloid. Newton is said to have solved the problem in a single evening after receiving it; he published his solution anonymously, but Bernoulli recognised the author, “as the lion by its claw”. The problem is the founding example of the *calculus of variations*, which you met in Section 4.7: finding the curve that minimises a quantity defined by an integral.

Huygens's cycloidal clock

Huygens knew that a simple pendulum keeps imperfect time because its period depends on its amplitude. In his *Horologium Oscillatorium* (1673) he showed how to make the bob move along a cycloid: hang the string between two curved “cheeks” shaped as cycloids, and the string wraps round them in such a way that the bob traces a cycloid itself (the *involute* of a cycloid is another cycloid). In practice the cheeks caused more error through friction than they removed, and clockmakers instead kept the swings small. But Huygens’s analysis invented the theory of evolutes and involutes, and his book was one of the great works of seventeenth-century mechanics.

The thinking skill

Find the variable in which the motion is simple

Motion along a curve is often simplest in terms of the distance along the curve. Look for the variable that makes the equation linear.

At university this becomes the idea of *generalised coordinates* in Lagrangian mechanics: any variables that describe the configuration can be used, and choosing them well can make a problem solvable.

Try this next

- 1 Different heights.** A particle is released from height h in the tautochrone bowl. Show that its speed at V is $\sqrt{2gh}$, as it must be by energy conservation, and check that this agrees with ωs_0 .
- 2 An ordinary bowl.** In a hemispherical bowl of radius R , compare the time taken to reach the bottom from a small height with the time from the rim. Which is longer? (Use the result of Section 11.3.)
- 3 Equal masses.** Repeat the second part of the question with two particles of equal mass m , released from heights h_1 and h_2 , and coefficient of restitution e . Do they still always collide at V ?
- 4 Cycloid arc length.** Show that one arch of a cycloid has length $8R$, and that the area under it is $3\pi R^2$, three times the area of the rolling circle.

Further reading

- **George Simmons, *Differential Equations with Applications and Historical Notes*** (CRC Press). Includes the tautochrone and the brachistochrone, with their history.

- **E. A. Whitman**, “**Some historical notes on the cycloid**”, *American Mathematical Monthly* 50 (1943). Why the cycloid was called “the Helen of geometers”.
- **Paul Nahin**, ***When Least Is Best*** (Princeton University Press). The brachistochrone and other optimisation problems, told with history.

11.5 Gravity in an Accelerating Train

The question

A train moves westwards on a straight horizontal track with constant acceleration a , where $a > 0$. Axes are chosen as follows: the origin is fixed in the train; the x -axis is in the direction of the track with the positive x -axis pointing to the East; and the positive y -axis points vertically upwards.

A smooth wire is fixed in the train. It lies in the x - y plane and is bent in the shape given by $ky = x^2$, where k is a positive constant. A small bead is threaded onto the wire. Initially, the bead is held at the origin. It is then released.

- (i) Explain why the bead cannot remain stationary relative to the train at the origin.
- (ii) Show that, in the subsequent motion, the coordinates (x, y) of the bead satisfy

$$\dot{x}(\ddot{x} - a) + \dot{y}(\ddot{y} + g) = 0$$

and deduce that $\frac{1}{2}(\dot{x}^2 + \dot{y}^2) - ax + gy$ is constant during the motion.

- (iii) Find an expression for the maximum vertical displacement, b , of the bead from its initial position in terms of a , k and g .
- (iv) Find the value of x for which the speed of the bead relative to the train is greatest and give this maximum speed in terms of a , k and g .

The hook

When a train pulls out of a station, you feel pressed back into your seat; a cup of coffee on the table tilts its surface; a ball on the floor rolls towards the back. From inside the train it looks exactly as if gravity had tilted, pointing slightly backwards as well as down. Einstein called the realisation that this is more than an analogy “the happiest thought of my life”: gravity and acceleration are, locally, indistinguishable.

This question puts a bead on a curved wire in an accelerating train. Seen from the train, the bead behaves exactly as if it were in a gravitational field tilted by the angle whose tangent is a/g . The bottom of the wire is no longer the lowest point, and the bead oscillates about a new equilibrium.

Before you start

Ask yourself The train accelerates westwards (in the negative x direction) at a .
 ? For an observer in the train, what “fictitious” force appears to act on a mass m ?

In the train’s frame, Newton’s second law holds if we add a fictitious force $-m \times$ (acceleration of the frame). The frame accelerates at $-a$ in the x direction, so the fictitious force is $+ma$, pointing east. Together with gravity, the bead feels an effective gravity of $(a, -g)$ per unit mass: tilted towards the east.

Working it through

STAGE 1 No rest at the origin

At the origin the wire $ky = x^2$ is horizontal. The normal reaction is vertical there, so it cannot cancel the eastward fictitious force ma . The net force along the wire is $ma \neq 0$, so the bead cannot stay at rest relative to the train. ■

STAGE 2 An energy equation

Ask yourself Write Newton’s law in the train’s frame, including the normal reaction $\mathbf{N}$. Then take the dot product with the velocity $(\dot{x}, \dot{y})$. Why does $\mathbf{N}$ disappear?

In the train’s frame,

$$m(\ddot{x}, \ddot{y}) = \mathbf{N} + (ma, -mg).$$

The wire is smooth, so $\mathbf{N}$ is perpendicular to the wire, and hence to the velocity $(\dot{x}, \dot{y})$, which is along the wire. Taking the dot product with the velocity kills $\mathbf{N}$:

$$m\dot{x}\ddot{x} + m\dot{y}\ddot{y} = m\dot{x}a - mg\dot{y}, \quad \text{so} \quad \dot{x}(\ddot{x} - a) + \dot{y}(\ddot{y} + g) = 0. \quad \blacksquare$$

The left side is the derivative of $\frac{1}{2}(\dot{x}^2 + \dot{y}^2) - ax + gy$, which is therefore constant. ■

This is conservation of energy in the train’s frame, with a potential energy $-ax + gy$ per unit mass that includes the fictitious force.

STAGE 3 How high?

Ask yourself The constant is 0, since the bead starts at rest at the origin. At the highest point the bead is momentarily at rest. What equation does that give?

At rest, $\frac{1}{2}(\dot{x}^2 + \dot{y}^2) = 0$, so $ax = gy$. With $y = \frac{x^2}{k}$, this gives $ax = \frac{gx^2}{k}$, so (apart from the

starting point $x = 0$) $x = \frac{ak}{g}$, and

$$b = \frac{a^2k}{g^2}. \quad \blacksquare$$

The bead swings east and up the wire to the point $(\frac{ak}{g}, \frac{a^2k}{g^2})$, then back to the origin, and so on.

STAGE 4 The fastest point



Ask yourself Write the speed squared as a function of x alone. Where is it greatest?

The speed squared is $2(ax - gy) = 2(ax - \frac{gx^2}{k})$, a quadratic in x that is greatest at $x = \frac{ak}{2g}$, halfway between the turning points. There the speed squared is $2(\frac{a^2k}{2g} - \frac{a^2k}{4g}) = \frac{a^2k}{2g}$, so the maximum speed is

$$a\sqrt{\frac{k}{2g}}. \quad \blacksquare$$



Ask yourself Find the gradient of the wire at $x = \frac{ak}{2g}$. How does it relate to the direction of the effective gravity ($a, -g$)?

The gradient is $\frac{2x}{k} = \frac{a}{g}$. So at this point the wire is perpendicular to the effective gravity: it is the “lowest” point of the wire in the tilted field, and the bead oscillates about it (Figure 11.5.1).

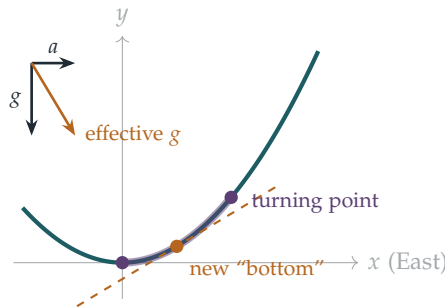


Figure 11.5.1 The wire $ky = x^2$ (teal, with $k = 1$ and $a/g = 0.6$) in the frame of the train. Effective gravity (amber arrow) is tilted east by the angle with tangent a/g . The bead oscillates (plum) between the origin and $x = ak/g$, about the point where the wire is perpendicular to the effective gravity, which acts as the new lowest point. The dashed line is a “level” line in the tilted field.

Looking back

- **Work in the convenient frame, and add the fictitious force.** In the train's frame the wire is fixed, and the only price is an extra constant force.
- **Dot with the velocity.** This removes constraint forces and gives an energy equation.
- **Interpret the answer.** The turning point and the fastest point have a geometric meaning: effective gravity is tilted.

What's really going on

Fictitious forces

Newton's laws hold in *inertial* frames, those that are not accelerating. In an accelerating frame they can be rescued by adding fictitious forces: $-m\mathbf{A}$ for a frame with acceleration $\mathbf{A}$, and, in a rotating frame, the centrifugal and Coriolis forces. They are called fictitious because no object exerts them; they are artefacts of the frame. But they are perfectly real to anyone inside it. The centrifugal force flattens the Earth at the poles, and the Coriolis force sets the direction in which hurricanes spin (anticlockwise in the northern hemisphere) and deflects long-range artillery shells.

The equivalence principle

A constant fictitious force $-m\mathbf{A}$ is proportional to mass, exactly like gravity. No experiment in a closed box can tell the difference between the box resting on a planet and the box accelerating in empty space: this is the *equivalence principle*. Einstein took it as the starting point of general relativity. Since light travelling across an accelerating box appears to bend, light must bend in a gravitational field too; and since clocks at the front and back of an accelerating rocket run at different rates, clocks at different heights in a gravitational field must too. Both predictions have been confirmed, and the second must be corrected for in every GPS satellite, whose clocks run faster than clocks on the ground by about 45 microseconds a day from this effect alone.

Einstein's happiest thought

In 1907, working at the patent office in Bern, Einstein was writing a review article on special relativity when, as he later recalled, “the happiest thought of my life” came to him: a person falling freely from a roof would not feel their own weight. Gravity could be “transformed away” by choosing a falling frame. It took him eight more years, and a great deal of new mathematics (the differential geometry of curved spaces, developed by Gauss and Riemann), to turn the thought into the general theory of relativity of 1915.

The thinking skill

Find a conserved quantity in a moving frame

In a frame with constant acceleration, a constant fictitious force acts. It can be included in a potential energy, and energy is conserved in that frame.

The trick of dotting the equation of motion with the velocity to eliminate constraint forces works in any frame. At university it becomes a general principle: a system whose equations do not depend explicitly on time has a conserved energy.

Try this next

- 1 **A cannon on a train.** (From STEP I, 2000.) A railway carriage moves north with acceleration a , and a child fires a toy cannon southwards at angle θ and speed V . Show that the range on the carriage floor is greatest when $\tan 2\theta = -\frac{g}{a}$, that this angle is approximately $\frac{\pi}{4} + \frac{a}{2g}$ when a is small, and that the maximum range is then approximately $\frac{V^2}{g} + \frac{V^2 a}{g^2}$.
- 2 **A headwind.** (From STEP II, 2006.) A projectile is fired north at speed u and angle θ , against a wind that exerts a constant southward force f per unit mass. Find where it lands. Show that if θ exceeds a certain angle α the wind blows it back towards O before it lands, and find α . Describe the motion when $f = g$ and $\theta = 45^\circ$.
- 3 **Coffee in a train.** Show that the surface of a cup of coffee in the accelerating train is a plane tilted at angle $\arctan(a/g)$ to the horizontal.
- 4 **Small oscillations.** Find the period of small oscillations of the bead about $x = \frac{ak}{2g}$. (*Hint:* near the new bottom the wire is approximately a circle of radius equal to its radius of curvature.)

Further reading

- **Daniel Kleppner & Robert Kolenkow, *An Introduction to Mechanics*** (Cambridge University Press), the chapter on non-inertial frames and fictitious forces.
- **Albert Einstein, *Relativity: The Special and the General Theory*** (various editions). Einstein's own popular account, including the accelerating box.
- **Pedro Ferreira, *The Perfect Theory*** (Little, Brown). A biography of general relativity.

11.6 Backspin

The question

By pressing a finger down on it, a uniform spherical marble of radius a is made to slide along a horizontal table top with an initial linear velocity v_0 and an initial *backward* angular velocity ω_0 about the horizontal axis perpendicular to v_0 . The frictional force between the marble and the table is constant (independent of speed).

For what value of $v_0/(a\omega_0)$ does the marble

- (i) slide to a complete stop,
- (ii) come to a stop and then roll back towards its initial position with linear speed $v_0/7$?

The hook

Press down on the back of a marble with your finger so that it squirts forwards while spinning backwards. It slides away from you, slows, stops, and then, as if it had changed its mind, rolls back. Snooker players use the same effect in the “screw” or “draw” shot, hitting the cue ball low so that it spins backwards, stops on hitting the object ball, and comes back.

The question asks for the exact amount of backspin that makes the marble stop dead, and the amount that makes it return at one-seventh of its starting speed. The calculation tracks two quantities, linear and angular velocity, both changed by the same frictional force. Behind it there is a beautiful shortcut: one combination of the two is not changed by friction at all.

A note on the syllabus

Rotational dynamics was removed from the STEP specification in 2019. All you need here is: a solid sphere of mass m and radius a has moment of inertia $I = \frac{2}{5}ma^2$ about a diameter, and a torque τ changes its angular velocity according to $I\dot{\omega} = \tau$, the rotational version of $m\dot{v} = F$.

Before you start



Ask yourself The marble moves forwards at v and spins backwards at ω . How fast is the bottom point of the marble moving over the table? Which way does friction act?

The bottom point moves at v due to the marble's motion, plus $a\omega$ due to the backward spin (a backward spin carries the bottom point forwards). So it slides forwards at $v + a\omega$, and friction acts backwards, opposing the sliding. The marble rolls without slipping only when $v + a\omega = 0$.

Working it through

STAGE 1 What friction does



Ask yourself Let the friction force be F , backwards. What does it do to v ? Its torque about the centre is Fa : does it increase or decrease the backspin?

Friction decelerates the marble: $m\dot{v} = -F$, so $v = v_0 - \frac{F}{m}t$. Acting backwards at the bottom of the marble, it also exerts a torque Fa that opposes the backspin: $I\dot{\omega} = -Fa$, so with $I = \frac{2}{5}ma^2$,

$$\omega = \omega_0 - \frac{5F}{2ma}t.$$

Both v and ω decrease steadily while the marble slides.

STAGE 2 Stopping dead



Ask yourself For the marble to stop completely, v and ω must reach zero at the same moment. When does each reach zero?

$v = 0$ at $t = \frac{mv_0}{F}$, and $\omega = 0$ at $t = \frac{2ma\omega_0}{5F}$. These are equal when $v_0 = \frac{2}{5}a\omega_0$, that is,

$$\frac{v_0}{a\omega_0} = \frac{2}{5}. \quad \blacksquare$$

At that moment the marble has no linear or angular velocity, and friction (which only acts while there is sliding) switches off: it stays put.

STAGE 3 Coming back



Ask yourself If the backspin is larger, v reaches zero while ω is still positive. The bottom point is still sliding forwards, so friction still acts backwards. What happens next, and when does sliding stop?

Friction continues to push the marble backwards, so v becomes negative, while ω keeps decreasing. Sliding stops when $v + a\omega = 0$:

$$v_0 + a\omega_0 - \left(\frac{F}{m} + \frac{5F}{2m}\right)t = 0, \quad t^* = \frac{2m(v_0 + a\omega_0)}{7F}.$$

At that moment,

$$v = v_0 - \frac{2(v_0 + a\omega_0)}{7} = \frac{5v_0 - 2a\omega_0}{7}.$$

From then on the marble rolls without slipping, friction stops acting, and v stays constant. For it to roll back at speed $\frac{v_0}{7}$ we need $5v_0 - 2a\omega_0 = -v_0$, so

$$\frac{v_0}{a\omega_0} = \frac{1}{3}. \quad \blacksquare$$

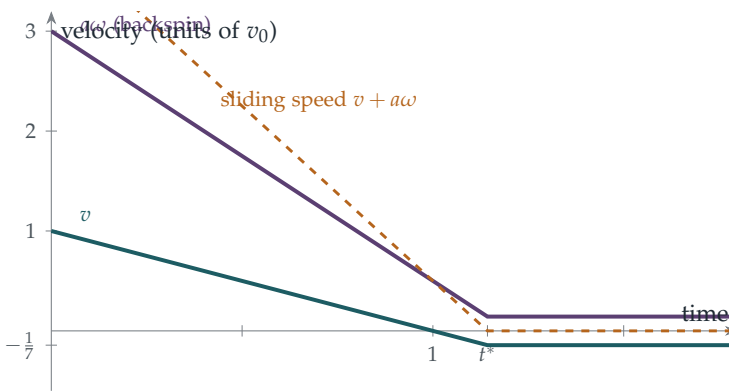


Figure 11.6.1 The case $v_0/(a\omega_0) = \frac{1}{3}$, with time in units of mv_0/F . Friction reduces the forward velocity v (teal) and the backspin $a\omega$ (plum) at constant rates, in the ratio 2 : 5. The marble stops moving forwards at $t = 1$, but is still sliding, so it is pushed backwards until the sliding speed (amber) reaches zero at $t^* = \frac{8}{7}$. It then rolls back at $\frac{v_0}{7}$.

Looking back

- **Track two quantities.** Friction changed both v and ω , at rates in a fixed ratio.
- **Know when a force switches off.** Sliding friction acts only while there is sliding; once the marble rolls, it stops.
- **The sliding speed is the key variable.** It decreased linearly, and its zero marked the change from sliding to rolling.

What's really going on

A conserved quantity

There is a quicker way to the final speed. Consider the angular momentum of the marble about the point of contact with the table. Friction acts *at* that point, so it has no moment about it, and the normal reaction and gravity both act along the vertical through it. So this angular momentum is conserved while the marble slides. It is mva (from the motion of the centre) minus $I\omega$ (the backspin counts negatively):

$$L = mav - \frac{2}{5}ma^2\omega.$$

When rolling at speed V , with $a\omega = -V$, it is $maV + \frac{2}{5}maV = \frac{7}{5}maV$. Equating with the initial value,

$$\frac{7}{5}maV = mav_0 - \frac{2}{5}ma^2\omega_0, \quad V = \frac{5v_0 - 2a\omega_0}{7},$$

the same answer with no reference to the size of the friction force, or even to whether it is constant. The final speed is determined by conservation alone; friction only decides how long it takes to get there.

Snooker, bowling and tennis

The same analysis explains many sports. A snooker ball struck centrally slides at first and then starts to roll, finishing at $\frac{5}{7}$ of its initial speed (set $\omega_0 = 0$ in the formula). A ten-pin bowling ball thrown with sidespin skids down the lane and then “hooks” when it starts to roll. A tennis ball hit with topspin dips sharply, though there the main effect is aerodynamic: a spinning ball deflects the air around it, the *Magnus effect*.

Coriolis and billiards

Gaspard-Gustave Coriolis, now remembered for the fictitious force in rotating frames (Section 11.5), wrote in 1835 a whole book on the mathematics of billiards, *Théorie mathématique des effets du jeu de billard*. He showed, among other things, that a ball struck with spin follows a parabola while it slides before rolling straight, and analysed exactly the screw shot of this question. The book is still cited by physicists who study the game.

The thinking skill

Track two quantities changed by one force

When a single force affects several quantities, write down how it changes each, and look for a combination it does not change at all.

Here friction changed v and ω but not the angular momentum about the contact point. At university, finding such combinations is a recurring theme: choosing

the point about which to take moments so that unknown forces drop out is the rotational version of choosing a frame.

Try this next

- 1 **Topspin.** If the marble starts with forward spin ω_0 instead of backspin, find its final rolling speed. For what value of $v_0/(a\omega_0)$ does it speed up?
- 2 **A snooker ball.** A ball struck centrally starts sliding at v_0 with no spin. Show that it ends up rolling at $\frac{5}{7}v_0$, and find the fraction of kinetic energy lost. (The kinetic energy of rolling is $\frac{1}{2}mV^2 + \frac{1}{2}I\omega^2$.)
- 3 **A hoop.** Repeat the question for a thin hoop rolling in its own plane, with $I = ma^2$. What ratio $v_0/(a\omega_0)$ now makes it return at $\frac{v_0}{7}$?
- 4 **The draw shot.** Where must a snooker cue strike the ball (at what height above the centre, in terms of a) for the ball to roll without sliding from the start? (*Hint:* take angular momentum about the contact point during the impulse.)

Further reading

- **David Morin, *Introduction to Classical Mechanics*** (Cambridge University Press), Chapter 8, on angular momentum and rolling with slipping.
- **Daniel Kleppner & Robert Kolenkow, *An Introduction to Mechanics*** (Cambridge University Press), the chapters on rigid body motion.
- **David Alciatore, *The Illustrated Principles of Pool and Billiards*** (Sterling). The physics of the game, with many diagrams.

11.7 Simpson's Rule in a Rod

The question

A body of mass m and centre of mass O is said to be *dynamically equivalent* to a system of particles of total mass m and centre of mass O if the moment of inertia of the system of particles is the same as the moment of inertia of the body, about any axis through O . Show that this implies that the moment of inertia of the system of particles is the same as that of the body about *any* axis.

Show that a uniform rod of length $2a$ and mass m is dynamically equivalent to a suitable system of three particles, one at each end of the rod, and one at the midpoint.

Use this result to deduce that a uniform rectangular lamina of mass M is dynamically equivalent to a system consisting of particles each of mass $\frac{1}{36}M$ at the corners, particles each of mass $\frac{1}{9}M$ at the midpoint of each side, and a particle of mass $\frac{4}{9}M$ at the centre. Hence find the moment of inertia of a square lamina, of side $2a$ and mass M , about one of its diagonals.

The mass per unit length of a thin rod of mass m is proportional to the distance from one end of the rod, and a dynamically equivalent system consists of one particle at each end of the rod and one at the midpoint. Write down a set of equations which determines these masses, and show that, in fact, only two particles are required.

The hook

How does a uniform rod respond to being pushed and spun? It turns out that, for every purpose in mechanics, a rod behaves exactly like three point masses: one-sixth of its mass at each end and two-thirds at the middle. Nothing you can do to the rod (push it, spin it about any axis) distinguishes it from the three particles.

The numbers $\frac{1}{6}, \frac{2}{3}, \frac{1}{6}$ may look familiar. They are the weights in *Simpson's rule*, $\int_{-a}^a f \approx 2a(\frac{1}{6}f(-a) + \frac{2}{3}f(0) + \frac{1}{6}f(a))$, the numerical integration method you may have met at A level. This is no coincidence: a dynamically equivalent system and a numerical integration rule are solving the same problem.

A note on the syllabus

Moments of inertia were removed from the STEP specification in 2019. The

moment of inertia of a body about an axis is $\sum m_i d_i^2$, the sum of each mass times the square of its distance from the axis (an integral for a continuous body). A uniform rod of mass m and length $2a$ has moment of inertia $\frac{1}{3}ma^2$ about a perpendicular axis through its centre. The *parallel axis theorem* says that the moment of inertia about any axis is the moment about the parallel axis through the centre of mass plus md^2 , where d is the distance between the axes.

Before you start



Ask yourself Two systems have the same mass, the same centre of mass, and the same moment of inertia about every axis through it. Why does the parallel axis theorem then give the same moment of inertia about every other axis too?

For any axis, both moments of inertia equal (moment about the parallel axis through O) $+md^2$. The first terms agree by assumption, and m and d are the same for both systems.

Working it through

STAGE 1 The rod



Ask yourself Put masses p at each end and q at the middle. By symmetry the centre of mass is right. What do the total mass and the moment of inertia about a perpendicular axis through the centre require?

We need $2p + q = m$, and the moment of inertia about a perpendicular axis through the middle must be $\frac{1}{3}ma^2$: $2pa^2 = \frac{1}{3}ma^2$. So $p = \frac{1}{6}m$ and $q = \frac{2}{3}m$.



Ask yourself Does this work for *every* axis through the centre, not just the perpendicular one?

For an axis through the centre at angle ψ to the rod, a point at distance r along the rod is at distance $r \sin \psi$ from the axis. So every moment of inertia (rod or particles) is $\sin^2 \psi$ times the value for the perpendicular axis, and the two agree for all ψ . (Axes not in a plane containing the rod work the same way: only the distance of each point from the axis matters, and for points on a line through O this is proportional to r .) So the rod is dynamically equivalent to masses $\frac{1}{6}m$, $\frac{2}{3}m$, $\frac{1}{6}m$.

STAGE 2 The rectangle

Ask yourself ? Cut the lamina into thin strips parallel to one side. Replace each strip by its three equivalent particles. What do you get, and what can you do next?

Each thin strip is a rod, equivalent to $\frac{1}{6}, \frac{2}{3}, \frac{1}{6}$ of its mass on the left edge, the centre line and the right edge. Adding over all the strips, the lamina is equivalent to three uniform rods (along the left edge, the middle and the right edge) with masses $\frac{1}{6}M, \frac{2}{3}M, \frac{1}{6}M$. Now replace each of these rods by three particles in the same ratio. The mass at each point is a product of two fractions (Figure 11.7.1):

$$\text{corners: } \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}, \quad \text{edge midpoints: } \frac{1}{6} \cdot \frac{2}{3} = \frac{1}{9}, \quad \text{centre: } \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9},$$

all times M . ■

(Strictly, each replacement must preserve the moment of inertia about every axis through the centre of the whole lamina, which it does: each replacement keeps the mass, the centre of mass and the moments of inertia about axes through the centre of mass of the piece, and the parallel axis theorem does the rest.)

STAGE 3 The square about its diagonal

For a square of side $2a$, the diagonal passes through two corners and the centre, which contribute nothing. The other two corners are at distance $\sqrt{2}a$ from it, and the four edge midpoints at distance $\frac{a}{\sqrt{2}}$. So

$$I = 2 \cdot \frac{1}{36}M \cdot 2a^2 + 4 \cdot \frac{1}{9}M \cdot \frac{a^2}{2} = \frac{1}{9}Ma^2 + \frac{2}{9}Ma^2 = \frac{1}{3}Ma^2. \quad \blacksquare$$

This is the same as the moment of inertia about an axis through the centre parallel to a side. (That is no accident: for a square, all axes through the centre in its plane give the same answer, by a symmetry argument in the exercises.)

STAGE 4 A non-uniform rod

Ask yourself ? Let the rod have length L and density $\frac{2mx}{L^2}$ at distance x from one end. Put masses p_0, p_1, p_2 at $x = 0, \frac{L}{2}, L$. Matching the total mass, the centre of mass and the moment of inertia about the end gives three equations. What are they?

The rod has total mass m , first moment $\int_0^L x \cdot \frac{2mx}{L^2} dx = \frac{2}{3}mL$ and second moment $\int_0^L x^2 \cdot \frac{2mx}{L^2} dx = \frac{1}{2}mL^2$ about the light end. Matching these:

$$p_0 + p_1 + p_2 = m, \quad \frac{1}{2}Lp_1 + Lp_2 = \frac{2}{3}mL, \quad \frac{1}{4}L^2p_1 + L^2p_2 = \frac{1}{2}mL^2.$$

Subtracting the third (divided by L^2) from the second (divided by L) gives $\frac{1}{4}p_1 = \frac{1}{6}m$, so $p_1 = \frac{2}{3}m$, then $p_2 = \frac{1}{3}m$ and $p_0 = 0$. Only two particles are needed: $\frac{2}{3}m$ at the midpoint and $\frac{1}{3}m$ at the heavy end. ■

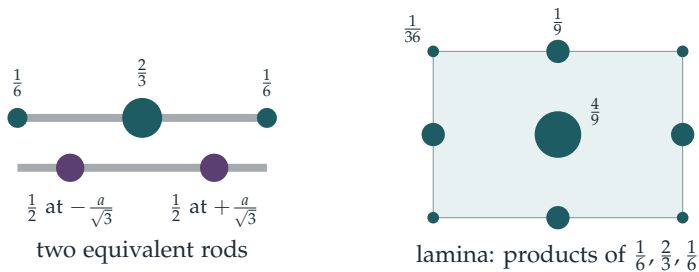


Figure 11.7.1 Left, top: a uniform rod is dynamically equivalent to masses $\frac{1}{6}$, $\frac{2}{3}$, $\frac{1}{6}$ (as fractions of its mass) at its ends and centre, the weights of Simpson's rule. Left, bottom: it is also equivalent to two halves at distance $\frac{a}{\sqrt{3}}$ from the centre, the nodes of two-point Gaussian integration. Right: a rectangular lamina is equivalent to nine masses, with areas of the dots proportional to the masses.

Looking back

- **Match moments.** Dynamical equivalence means matching the zeroth, first and second moments of the mass distribution.
- **Build up dimension by dimension.** A rectangle is a rod of rods, so its weights are products.
- **Count equations and unknowns.** Three moments to match, three unknown masses; and sometimes a solution happens to have a zero.

What's really going on

Simpson's rule

Matching moments of a mass distribution is the same problem as designing a numerical integration rule. A rule $\int_{-a}^a f(x) dx \approx w_0 f(-a) + w_1 f(0) + w_2 f(a)$ that is exact for $f = 1, x, x^2$ must satisfy

$$w_0 + w_1 + w_2 = 2a, \quad -aw_0 + aw_2 = 0, \quad a^2w_0 + a^2w_2 = \frac{2}{3}a^3,$$

exactly the conditions for the rod (with $f = 1, x, x^2$ playing the roles of mass, first moment and moment of inertia). The solution is $w = 2a(\frac{1}{6}, \frac{2}{3}, \frac{1}{6})$: Simpson's rule. A bonus comes from symmetry: the rule is also exact for $f = x^3$, since both sides vanish, so Simpson's rule integrates cubics exactly, and its error for smooth functions is proportional to the fourth derivative. A rod therefore matches its three particles not just in mass, centre and moment of inertia, but in the third moment too.

Gaussian quadrature and the inertia tensor

If the positions as well as the masses are free, two particles suffice: half the mass at each of the points $\pm \frac{a}{\sqrt{3}}$. Check: $2 \cdot \frac{1}{2}m \cdot \frac{a^2}{3} = \frac{1}{3}ma^2$. This is the two-point *Gauss–Legendre rule*, $\int_{-a}^a f \approx a \left(f\left(-\frac{a}{\sqrt{3}}\right) + f\left(\frac{a}{\sqrt{3}}\right) \right)$, which is also exact for cubics. Carl Friedrich Gauss showed in 1814 that n cleverly placed points can integrate polynomials of degree $2n - 1$ exactly, and Gaussian quadrature is now a standard method of numerical analysis.

In mechanics, the moments of inertia of a body about all axes through its centre are encoded in a symmetric 3×3 matrix, the *inertia tensor*. Two bodies with the same mass, centre of mass and inertia tensor respond identically to all forces, which is why engineers can replace a complicated part by a few point masses in a simulation. The eigenvectors of the inertia tensor are the *principal axes*, the axes about which a body can spin steadily.

The thinking skill

Replace a system by an equivalent one

If two systems agree in every quantity that matters for the problem, replace the complicated one by the simple one.

Here only three moments mattered, so a continuous rod could be replaced by three particles. At university, this idea appears as Gaussian quadrature, as multipole expansions in electromagnetism and gravitation (far away, only a few moments of a charge or mass distribution matter), and as the method of moments in statistics.

Try this next

- 1 Simpson's rule.** Use Simpson's rule with a single interval to estimate $\int_0^1 \frac{dx}{1+x^2}$, and compare with $\frac{\pi}{4}$. Check that the rule is exact for $\int_{-1}^1 x^3 dx$ and $\int_{-1}^1 x^2 dx$ but not for $\int_{-1}^1 x^4 dx$.
- 2 All axes alike.** Show that the moment of inertia of a square lamina about any axis through its centre in its plane is $\frac{1}{3}Ma^2$. (*Hint: the nine equivalent particles are symmetric under rotation by 90° ; what does that say about a quadratic expression in $\cos \psi$ and $\sin \psi$?*)
- 3 A triangle.** Show that a uniform triangular lamina of mass M is dynamically equivalent to three particles of mass $\frac{1}{3}M$ at the midpoints of its sides.
- 4 A solid cube.** Find a dynamically equivalent system of 27 particles for a uniform solid cube, and use it to find the moment of inertia of a cube of side $2a$ about an

edge.

Further reading

- ▶ **Daniel Kleppner & Robert Kolenkow, *An Introduction to Mechanics*** (Cambridge University Press), the chapter on rigid bodies and moments of inertia.
- ▶ **Endre Süli & David Mayers, *An Introduction to Numerical Analysis*** (Cambridge University Press), the chapters on numerical integration, including Simpson's rule and Gaussian quadrature.
- ▶ **Philip Davis & Philip Rabinowitz, *Methods of Numerical Integration*** (Dover). The classic reference.

11.8 Tilted Equilibria

The question

In this question, $n \geq 2$.

- (i) A solid, of uniform density, is formed by rotating through 360° about the y -axis the region bounded by the part of the curve $r^{n-1}y = r^n - x^n$ with $0 \leq x \leq r$, and the x - and y -axes. Show that the y -coordinate of the centre of mass of this solid is $\frac{nr}{2(n+1)}$.
- (ii) Show that the normal to the curve $r^{n-1}y = r^n - x^n$ at the point $(rp, r(1 - p^n))$, where $0 < p \leq 1$, meets the y -axis at $(0, Y)$, where $Y = r\left(1 - p^n - \frac{1}{np^{n-2}}\right)$.

In the case $n = 4$, show that the greatest value of Y is $\frac{1}{4}r$.

- (iii) A solid is formed by rotating through 360° about the y -axis the region bounded by the curves $r^3y = r^4 - x^4$ and $ry = -(r^2 - x^2)$, both for $0 \leq x \leq r$. A and B are the points $(0, -r)$ and $(0, r)$, respectively, on the surface of the solid.

Show that the solid can rest in equilibrium on a horizontal surface with the vector $\overrightarrow{AB}$ at three different, non-zero, angles to the upward vertical. You should not attempt to find these angles.

The hook

Put a rugby ball on a table and it settles on its side; stand it on its end and, with great care, it balances there too, but the slightest nudge topples it. Every rigid body resting on a table has a set of equilibrium positions, some stable and some unstable, and which ones exist depends only on its shape and its centre of mass. How many can a body have? How few?

This question builds a solid, shaped like a spinning top with a domed upper half and a pointed lower half, and shows that besides resting upright or upside down it can rest at three different tilts. The key idea is a single geometric condition: a body rests on a point of its surface exactly when the normal to the surface at that point passes through the centre of mass. Behind it lies a recent discovery: a convex shape, the Gömböc, with only one stable and one unstable position.

Before you start



Ask yourself A solid rests on a table, touching at a single point X of its curved surface. The forces on it are its weight (acting at the centre of mass G) and the reaction at X . When are these in equilibrium?

The table is tangent to the surface at X , and the reaction acts at X , perpendicular to the table: along the normal to the surface. The weight acts vertically through G . Two forces balance only if they act along the same line, so the vertical through X , which is the normal at X , must pass through G . So: *the body can rest touching at X exactly when the normal at X passes through G .*

Working it through

STAGE 1 The centre of mass



Ask yourself Slice the solid into thin cylindrical shells of radius x , thickness δx and height y . What is the volume of a shell, and where is its centre of mass?

With $y = \frac{r^n - x^n}{r^{n-1}}$, a shell has volume $2\pi xy \delta x$ and its centre of mass at height $\frac{y}{2}$. So

$$V = \int_0^r 2\pi xy \, dx = \frac{2\pi}{r^{n-1}} \left(\frac{r^{n+2}}{2} - \frac{r^{n+2}}{n+2} \right) = \frac{n\pi r^3}{n+2},$$

and the moment about the base is

$$\begin{aligned} \int_0^r 2\pi x \cdot \frac{y^2}{2} \, dx &= \frac{\pi}{r^{2n-2}} \int_0^r x(r^n - x^n)^2 \, dx \\ &= \pi r^4 \left(\frac{1}{2} - \frac{2}{n+2} + \frac{1}{2n+2} \right) = \frac{n^2 \pi r^4}{2(n+1)(n+2)}. \end{aligned}$$

(The bracket simplifies because $(n+1)(n+2) - 4(n+1) + (n+2) = n^2$.) Dividing,

$$\bar{y} = \frac{nr}{2(n+1)}. \quad \blacksquare$$

For $n = 2$ (a paraboloid) this is $\frac{r}{3}$, the familiar result; as $n \rightarrow \infty$ the solid becomes a cylinder and $\bar{y} \rightarrow \frac{r}{2}$.

STAGE 2 Where normals meet the axis

The gradient of the curve at $x = rp$ is $-\frac{nx^{n-1}}{r^{n-1}} = -np^{n-1}$, so the normal has gradient $\frac{1}{np^{n-1}}$. Moving along the normal from $(rp, r(1-p^n))$ to the y -axis, x decreases by rp and y by $\frac{rp}{np^{n-1}} = \frac{r}{np^{n-2}}$. So

$$Y = r \left(1 - p^n - \frac{1}{np^{n-2}} \right). \quad \blacksquare$$

Ask yourself For $n = 4$, $Y = r(1 - p^4 - \frac{1}{4p^2})$. Where is it greatest? What happens as $p \rightarrow 0$, and at $p = 1$?

$\frac{dY}{dp} = r(-4p^3 + \frac{1}{2p^3}) = 0$ when $p^6 = \frac{1}{8}$, that is $p^2 = \frac{1}{2}$. Then $Y = r(1 - \frac{1}{4} - \frac{1}{2}) = \frac{1}{4}r$. ■

As $p \rightarrow 0$, $Y \rightarrow -\infty$ (near the top the dome is very flat, and its normals are almost vertical, meeting the axis far below); at $p = 1$, $Y = -\frac{1}{4}r$. So as p runs from 0 to 1, Y rises from $-\infty$ to $\frac{1}{4}r$ and falls back to $-\frac{1}{4}r$.

STAGE 3 The spinning top

Ask yourself The solid in part (iii) is a dome ($n = 4$) on top of an upside-down paraboloid ($n = 2$). Where is its centre of mass?

From Stage 1, the dome has volume $\frac{2\pi r^3}{3}$ and centre of mass at $\frac{2r}{5}$; the paraboloid has volume $\frac{\pi r^3}{2}$ and, being upside down, centre of mass at $-\frac{r}{3}$. So

$$\bar{y} = \frac{\frac{2}{3} \cdot \frac{2}{5} - \frac{1}{2} \cdot \frac{1}{3}}{\frac{2}{3} + \frac{1}{2}} r = \frac{\frac{4}{15} - \frac{1}{6}}{\frac{7}{6}} r = \frac{3r}{35}.$$

G is at $(0, \frac{3r}{35})$, slightly above the widest point.

STAGE 4 Counting equilibria

Ask yourself A tilted equilibrium needs a point (not A or B) whose normal passes through G. On each part of the surface, how many points have normals meeting the axis at height $\frac{3r}{35}$?

The dome. Its normals meet the axis at heights Y which rise from $-\infty$ to $\frac{1}{4}r$ and fall to $-\frac{1}{4}r$. Since $-\frac{1}{4}r < \frac{3r}{35} < \frac{1}{4}r$, by the intermediate value theorem the value $\frac{3r}{35}$ is taken once on the rising part and once on the falling part: two points.

The pointed base. By part (ii) with $n = 2$, reflected in the x -axis, the normal at the point with parameter p meets the axis at $-r(1 - p^2 - \frac{1}{2}) = r(p^2 - \frac{1}{2})$, which increases from $-\frac{1}{2}r$ to $\frac{1}{2}r$. It equals $\frac{3r}{35}$ exactly once, at $p^2 = \frac{41}{70}$.

Each of these three points gives an equilibrium, and the three normals make different angles with the axis, so $\overline{AB}$ makes three different non-zero angles with the vertical (Figure 11.8.1). ■

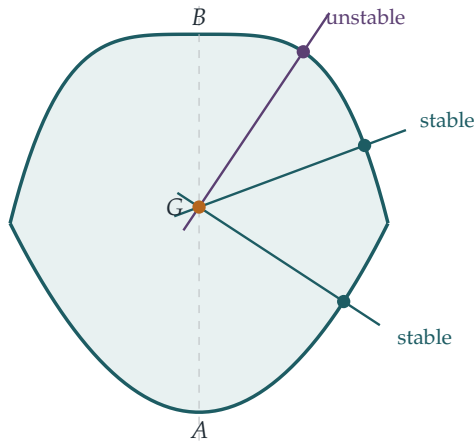


Figure 11.8.1 Cross-section of the solid of part (iii), with its centre of mass G (amber). At each of the three marked points the normal to the surface passes through G , so the solid can rest on that point. Two of these equilibria are stable (the surface is locally nearest to G there) and one is unstable (locally furthest).

Looking back

- **Turn a physical condition into geometry.** Equilibrium became “the normal passes through G ”.
- **Existence without solving.** We never found the angles. The intermediate value theorem, with the shape of the function $Y(p)$, showed how many there are.
- **Reuse a general result.** Part (ii) for general n handled both pieces of the surface in part (iii).

What’s really going on

Equilibria as critical points

When the solid rests on the point X , the height of G above the table is the distance from G to the tangent plane at X , which for a convex body is $|GX|$ when GX is along the normal. In general, the equilibria are the *critical points* of the distance from G to the surface: a point X where $|GX|$ is stationary is exactly a point whose normal passes through G . Stable equilibria are local minima of this distance (G as low as possible), and unstable ones are local maxima or saddle points.

Going round the cross-section from A to B , the distance from G is: large at A (a maximum: balanced on its point, unstable), smaller at the tilted point on the base (a minimum, stable), larger at the rim, smaller at the lower tilted point of the dome (stable), larger at the upper one (unstable), and smaller at B (stable: resting on the top of the dome). Minima and maxima must alternate, as the figure shows. (In fact there is one more equilibrium that the question does not count: the solid can balance, unstably,

on its sharp rim, where the surface has no single normal.)

The Gömböc

How few equilibria can a convex body have? Any body has at least one stable and one unstable position: the nearest and furthest points of the surface from G . A two-dimensional convex shape of uniform density always has at least four (two minima and two maxima of distance, alternating). In 1995 Vladimir Arnold asked whether a three-dimensional convex body of uniform density could have just two: one stable, one unstable. In 2006 Gábor Domokos and Péter Várkonyi constructed one, the *Gömböc*: put it down in any position, and it rolls and rocks until it reaches its single stable position. Domokos later noticed that some tortoises have shells close to this shape, which helps them right themselves when turned over.

Archimedes and floating bodies

The study of the stability of bodies goes back to Archimedes, who in *On Floating Bodies* analysed when a floating paraboloid (a model of a ship's hull) is stable upright and when it capsizes, in terms of its shape and density. His answer, worked out with geometry alone, involves exactly the kind of centre-of-mass and normal calculations used here. Naval architects still use his ideas, in the form of the *metacentre*: a ship is stable if the point where the vertical through the centre of buoyancy meets the ship's axis lies above the centre of mass.

The thinking skill

Potential energy landscapes

Equilibria are the stationary points of potential energy, and stable equilibria are its minima. Count stationary points instead of solving for them.

Here the potential energy was proportional to the height of G , the distance from G to the surface. At university this viewpoint runs through mechanics, chemistry (molecules settle into minima of energy) and topology, where *Morse theory* shows that the numbers of minima, maxima and saddles of any function on a surface are constrained by the surface's shape.

Try this next

- 1 A self-righting toy.** (From STEP II, 1996.) A toy is made of a solid cone of height λa fixed to a solid hemisphere of radius a along their common circular face. Show that if $\lambda < \sqrt{3}$ the toy returns upright when placed with its hemisphere on a table, but that if $\lambda > \sqrt{3}$ it may overbalance. (*Hint*: where must G be relative to the centre of

the hemisphere?) Show that if it is placed with the cone touching the table, it stays there whatever λ .

- 2 **A sphere.** For a uniform solid sphere, every normal passes through the centre. What does that say about its equilibria? What if the sphere's centre of mass is displaced slightly from its centre?
- 3 **An ellipse.** Show that an elliptical disc rolling in its own plane on a line has exactly four equilibria, and identify which are stable.
- 4 **Two paraboloids.** Replace the dome in part (iii) by a paraboloid $ry = r^2 - x^2$, so that the solid is symmetric about the x -axis. Show that G is at the origin, find the tilted equilibria, and explain why they come in pairs with equal tilts.

Further reading

- **Péter Várkonyi & Gábor Domokos**, “Static equilibria of rigid bodies: dice, pebbles, and the Poincaré–Hopf theorem”, *Journal of Nonlinear Science* 16 (2006). The Gömböc paper.
- **Daniel Kleppner & Robert Kolenkow**, *An Introduction to Mechanics* (Cambridge University Press), on equilibrium and stability.
- **Chris Rorres**, “Completing Book II of Archimedes's *On Floating Bodies*”, *Mathematical Intelligencer* 26 (2004). Archimedes' stability analysis, with modern pictures.

11.9 The Safety Parabola

The question

A particle is projected from a point O on horizontal ground with initial speed u and at an angle of θ above the ground. The motion takes place in the x - y plane, where the x -axis is horizontal, the y -axis is vertical and the origin is O . Obtain the Cartesian equation of the particle's trajectory in terms of u , g and λ , where $\lambda = \tan \theta$.

Now consider the trajectories for different values of θ with u fixed. Show that for a given value of x , the coordinate y can take all values up to a maximum value, Y , which you should determine as a function of x , u and g .

Sketch a graph of Y against x and indicate on your graph the set of points that can be reached by a particle projected from O with speed u .

Hence find the furthest distance from O that can be achieved by such a projectile.

The hook

A fountain sprays water at the same speed in every direction. The jets form a glittering dome with a sharp outer edge: inside, water passes through every point; outside, none reaches. That edge is a parabola, called the *parabola of safety*, because a target outside it cannot be hit by a cannon of the given muzzle speed, whatever the angle.

The surprising thing is that the boundary is not itself one of the trajectories. It is the *envelope* of the whole family, a curve that touches every trajectory once, exactly like the line $y = 0$ touching the family of parabolas in Section 8.1. This question finds it with nothing more than the discriminant of a quadratic.

Before you start



Ask yourself Write x and y in terms of t , and eliminate t . How can you write $\sec^2 \theta$ in terms of $\lambda = \tan \theta$?

With $x = ut \cos \theta$ and $y = ut \sin \theta - \frac{1}{2}gt^2$, substitute $t = \frac{x}{u \cos \theta}$:

$$y = x \tan \theta - \frac{gx^2}{2u^2 \cos^2 \theta} = \lambda x - \frac{gx^2}{2u^2} (1 + \lambda^2),$$

using $\sec^2 \theta = 1 + \tan^2 \theta$.

Working it through

STAGE 1 Fix x , vary the angle



Ask yourself For a fixed $x > 0$, think of y as a function of λ . What kind of function is it? Which values does it take as λ ranges over all real numbers?

For fixed x , y is a quadratic in λ with negative leading coefficient $-\frac{gx^2}{2u^2}$: a downward parabola in λ . It has a maximum, and as $\lambda \rightarrow \pm\infty$, $y \rightarrow -\infty$. Since it is continuous, it takes every value up to its maximum. The maximum is where $\frac{dy}{d\lambda} = x - \frac{gx^2}{u^2}\lambda = 0$, that is $\lambda = \frac{u^2}{gx}$, and there

$$Y = \frac{u^2}{gx}x - \frac{gx^2}{2u^2}\left(1 + \frac{u^4}{g^2x^2}\right) = \frac{u^2}{2g} - \frac{gx^2}{2u^2}. \quad \blacksquare$$

(Equivalently: the point (x, y) can be reached exactly when the quadratic $\frac{gx^2}{2u^2}\lambda^2 - x\lambda + \frac{gx^2}{2u^2} + y = 0$ has a real root λ , and the discriminant condition gives $y \leq Y$.)

STAGE 2 The region

The curve $Y = \frac{u^2}{2g} - \frac{gx^2}{2u^2}$ is a downward parabola with its vertex at $(0, \frac{u^2}{2g})$, the height reached by firing straight up, crossing the ground at $x = \pm\frac{u^2}{g}$, the maximum range. Every point on or below it (and above the ground) can be reached, and no point above it can (Figure 11.9.1).



Ask yourself Through a point strictly below the curve, how many trajectories pass? And through a point on it?

A point below the curve corresponds to a quadratic in λ with two real roots: two angles of projection, one high and one low. A point on the curve corresponds to a repeated root: exactly one trajectory, which touches the curve there.

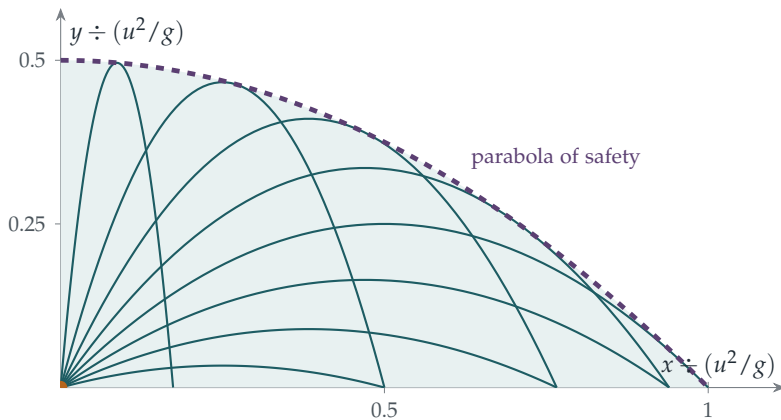


Figure 11.9.1 Trajectories at a fixed speed u and angles $15^\circ, 25^\circ, \dots, 85^\circ$ (teal). Each touches the dashed parabola of safety $y = \frac{u^2}{2g} - \frac{gx^2}{2u^2}$ exactly once, and the region beneath it (shaded) is exactly the set of points that can be reached. The origin O (amber) is the focus of the parabola of safety.

STAGE 3 The furthest point

Ask yourself The furthest reachable point from O (on or above the ground) lies on the boundary. Write the squared distance $x^2 + Y^2$ as a function of x^2 . Is it increasing or decreasing?

Let $X = x^2$. Then

$$x^2 + Y^2 = X + \left(\frac{u^2}{2g} - \frac{gX}{2u^2} \right)^2, \quad \frac{d}{dX}(x^2 + Y^2) = 1 - \frac{g}{u^2} \left(\frac{u^2}{2g} - \frac{gX}{2u^2} \right) = \frac{1}{2} + \frac{g^2 X}{2u^4} > 0.$$

So the distance increases as we move out along the boundary, and the furthest point is where the boundary meets the ground, at $x = \frac{u^2}{g}$. The furthest distance is

$$\frac{u^2}{g},$$

achieved by the ordinary maximum-range shot at 45° . ■

Looking back

- **Turn the question around.** Instead of asking where a given trajectory goes, ask which trajectories pass through a given point.
- **A quadratic in the parameter.** For each point, the equation for $\lambda = \tan \theta$ is a quadratic, and its discriminant decides whether the point can be reached.
- **Two, one or none.** Two angles reach a point inside, one reaches a point on the boundary, none reaches a point outside.

What's really going on

Envelopes

A one-parameter family of curves $F(x, y, \lambda) = 0$ often has an *envelope*: a curve touching every member of the family. It is found by solving $F = 0$ and $\frac{\partial F}{\partial \lambda} = 0$ together, which is exactly what we did by maximising y over λ . Envelopes appear throughout science. The bright curve of light on the surface of a cup of coffee is the envelope of reflected rays, a *caustic*. A supersonic aircraft's sonic boom is carried by the envelope of the sound waves it emits, the *Mach cone*. And in Section 8.1, the envelope of a family of solutions of a differential equation was itself a solution, a “singular” one.

The focus

The parabola of safety has a remarkable property: its focus is the launch point O . (A parabola $y = c - \frac{x^2}{4c}$ has its focus at the origin; here $c = \frac{u^2}{2g}$.) Every trajectory, too, is a parabola, and the foci of all the trajectories lie on a circle of radius $\frac{u^2}{2g}$ about O . There is an energy explanation: the speed of the projectile at height y is $\sqrt{u^2 - 2gy}$, which is zero at the height $\frac{u^2}{2g}$ of the directrix, and parabolas whose directrix is the same line and which pass through O have their foci on that circle.

Torricelli's parabola

Galileo showed around 1609 that projectiles move along parabolas, and that the maximum range on level ground is at 45° . His student Evangelista Torricelli, better known for inventing the barometer, found the parabola of safety in 1644, in his *De motu gravium*. For gunners it answered a practical question: which targets are within range at all? Torricelli also showed that the trajectories through a given point come in pairs with angles symmetric about the direction of the point and the vertical, the high and low shots of the question.

The thinking skill

Envelopes of families of curves

To find the boundary of the region covered by a family of curves, fix a point and ask for which parameter values a curve passes through it. The boundary is where the number of solutions changes.

The same idea, counting solutions as a point moves, is the basis of the study of caustics in optics, of bifurcations in dynamical systems, and of catastrophe theory, which classifies how the number of solutions of an equation can change.

Try this next

- 1 **An exploding firework.** (From STEP II, 2023.) An explosion at the origin sends particles in all directions with speed u . Show that at time t they lie on a sphere of radius ut whose centre falls freely from the origin. Show that a particle projected horizontally with speed u from the top of the paraboloid of safety moves along its surface for ever, and find when it must be launched to collide with the particle fired at angle α .
- 2 **Firing up a slope.** Find the maximum range of a projectile with speed u up a plane inclined at angle β . (*Hint:* where does the line $y = x \tan \beta$ meet the parabola of safety?) Show that the best angle bisects the angle between the slope and the vertical.
- 3 **Two angles.** Show that the two angles that reach a point P are symmetric about the bisector of the angle between OP and the vertical.
- 4 **The foci.** Show that the trajectory with angle θ has its focus at $(\frac{u^2}{2g} \sin 2\theta, -\frac{u^2}{2g} \cos 2\theta)$, and hence that the foci lie on a circle about O .

Further reading

- **David Morin, *Introduction to Classical Mechanics*** (Cambridge University Press), Chapter 3, which discusses the envelope of trajectories.
- **J. W. Bruce, P. J. Giblin & C. G. Gibson, “On caustics of plane curves”, *American Mathematical Monthly* 88 (1981).** Envelopes and caustics in geometry.
- **E. H. Lockwood, *A Book of Curves*** (Cambridge University Press). Envelopes, caustics and many classical curves, with constructions.

11.10 A Ball That Comes Back Towards You

The question

A particle P is projected from a point O on horizontal ground with speed u and angle of projection α , where $0 < \alpha < \frac{1}{2}\pi$.

- (i) Show that if $\sin \alpha < \frac{2\sqrt{2}}{3}$, then the distance OP is increasing throughout the flight.

Show also that if $\sin \alpha > \frac{2\sqrt{2}}{3}$, then OP will be decreasing at some time before the particle lands.

- (ii) At the same time as P is projected, a particle Q is projected horizontally from O with speed v along the ground in the opposite direction from the trajectory of P . The ground is smooth. Show that if

$$2\sqrt{2}v > (\sin \alpha - 2\sqrt{2}\cos \alpha)u,$$

then QP is increasing throughout the flight of P .

The hook

Throw a ball forwards and it moves away from you until it lands. Obviously. Throw it straight up and it comes back. Somewhere in between, something must change: there must be angles at which the ball moves away, then comes *closer*, then moves away again before landing. This question finds the exact threshold, $\sin \alpha = \frac{2\sqrt{2}}{3}$, or about 70.5° .

The result is a nice example of a common situation in mathematics: a quantity everyone assumes is monotonic, which is not. It also shows a standard technique, used constantly at university: to find out whether a distance is increasing, differentiate its *square*, which avoids square roots.

Before you start



Ask yourself Why is it easier to work with OP^2 than with OP ? Is “ OP is increasing” the same as “ OP^2 is increasing”?

$OP^2 = x^2 + y^2$ is a polynomial in t , while OP involves a square root. Since $OP \geq 0$, and squaring is increasing on non-negative numbers, OP increases exactly when OP^2 does.

Working it through

STAGE 1 The squared distance

With $x = ut \cos \alpha$ and $y = ut \sin \alpha - \frac{1}{2}gt^2$,

$$OP^2 = u^2 t^2 \cos^2 \alpha + u^2 t^2 \sin^2 \alpha - ugt^3 \sin \alpha + \frac{1}{4}g^2 t^4 = u^2 t^2 - ugt^3 \sin \alpha + \frac{1}{4}g^2 t^4.$$



Ask yourself Differentiate with respect to t . After taking out a factor of t , what is left? When is it positive for all t ?

$$\frac{d}{dt}OP^2 = 2u^2 t - 3ugt^2 \sin \alpha + g^2 t^3 = t(g^2 t^2 - 3ug \sin \alpha t + 2u^2).$$

For $t > 0$, the sign is that of the quadratic $g^2 t^2 - 3ug \sin \alpha t + 2u^2$. This quadratic has positive leading coefficient, so it is positive for every t if its discriminant is negative:

$$9u^2 g^2 \sin^2 \alpha - 8u^2 g^2 < 0 \iff \sin^2 \alpha < \frac{8}{9} \iff \sin \alpha < \frac{2\sqrt{2}}{3}.$$

In that case OP is increasing throughout the flight. ■

STAGE 2 Steeper throws



Ask yourself If $\sin \alpha > \frac{2\sqrt{2}}{3}$, the quadratic has two real roots. Are they positive? Is the smaller one before the particle lands?

The roots have sum $\frac{3u \sin \alpha}{g} > 0$ and product $\frac{2u^2}{g^2} > 0$, so both are positive, and between them the quadratic, and so $\frac{d}{dt}OP^2$, is negative. The particle lands at $T = \frac{2u \sin \alpha}{g}$. The smaller root is

$$t_1 = \frac{u}{2g} (3 \sin \alpha - \sqrt{9 \sin^2 \alpha - 8}) < \frac{u}{2g} \cdot 3 \sin \alpha < \frac{2u \sin \alpha}{g} = T.$$

So just after t_1 (and before landing) OP is decreasing. ■

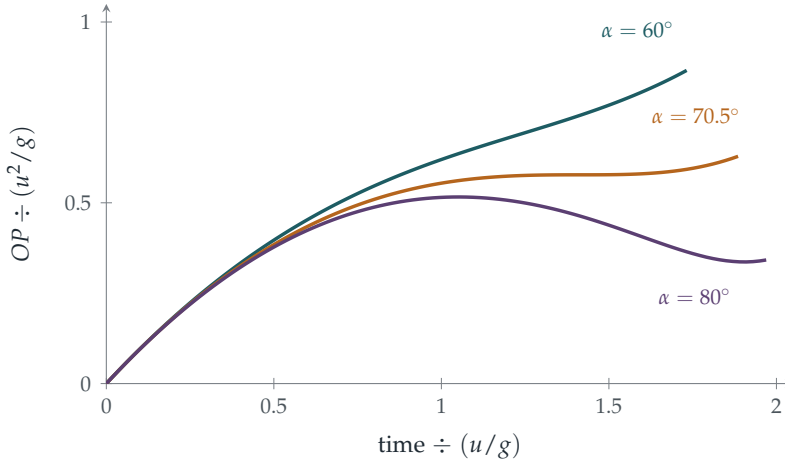


Figure 11.10.1 Distance from the thrower against time, until landing, for three angles. At 60° (teal) it always increases. At the critical angle $\arcsin \frac{2\sqrt{2}}{3} \approx 70.5^\circ$ (amber) it pauses for an instant. At 80° (plum) the ball moves away, comes back towards the thrower for a while, and moves away again just before landing.

STAGE 3 A moving observer



Ask yourself Q moves away from O along the ground in the opposite direction, at speed v . Measured from Q , the horizontal separation is $x + vt$. What does that do to the calculation?

The separation QP has horizontal component $(u \cos \alpha + v)t$ and vertical component $ut \sin \alpha - \frac{1}{2}gt^2$. So

$$QP^2 = ((u \cos \alpha + v)^2 + u^2 \sin^2 \alpha)t^2 - ugt^3 \sin \alpha + \frac{1}{4}g^2t^4,$$

the same as before with u^2 replaced by $w^2 = (u \cos \alpha + v)^2 + u^2 \sin^2 \alpha$ in the first term. The derivative is $t(g^2t^2 - 3ug \sin \alpha t + 2w^2)$, and the condition for it to be positive for all $t > 0$ is

$$9u^2 \sin^2 \alpha < 8w^2 \iff u^2 \sin^2 \alpha < 8(u \cos \alpha + v)^2,$$

after cancelling $8u^2 \sin^2 \alpha$ from both sides. Taking square roots (both $u \sin \alpha$ and $u \cos \alpha + v$ are positive), this is $u \sin \alpha < 2\sqrt{2}(u \cos \alpha + v)$, which rearranges to

$$2\sqrt{2}v > (\sin \alpha - 2\sqrt{2}\cos \alpha)u. \quad \blacksquare$$

(This is sufficient for QP to increase throughout the flight, since the derivative is then positive for all $t > 0$.)

Looking back

- **Differentiate the square.** Squared distances are polynomials, and their derivatives have the same sign as the derivative of the distance.

- **A quadratic that is always positive.** The condition “increasing for all time” became a discriminant condition.
- **A new problem as an old one.** The moving observer only changed one coefficient, so the earlier analysis applied directly.

What’s really going on

A geometric view

OP decreases exactly when the velocity of P points “inwards”, making an obtuse angle with the vector $\vec{OP}$, that is, when $\vec{OP} \cdot \mathbf{v} < 0$. For a steep throw, near the top of the flight the particle is high above and slightly in front of the thrower, moving almost horizontally away, but starting to fall; once the downward part of its velocity outweighs the outward part, relative to the direction of $\vec{OP}$, it comes closer. For shallow throws the particle is always far enough in front that its velocity never points back towards O . The critical case is when the velocity is perpendicular to $\vec{OP}$ at exactly one instant, the moment when the distance pauses in Figure 11.10.1.

Monotonicity is not automatic

Many quantities that “obviously” increase do not. The distance of the Moon from the Earth oscillates; the distance between two cars on a winding road can decrease even when one is always ahead; the population of a predator can fall while its prey multiplies. Proving that a quantity is monotonic usually means showing that a derivative has a fixed sign, and the technique here (reduce to a polynomial, then check its roots or discriminant) is the standard way to do it. In analysis, monotonicity is often the key to proving convergence: a bounded monotonic sequence converges, as you saw in Chapter 2.

The thinking skill

Check monotonicity through a derivative

To decide whether a quantity always increases, differentiate it (or a convenient function of it, such as its square) and ask whether the derivative can change sign.

At university, sign arguments of this kind are used to prove inequalities, to locate maxima and minima, and to study the stability of solutions of differential equations, where a quantity that always decreases (as in Section 8.2) proves that a system settles down.

Try this next

- 1 **How close?** For $\alpha = 80^\circ$, find the times at which OP has its local maximum and minimum, and the distances there, in units of u^2/g .
- 2 **On a slope.** Repeat part (i) for a ball thrown up a plane inclined at angle β to the horizontal, landing on the plane. Does the critical angle (measured from the plane) change?
- 3 **With air resistance.** Suppose the ball also experiences a resistance proportional to its velocity. Without calculating, would you expect the critical angle to be larger or smaller than 70.5° ? Why?
- 4 **Throwing from a height.** If the ball is thrown from the top of a cliff and lands far below, show that OP eventually increases for any angle, and find whether there are angles for which it decreases at some stage.

Further reading

- **James Walker, *The Flying Circus of Physics*** (Wiley). Hundreds of surprising everyday physics puzzles.
- **David Morin, *Introduction to Classical Mechanics*** (Cambridge University Press), Chapter 3, on projectiles, with several counter-intuitive problems.
- **J. M. Steele, *The Cauchy–Schwarz Master Class*** (Cambridge University Press). The art of proving that one thing is always bigger than another.

11.11 How Fielders Catch

The question

A fielder, who is perfectly placed to catch a ball struck by the batsman in a game of cricket, watches the ball in flight. Assuming that the ball is struck at the fielder's eye level and is caught just in front of her eye, show that $\frac{d}{dt}(\tan \theta)$ is constant, where θ is the angle between the horizontal and the fielder's line of sight.

In order to catch the next ball, which is also struck towards her but at a different velocity, the fielder runs at constant speed v towards the batsman. Assuming that the ground is horizontal, show that the fielder should choose v so that $\frac{d}{dt}(\tan \theta)$ remains constant.

The hook

A ball is struck high into the outfield. A good fielder sets off at once and arrives just as it comes down. How? Nobody can solve the equations of projectile motion in their head in a fraction of a second, and in any case the fielder does not know the ball's speed or angle. Yet fielders, and dogs catching frisbees, do it routinely.

In 1968 the physicist Seville Chapman proposed an answer, and this question proves it. If the fielder watches the angle of elevation θ of the ball, and moves so that $\tan \theta$ increases at a *constant rate*, she will arrive exactly where the ball lands. She needs no knowledge of physics, only the ability to notice whether the ball seems to be rising faster or slower than steadily. A hard prediction problem is replaced by a simple feedback rule.

Before you start



Ask yourself Put the batsman at the origin and the fielder at distance D . If the ball is at (x, y) , what is $\tan \theta$?

The ball is at height y above the fielder's eye, and at horizontal distance $D - x$ in front of her (it is coming towards her). So $\tan \theta = \frac{y}{D-x}$.

Working it through

STAGE 1 The perfectly placed fielder



Ask yourself Let the ball have horizontal speed U and initial vertical speed W , and let it land (at eye level) at time T . Write x , y and D in terms of U , W , T , g and t . Can you factorise y ?

We have $x = Ut$ and $y = Wt - \frac{1}{2}gt^2$, and the ball returns to eye level at $T = \frac{2W}{g}$, where the fielder is: $D = UT$. Since $W = \frac{1}{2}gT$,

$$y = \frac{1}{2}gt(T - t), \quad D - x = U(T - t).$$

The factor $T - t$ cancels:

$$\tan \theta = \frac{\frac{1}{2}gt(T - t)}{U(T - t)} = \frac{g}{2U} t.$$

So $\frac{d}{dt} \tan \theta = \frac{g}{2U}$, a constant. ■

STAGE 2 The running fielder



Ask yourself Now the fielder starts at distance D_0 and runs towards the batsman at speed v , so her position is $D_0 - vt$. She catches the ball at time T . What is D_0 ? Repeat the calculation.

To catch the ball at time T , she must be where it lands: $D_0 - vT = UT$, so $D_0 = (U + v)T$. Her horizontal distance from the ball is $D_0 - vt - Ut = (U + v)(T - t)$, and again the factor cancels:

$$\tan \theta = \frac{\frac{1}{2}gt(T - t)}{(U + v)(T - t)} = \frac{g}{2(U + v)} t.$$

So if she is running at the right speed, $\tan \theta$ increases at a constant rate.



Ask yourself The question asks for the converse: if she runs so that $\frac{d}{dt} \tan \theta$ is constant, will she catch the ball?

Suppose she runs at speed v and $\tan \theta = \frac{\frac{1}{2}gt(T - t)}{D_0 - (U + v)t}$ increases at a constant rate c , so $\tan \theta = ct$ (it is 0 at $t = 0$). Then

$$\frac{1}{2}g(T - t) = c(D_0 - (U + v)t) \quad \text{for all } t.$$

Comparing the coefficients of t and the constant terms: $\frac{1}{2}g = c(U + v)$ and $\frac{1}{2}gT = cD_0$. Dividing, $D_0 = (U + v)T$: exactly the condition to be at the landing point at time T . So she should choose v to keep $\tan \theta$ rising steadily, and then she catches the ball. ■

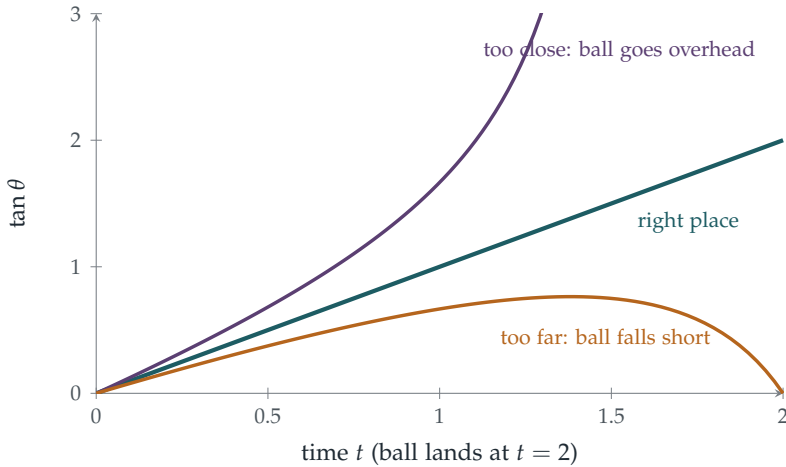


Figure 11.11.1 What a stationary fielder sees, for a ball that lands at distance 2 (with $g = 2$ and horizontal speed 1). In the right place, $\tan \theta$ rises in a straight line (teal). Standing too close, $\tan \theta$ accelerates upwards and the ball goes over her head (plum); too far back, it rises ever more slowly and then falls (amber). The rule: move so that the image of the ball rises steadily.

Looking back

- **Look for a common factor.** Both the height and the horizontal distance contained $T - t$, and cancelling it made $\tan \theta$ linear.
- **Prove the converse.** Showing that the rule *works* (it leads to the catch) needed the reverse argument, comparing coefficients.
- **An invariant that is easy to perceive.** A steady rise in $\tan \theta$ can be judged by eye; the landing point cannot.

What's really going on

Optical acceleration cancellation

Chapman's rule is now called *optical acceleration cancellation*: the fielder moves so as to keep the "optical acceleration" $\frac{d^2}{dt^2} \tan \theta$ at zero. If the ball's image accelerates upwards, the ball will go over her head, so she backs away; if it decelerates, the ball will fall short, so she runs in (Figure 11.11.1). The strategy does not tell her where the ball will land, but it gets her there. It is a feedback control law, like the ones that steer guided missiles: missiles use *proportional navigation*, keeping the line-of-sight angle to the target constant, which is the sideways version of the same idea.

What fielders really do

Do real fielders use Chapman's rule? Experiments in the 1990s, with cameras mounted on the heads of baseball outfielders, gave mixed results. Michael McBeath and colleagues argued in *Science* in 1995 that, for balls hit to one side, fielders run along curved paths that keep the ball's image moving in a straight line across their field of view (a "linear optical trajectory"), which also leads them to the catch. Later work has suggested that both strategies, and a mixture of them, are used, and that fielders often do not run at constant speed. A 2004 study found that dogs catching frisbees follow paths consistent with the same optical rules. Whatever the details, the remarkable conclusion stands: catching does not require prediction, only the right feedback.

Chapman's paper

Seville Chapman's "Catching a baseball", published in the *American Journal of Physics* in 1968, is three pages long and contains essentially the calculation in this question. He noted that the strategy works only without air resistance, and that a real baseball's trajectory is distorted by drag and spin, which would complicate things. Nevertheless, the idea that the brain uses simple invariants of the visual field to solve complicated control problems has become important in psychology and robotics.

The thinking skill

Find an invariant that is easy to observe

When a quantity is hard to predict, look for a related quantity that can be observed directly and that behaves simply exactly when things are on track.

Here the landing point was hard to predict, but the rate of change of $\tan \theta$ was easy to monitor. At university this is the idea of feedback control, and of choosing an observable that encodes the information you need.

Try this next

- 1 Running backwards.** If the ball will land behind the fielder's starting position, show that she should run away from the batsman, and that the same rule applies.
- 2 Not θ itself.** Show that, for a perfectly placed fielder, $\frac{d\theta}{dt}$ is not constant. Why is $\tan \theta$ the right quantity to watch?
- 3 Eye level.** If the ball is struck from a height h below the fielder's eye level, is $\tan \theta$ still linear in t for a well-placed fielder? (Assume she catches it at eye level.)

- 4 **Air resistance.** With linear air resistance, the horizontal velocity decays exponentially. Is $\tan \theta$ still linear in t for a well-placed fielder?

Further reading

- Seville Chapman, “Catching a baseball”, *American Journal of Physics* 36 (1968).
- Michael McBeath, Dennis Shaffer & Mary Kaiser, “How baseball outfielders determine where to run to catch fly balls”, *Science* 268 (1995).
- Dennis Shaffer et al., “How dogs navigate to catch Frisbees”, *Psychological Science* 15 (2004).

11.12 The Ant on the Rubber Rope

The question

A uniform elastic string lies on a smooth horizontal table. One end of the string is attached to a fixed peg, and the other end is pulled at constant speed u . At time $t = 0$, the string is taut and its length is a . Obtain an expression for the speed, at time t , of the point on the string which is a distance x from the peg at time t .

An ant walks along the string starting at $t = 0$ at the peg. The ant walks at constant speed v along the string (so that its speed relative to the peg is the sum of v and the speed of the point on the string beneath the ant). At time t , the ant is a distance x from the peg. Write down a first order differential equation for x , and verify that

$$\frac{d}{dt} \left(\frac{x}{a + ut} \right) = \frac{v}{a + ut}.$$

Show that the time T taken for the ant to reach the end of the string is given by

$$uT = a(e^k - 1),$$

where $k = u/v$.

On reaching the end of the string, the ant turns round and walks back to the peg. Find in terms of T and k the time taken for the journey back.

The hook

An ant sets off along a rubber rope one kilometre long, walking at one centimetre per second. Every second, the rope is stretched uniformly by another kilometre. The far end races away at a thousand metres per second, a hundred thousand times faster than the ant. Does the ant ever reach the end?

Astonishingly, yes. The stretching carries the ant along with it, and although the ant's progress as a *fraction* of the rope shrinks, the fractions add up like the harmonic series, which diverges (Section 2.3). The ant arrives, though only after a time that dwarfs the age of the universe. This question solves the continuous version exactly, and finds that the time grows exponentially with the ratio of the speeds.

Before you start



Ask yourself The string is stretched uniformly: every piece is stretched by the same factor. At time t its length is $a + ut$. If the point at distance x from the peg is a fraction $\frac{x}{a+ut}$ of the way along, how fast does it move?

Each point on the string keeps the same fraction f of the length, so its distance from the peg is $f(a + ut)$, and its speed is $fu = \frac{ux}{a+ut}$. Points near the peg barely move; the far end moves at u .

Working it through

STAGE 1 The ant's equation

The ant moves at v relative to the string, plus the speed of the string beneath it:

$$\frac{dx}{dt} = v + \frac{ux}{a + ut}.$$



Ask yourself Differentiate $\frac{x}{a+ut}$ by the quotient rule and use the equation for $\dot{x}$.

$$\frac{d}{dt}\left(\frac{x}{a + ut}\right) = \frac{\dot{x}}{a + ut} - \frac{ux}{(a + ut)^2} = \frac{1}{a + ut}\left(v + \frac{ux}{a + ut}\right) - \frac{ux}{(a + ut)^2} = \frac{v}{a + ut}. \quad \blacksquare$$

The fraction of the string the ant has covered increases at the rate $\frac{v}{a+ut}$: its own speed divided by the current length. The stretching drops out completely, because it does not change anyone's fraction.

STAGE 2 Reaching the end



Ask yourself Integrate from 0 to t . When does the fraction reach 1?

The ant starts at the peg, so

$$\frac{x}{a + ut} = \int_0^t \frac{v \, ds}{a + us} = \frac{v}{u} \ln \frac{a + ut}{a}.$$

The ant reaches the end when the fraction is 1: $\ln \frac{a+uT}{a} = \frac{u}{v} = k$, so $a + uT = ae^k$, that is

$$uT = a(e^k - 1). \quad \blacksquare$$

The logarithm grows without limit, so the ant always arrives, however fast the rope is stretched.

STAGE 3 The way back



Ask yourself Walking back, the ant's speed relative to the string is $-v$. What is the rate of change of its fraction now? It starts at fraction 1 at time T .

The same calculation with $-v$ gives $\frac{d}{dt} \left(\frac{x}{a+ut} \right) = -\frac{v}{a+ut}$, so

$$\frac{x}{a+ut} = 1 - \frac{v}{u} \ln \frac{a+ut}{a+uT}.$$

The fraction reaches 0 when $a+ut = (a+uT)e^k$. Since $a+uT = ae^k$, the return time is

$$t - T = \frac{(a+uT)(e^k - 1)}{u} = \frac{ae^k \cdot uT/a}{u} = Te^k. \quad \blacksquare$$

The journey back takes e^k times as long as the journey out: by the time the ant turns round, the string is e^k times longer than at the start.

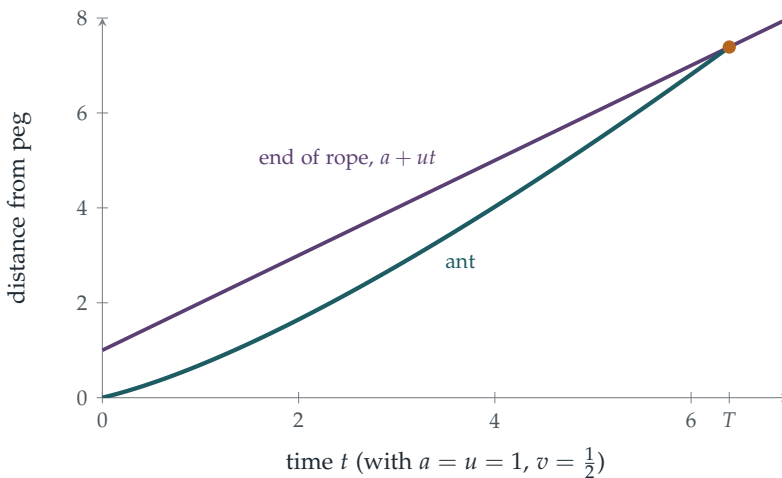


Figure 11.12.1 The ant (teal) walking along a rope whose end (plum) moves at twice the ant's walking speed ($k = 2$). The ant is carried along by the stretching, and its distance from the peg grows faster and faster. It reaches the end at $T = e^2 - 1 \approx 6.39$.

Looking back

- **Use the right variable.** The fraction of the rope covered was the natural quantity: the stretching does not change it.
- **Integrating factor in disguise.** Multiplying by $\frac{1}{a+ut}$ turned a linear differential equation into an exact derivative.
- **Divergence beats speed.** The fraction grew like a logarithm, which is unbounded, so the ant always arrives.

What's really going on

The harmonic series version

The puzzle is usually told with discrete stretches. A rope is 1 km long; the ant walks 1 cm in the first second; then the rope is stretched by 1 km; then the ant walks another centimetre, and so on. In the n th second, the ant covers 1 cm of a rope n km long: a fraction $\frac{1}{100\,000n}$. Stretching does not change its fraction, so after N seconds it has covered

$$\frac{1}{100\,000} \left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{N} \right) \approx \frac{\ln N + 0.577}{100\,000}$$

of the rope. It arrives when $H_N \approx 100\,000$, after about $e^{100\,000}$ seconds, a number with over 43 000 digits. (For comparison, the universe is about 4×10^{17} seconds old.) Our continuous answer agrees: $k = \frac{u}{v} = 100\,000$, and $T \approx \frac{a}{u} e^k$.

Light in an expanding universe

The ant on the rope is a model of light crossing an expanding universe. Distant galaxies recede from us, carried by the expansion of space itself, and light from them travels towards us at speed c *through* the expanding space, as the ant walks along the stretching rope. Whether light from a distant galaxy ever reaches us depends on how the expansion rate changes with time. In a universe expanding at a steady rate, like the rope, light always arrives eventually. In a universe whose expansion accelerates, as ours appears to, there are galaxies whose light will never reach us: there is a *cosmic event horizon*. The analogue for the ant is a rope stretched exponentially, so that the end's speed grows in proportion to the length (try it in the exercises).

Gardner's ant

The puzzle seems to have been invented by the French mathematician Denys Wilquin in 1972, as a worm on an elastic band, and was popularised by Martin Gardner in his *Scientific American* column and in his book *Time Travel and Other Mathematical Bewilderments*. It is one of the best-known examples of a result that contradicts intuition and is nevertheless easy to prove, and it is a favourite illustration of the divergence of the harmonic series.

The thinking skill

Find the quantity the complication does not affect

When part of a problem changes everything (here, the stretching), look for a quantity it leaves alone.

Stretching changes every distance but no fraction. Working with the fraction removed the stretching from the problem. At university, this idea appears as *comoving coordinates* in cosmology and fluid mechanics, and more generally as the search for invariants under a transformation.

Try this next

- 1 **Halfway.** When is the ant halfway along the rope? Is that before or after time $\frac{T}{2}$?
- 2 **Exponential stretching.** Suppose instead the rope's length is $ae^{\lambda t}$. Show that the ant reaches the end only if $v > \lambda a$. (This is the cosmic event horizon.)
- 3 **An accelerating rope.** If the length is $a + \frac{1}{2}bt^2$, does the ant always arrive?
- 4 **The discrete puzzle.** In the discrete version, how many seconds does the ant need to cover 1% of the rope? Use $H_N \approx \ln N + 0.577$.

Further reading

- **Martin Gardner**, *Time Travel and Other Mathematical Bewilderments* (W.H. Freeman). Includes the ant on the rubber rope.
- **Julian Havil**, *Gamma: Exploring Euler's Constant* (Princeton University Press). The harmonic series, and the ant.
- **Barbara Ryden**, *Introduction to Cosmology* (Cambridge University Press). Expanding space, horizons and comoving coordinates, at undergraduate level.

11.13 The Snowplough Problem

The question

Two identical snowploughs plough the same stretch of road in the same direction. The first starts at time $t = 0$ when the depth of snow is h metres, and the second starts from the same point T seconds later. Snow falls so that the depth of snow increases at a constant rate of $k \text{ m s}^{-1}$. It may be assumed that each snowplough moves at a speed equal to $k/(\alpha z) \text{ m s}^{-1}$ with α a constant, where z is the depth of snow it is ploughing, and that it clears all the snow.

- (i) Show that the time taken for the first snowplough to travel x metres is

$$(e^{\alpha x} - 1) \frac{h}{k} \text{ seconds.}$$

- (ii) Show that at time $t > T$, the second snowplough has moved y metres, where t satisfies

$$\frac{1}{\alpha} \frac{dt}{dy} = t - (e^{\alpha y} - 1) \frac{h}{k}.$$

Verify that the required solution of this equation is

$$t = (e^{\alpha y} - 1) \frac{h}{k} + \left(T - \frac{\alpha h y}{k}\right) e^{\alpha y}$$

and deduce that the snowploughs collide when they have moved a distance $kT/(\alpha h)$ metres.

The hook

Snow is falling steadily. A snowplough sets off down a road, slowing down as the snow it meets gets deeper. A second plough follows a little later along the cleared road. It meets only the snow that has fallen since the first plough passed, which is shallower, so it travels faster, and eventually it catches up. When?

This is a version of a famous problem published by Ralph Agnew in 1942, which became one of the best-known puzzles in differential equations. Its charm is that the second plough's speed depends on *when the first plough passed each point*, so the solution to part (i) feeds into the equation for part (ii). And the equation is cleverest when written "upside down", with time as a function of distance.

Before you start



Ask yourself The first plough meets snow of depth $h + kt$ at time t . What is its speed? Is it easier to find x as a function of t , or t as a function of x ?

Its speed is $\frac{dx}{dt} = \frac{k}{\alpha(h+kt)}$. This separates neatly. Turning it over, $\frac{dt}{dx} = \frac{\alpha(h+kt)}{k}$, which is a simple linear equation for t as a function of x . Either way works for the first plough; for the second, the upside-down form is essential.

Working it through

STAGE 1 The first plough

From $\frac{dt}{dx} = \alpha(\frac{h}{k} + t)$, the quantity $\frac{h}{k} + t$ grows exponentially in x : $\frac{h}{k} + t = Ce^{\alpha x}$. At $x = 0, t = 0$, so $C = \frac{h}{k}$ and

$$t_1(x) = \frac{h}{k}(e^{\alpha x} - 1). \quad \blacksquare$$

The first plough slows down exponentially: equal distances take exponentially increasing times.

STAGE 2 The second plough's equation



Ask yourself The second plough reaches position y at time t . How deep is the snow there? (What was the depth just after the first plough passed?)

The first plough cleared all the snow at position y at time $t_1(y)$. Since then, snow has accumulated at rate k , so the depth at time t is $z = k(t - t_1(y))$. The second plough's speed is therefore

$$\frac{dy}{dt} = \frac{k}{\alpha k(t - t_1(y))} = \frac{1}{\alpha(t - t_1(y))},$$

and turning it over,

$$\frac{1}{\alpha} \frac{dt}{dy} = t - \frac{h}{k}(e^{\alpha y} - 1). \quad \blacksquare$$

STAGE 3 Solving



Ask yourself This is linear in t : $\frac{dt}{dy} - \alpha t = -\frac{\alpha h}{k}(e^{\alpha y} - 1)$. What integrating factor works? What is the starting condition?

Multiplying by $e^{-\alpha y}$,

$$\frac{d}{dy}(te^{-\alpha y}) = -\frac{\alpha h}{k}(1 - e^{-\alpha y}), \quad te^{-\alpha y} = -\frac{\alpha h}{k}y - \frac{h}{k}e^{-\alpha y} + C.$$

The second plough starts at $y = 0$ at $t = T$, so $T = -\frac{h}{k} + C$, giving $C = T + \frac{h}{k}$. Multiplying back by $e^{\alpha y}$ and rearranging,

$$t_2(y) = \frac{h}{k}(e^{\alpha y} - 1) + \left(T - \frac{\alpha hy}{k}\right)e^{\alpha y}. \quad \blacksquare$$

(To verify, as the question asks: differentiate and substitute; the terms combine to $\alpha(t - t_1(y))$.)

STAGE 4 The collision

Ask yourself The ploughs collide when both reach the same point at the same time. Compare $t_2(y)$ with $t_1(y)$.

The difference is $t_2(y) - t_1(y) = (T - \frac{\alpha hy}{k})e^{\alpha y}$, the time gap between the two ploughs at position y . It shrinks to zero when

$$y = \frac{kT}{\alpha h}. \quad \blacksquare$$

At that point the second plough reaches the first: the snow in front of it has zero depth, and in the model its speed becomes infinite (Figure 11.13.1).

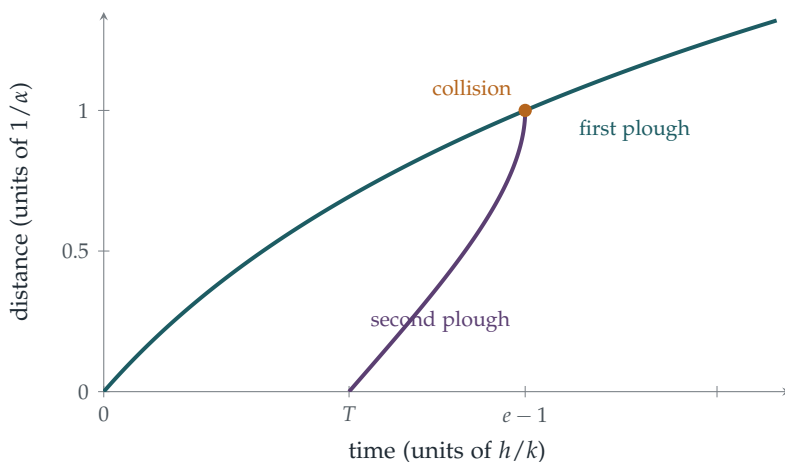


Figure 11.13.1 Distance against time for the two ploughs, with $T = \frac{h}{k}$ (so the second starts once the snow depth has doubled). The first plough (teal) slows as the snow deepens. The second (plum) starts later on a partly cleared road and speeds up as it closes the gap, meeting the first at distance $\frac{kT}{\alpha h} = \frac{1}{\alpha}$, at time $(e - 1)\frac{h}{k}$.

Looking back

- **Turn the equation upside down.** With t as a function of position, both equations were linear.

- **One solution feeds the next.** The first plough's timetable $t_1(y)$ appeared in the second plough's equation.
- **Compare, do not solve twice.** The collision condition came from the difference $t_2 - t_1$, which had a simple factor.

What's really going on

Agnew's snowplough

In Ralph Palmer Agnew's version, from his 1942 textbook *Differential Equations*, it starts snowing heavily and steadily one morning; a snowplough sets out at noon and covers 2 miles in the first hour and 1 mile in the second; when did it start snowing? The plough's speed is inversely proportional to the depth, which is proportional to the time since the snow started. The answer (an exercise below) involves the golden ratio: it started snowing about 37 minutes before noon. Versions with two ploughs, like this question, followed.

Information that travels

The second plough's motion depends on the history recorded in the road: the time at which each point was last cleared. This is a simple example of a system with *memory*, in which the state at one place depends on what happened there earlier. Traffic flow, the motion of glaciers and the spreading of a forest fire all have this character, and their mathematics, the theory of first-order partial differential equations and their *characteristics*, is a standard second-year topic. The idea of writing time as a function of position ("when does the front arrive here?") is exactly the method of *arrival times* used for such problems.

The thinking skill

Choose the independent variable

A differential equation may be much simpler with the roles of the variables reversed. If the rate depends on the time at which something happened at a place, write time as a function of place.

Both equations here became linear when written for $t(y)$. At university, the freedom to choose which variable is independent (and to change it when convenient) is used constantly, from the hodograph method in fluid mechanics to the use of arc length as a parameter in geometry.

Try this next

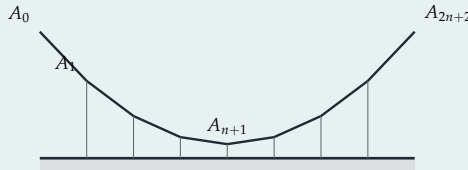
- 1 **Agnew's problem.** Snow starts falling at a steady rate at an unknown time before noon. A snowplough whose speed is inversely proportional to the depth starts at noon and goes 2 miles in the first hour and 1 mile in the second. Show that if it started snowing t_0 hours before noon, then $(1 + \frac{2}{t_0})^2 = (1 + \frac{1}{t_0})^3$, deduce that $\frac{1}{t_0}$ is the golden ratio, and find when it started snowing.
- 2 **Verification.** Verify by differentiation that $t_2(y)$ satisfies the differential equation of part (ii).
- 3 **The time gap.** Show that the time gap $t_2(y) - t_1(y)$ first increases and then decreases to zero, and find where it is greatest. Interpret this.
- 4 **Three ploughs.** A third plough starts at time $2T$. Write down the equation for its motion. Does it catch the second plough before or after the second catches the first?

Further reading

- **Ralph Palmer Agnew, *Differential Equations*** (McGraw-Hill, 1942). The original snowplough problem.
- **George Simmons, *Differential Equations with Applications and Historical Notes*** (CRC Press), which discusses the snowplough problem among many others.
- **Richard Haberman, *Mathematical Models*** (SIAM). Traffic flow and characteristics, at an accessible level.

11.14 Parabola, Not Catenary

The question



The above diagram represents a suspension bridge. A heavy uniform horizontal roadway is attached by vertical struts to a light flexible chain at points $A_1 = (x_1, y_1)$, $A_2 = (x_2, y_2), \dots, A_{2n+1} = (x_{2n+1}, y_{2n+1})$, where the coordinates are referred to horizontal and vertically upward axes Ox, Oy . The chain is fixed to external supports at points

$$A_0 = (x_0, y_0) \quad \text{and} \quad A_{2n+2} = (x_{2n+2}, y_{2n+2})$$

at the same height. The weight of the chain and struts may be neglected. Each strut carries the same weight w . The horizontal spacing h between A_i and A_{i+1} (for $0 \leq i \leq 2n+1$) is constant. Write down equations satisfied by the tensions T_i in the portion $A_{i-1}A_i$ of the chain for $1 \leq i \leq n+1$. Hence or otherwise show that

$$\frac{h}{y_n - y_{n+1}} = \frac{3h}{y_{n-1} - y_n} = \dots = \frac{(2n+1)h}{y_0 - y_1}.$$

Verify that the points $A_0, A_1, \dots, A_{2n+1}, A_{2n+2}$ lie on a parabola.

The hook

Hold a chain by its two ends and it hangs in a curve called the *catenary* (from the Latin for chain). Galileo thought it was a parabola; it is not, as the teenage Huygens showed in 1646. But hang a heavy roadway from the chain by evenly spaced struts, as in a suspension bridge, and the chain *does* form a parabola. The difference is in how the weight is spread: along the curved chain itself for the catenary, evenly along the horizontal for the bridge.

This question proves the parabola with nothing more than resolving forces at each joint. The key is that the horizontal component of tension is the same all the way along, so each segment's slope is proportional to the weight it supports. The result is a beautiful pattern: the drops between successive joints go up in the odd numbers $1, 3, 5, 7, \dots$.

whose partial sums are the perfect squares.

Before you start



Ask yourself At each joint A_i , three forces act: the tensions in the two neighbouring segments, and the weight w hanging from the strut. Which component of tension can the weight not affect?

The weight is vertical, so at each joint the horizontal components of the two tensions must balance. So the horizontal component of tension, H say, is the same in every segment of the chain.

Working it through

STAGE 1 The equations at each joint

Let segment $A_{i-1}A_i$ make angle ϕ_i below the horizontal (going from A_{i-1} towards the middle), so its tension T_i has horizontal component $T_i \cos \phi_i$ and vertical component $T_i \sin \phi_i$. Resolving at A_i for $1 \leq i \leq n$:

$$T_i \cos \phi_i = T_{i+1} \cos \phi_{i+1} = H, \quad T_i \sin \phi_i - T_{i+1} \sin \phi_{i+1} = w.$$

(The segment on the outer side pulls up and out with the larger vertical component; the inner segment pulls in and slightly down; the difference supports the strut.)



Ask yourself What happens at the middle joint A_{n+1} ? Use the symmetry of the bridge.

By symmetry, the two segments meeting at A_{n+1} are mirror images, so each supports half of its strut: $T_{n+1} \sin \phi_{n+1} = \frac{1}{2}w$.

STAGE 2 Slopes from weights

Working outwards from the middle, the vertical components are

$$T_{n+1} \sin \phi_{n+1} = \frac{1}{2}w, \quad T_n \sin \phi_n = \frac{3}{2}w, \quad \dots, \quad T_1 \sin \phi_1 = \frac{2n+1}{2}w.$$

Dividing each by H , the slopes are $\tan \phi_i = \frac{(2(n+1-i)+1)w}{2H}$. The slope of segment $A_{i-1}A_i$ is also $\frac{y_{i-1}-y_i}{h}$, so

$$\frac{y_n - y_{n+1}}{h} = \frac{w}{2H}, \quad \frac{y_{n-1} - y_n}{h} = \frac{3w}{2H}, \quad \dots, \quad \frac{y_0 - y_1}{h} = \frac{(2n+1)w}{2H}.$$

Inverting and multiplying by the odd numbers, each of $\frac{h}{y_n - y_{n+1}}, \frac{3h}{y_{n-1} - y_n}, \dots, \frac{(2n+1)h}{y_0 - y_1}$ equals $\frac{2H}{w}$. ■

STAGE 3 The parabola

Ask yourself Let j count joints outwards from the middle, so that A_{n+1-j} is at horizontal distance jh from A_{n+1} . How far is it above A_{n+1} ?

Adding the drops, with $c = \frac{wh}{2H}$:

$$y_{n+1-j} - y_{n+1} = c(1 + 3 + 5 + \cdots + (2j - 1)) = cj^2.$$

With x measured from the middle, $x = jh$, so the height above the lowest joint is $\frac{c}{h^2}x^2 = \frac{w}{2Hh}x^2$. By symmetry the same holds on the other side, and the supports ($j = n + 1$) fit the same formula. So all the points lie on the parabola

$$y = y_{n+1} + \frac{w}{2Hh}(x - x_{n+1})^2. \quad \blacksquare$$

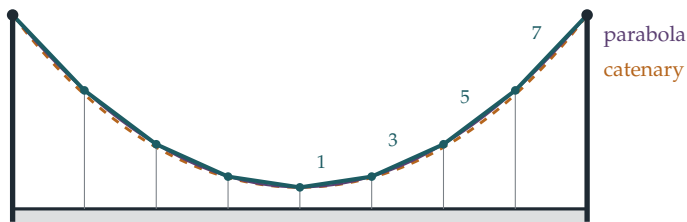


Figure 11.14.1 A chain carrying equal loads at equal horizontal spacings (teal). Its joints lie on a parabola (plum), and the drops between successive joints, from the middle outwards, are in the ratio 1 : 3 : 5 : 7. A free chain under its own weight, with the same ends and lowest point, would hang in a catenary (amber, dashed), which is flatter at the bottom and steeper near the supports.

Looking back

- **Find what is constant.** With only vertical loads, the horizontal tension is the same everywhere.
- **Accumulate from the middle.** Each segment supports all the weight between it and the middle, so the vertical components grow in steps of w .
- **Odd numbers sum to squares.** $1 + 3 + \cdots + (2j - 1) = j^2$ turned a list of slopes into a parabola.

What's really going on

Parabola or catenary?

The continuous version of the argument explains both curves. For a chain $y(x)$ under constant horizontal tension H , the vertical component of tension is Hy' , and over a short stretch it changes by the load carried there: $H dy' = dW$.

- If the load is spread evenly *horizontally* (a heavy roadway, weight ρ per unit length of road), then $Hy'' = \rho$, so y is a quadratic: a parabola.
- If the load is the chain's own weight, spread evenly *along the chain* (ρ per unit length of chain), then $Hy'' = \rho\sqrt{1 + y'^2}$, whose solution is the catenary $y = c \cosh \frac{x}{c}$, with $c = H/\rho$.

Near the bottom, where the chain is nearly horizontal, the two are almost the same; they differ near the supports, where the chain is steep and a given horizontal distance contains more chain. Real suspension bridges, whose deck is much heavier than their cables, are very close to parabolas.

Upside-down chains and arches

A chain is in pure tension; turn its shape upside down and you have an arch in pure compression, with no tendency to bend. Robert Hooke published this in 1675 as an anagram, which decoded reads: "As hangs the flexible line, so but inverted will stand the rigid arch." Christopher Wren and Hooke used the idea for the dome of St Paul's Cathedral, and Antoni Gaudí designed the Sagrada Família in Barcelona by hanging weighted strings from the ceiling and photographing them upside down. The Gateway Arch in St Louis is a (weighted) inverted catenary.

The catenary challenge

In 1690 Jakob Bernoulli challenged mathematicians to find the shape of a hanging chain. Within a year, solutions were published by his brother Johann, by Leibniz and by Huygens. Johann Bernoulli later recalled with glee that his brother, who had set the problem, had failed to solve it. The catenary was one of the first triumphs of the new calculus, and it led to the hyperbolic functions $\cosh$ and $\sinh$, studied systematically by Vincenzo Riccati and Johann Heinrich Lambert in the eighteenth century.

The thinking skill

Resolve forces at every joint

For a structure made of joints and straight members, balance forces at each joint. Constant components and accumulating loads often reveal the shape.

At university this becomes the analysis of trusses and frameworks in engineering, and, in the continuous limit, the differential equations of cables, beams and membranes. The step from a discrete chain to a continuous one is the same move from sums to integrals seen in Section 10.3.

Try this next

- 1 An even number of struts.** Repeat the question with $2n$ struts, so that the lowest point of the chain is a horizontal segment in the middle. Do the joints still lie on a parabola?
- 2 The tension.** Find the greatest tension in the chain, in terms of H , w and n . Where is it?
- 3 The catenary.** Verify that $y = c \cosh \frac{x}{c}$ satisfies $y'' = \frac{1}{c} \sqrt{1 + y'^2}$. Show that the length of the catenary from $x = 0$ to $x = X$ is $c \sinh \frac{X}{c}$.
- 4 Galileo's error.** A chain of length $2L$ hangs between two supports at distance $2d$ apart. For L only slightly larger than d , show that the sag of the catenary is approximately the same as that of a parabola. Why did Galileo's experiments not reveal the difference?

Further reading

- **E. H. Lockwood, *A Book of Curves*** (Cambridge University Press). The catenary, the parabola and their relatives.
- **Jacques Heyman, *The Stone Skeleton*** (Cambridge University Press). How masonry arches and domes stand up, from Hooke's chain onwards.
- **David Billington, *The Tower and the Bridge*** (Princeton University Press). The art and engineering of bridges.

What We Have Learned

We began with things that swing, things that spin, and things seen from somewhere that moves. We asked three questions: when is motion simple, what is the right point of view, and can intuition be trusted? Here is what the chapter's fourteen questions showed.

1. When is motion simple?

Simple harmonic motion, with a single frequency independent of amplitude, is the ideal. Driving it at its natural frequency produced resonance (Section 11.1), with beats just off resonance and unlimited growth exactly on it. Two coupled oscillators turned out to be two independent simple oscillators in disguise (Section 11.2), but with frequencies in the irrational ratio $\sqrt{3}$, so the combined motion never repeats. The real pendulum is not simple at all (Section 11.3): its period grows with amplitude, through an elliptic integral that Gauss's arithmetic-geometric mean computes with astonishing speed. And the tautochrone (Section 11.4) showed how to make motion exactly simple again, by bending the track into a cycloid.

2. What is the right point of view?

Almost every question was solved by choosing the right variable or the right frame.

- **Coordinates.** Arc length made the tautochrone simple; the combinations $x \pm y$ decoupled the springs; the fraction of the rope removed the stretching from the ant's problem (Section 11.12); writing time as a function of distance made the snowploughs' equations linear (Section 11.13).
- **Frames.** In the accelerating train (Section 11.5), a fictitious force tilted gravity, and the bead oscillated about a new lowest point. Einstein turned this observation into general relativity.
- **Points to take moments about.** For the spinning marble (Section 11.6), angular momentum about the contact point was conserved, because friction acts there, giving the final speed without knowing the friction.
- **Equivalent systems.** A rod behaves exactly like three particles (Section 11.7), with the weights of Simpson's rule; and the equilibria of a solid are where its normals pass through the centre of mass (Section 11.8).

3. Can intuition be trusted?

Often not. The boundary of the region a cannon can reach is not a trajectory but their envelope (Section 11.9). A ball thrown steeply enough comes back towards you before it lands (Section 11.10). A fielder can catch a ball without predicting where it will land, by keeping its image rising steadily (Section 11.11). An ant walking slowly along a rope stretched a hundred thousand times faster still reaches the end (Section 11.12). A hanging chain is not a parabola, but a chain carrying a bridge is (Section 11.14). In each case, a short calculation replaced a confident but wrong intuition with a precise statement, and usually with an explanation.

4. A toolkit for oscillations and frames

Method	Where you used it
Particular plus complementary solution	Resonance (11.1)
Decoupling by eigenvectors	Normal modes (11.2)
Energy integral and substitution	Pendulum (11.3)
The natural variable	Tautochrone (11.4); ant (11.12)
Fictitious forces	Train (11.5)
Angular momentum about a point	Backspin (11.6)
Matching moments	Simpson's rod (11.7)
Counting stationary points	Tilted equilibria (11.8)
Discriminants and envelopes	Safety parabola (11.9)
Sign of a derivative	Coming back (11.10)
Comparing coefficients	Fielders (11.11)
Integrating factors	Ant (11.12); snowplough (11.13)
Resolving at joints	Bridge (11.14)

Where next

- The next two chapters turn to probability and statistics, starting with Section 12.1.
- Normal modes are eigenvectors (Section 7.2), and the quantum harmonic oscillator of Section 8.6 is the oscillator of this chapter in quantum mechanics.
- Limit cycles (Section 8.4) are oscillations that sustain themselves, and chaos (Section 8.7) is what nonlinear oscillators can do when driven.

If you want to go further before university

- **A. P. French, *Vibrations and Waves*:** the classic introduction to oscillations, resonance and normal modes.
- **David Morin, *Introduction to Classical Mechanics*:** rotation, frames and many surprising problems.

- ▶ **Steven Strogatz, *Sync*:** how oscillators fall into step, from fireflies to bridges.
- ▶ **George Simmons, *Differential Equations with Applications and Historical Notes*:** the tautochrone, the brachistochrone and the snowplough, with their history.

Coleman joked that a physicist's career consists of treating the harmonic oscillator in ever more abstract ways. This chapter has tried to show why: the oscillator is the simplest motion there is, and almost everything looks like one if you choose the right point of view. Learning to find that point of view, a variable, a frame or a combination, is the most useful skill that mechanics teaches.