Waves III: Specific Waves

Sound waves are covered in chapter 19 of Halliday and Resnick, while light waves are covered in chapters 39 and 44. For more about light waves, see chapter 9 of Purcell. For water waves, and many other neat wave phenomena, see chapter I-51 of the Feynman lectures and chapter 7 of Crawford's *Waves*. For more about polarization, see chapter I-33 of the Feynman lectures, or for more detail, chapter 8 of Hecht's *Optics*. Basic geometric optics is covered in chapter 40 of Halliday and Resnick, and chapter 1 of Wang and Ricardo, volume 2. There is a total of **74** points.

1 Sound and Longitudinal Waves

[3] **Problem 1.** In this problem, you'll work through Newton's slick derivation of the speed of sound. Instead of considerig how parcels of gas move, we consider the force the gas exerts when squeezed.

(a) In **M4**, we showed that the speed v of longitudinal waves in a spring of length L, mass m, and spring constant obeys $v^2 = kL^2/m$. For a cylinder of gas of length L and area A, show that the effective spring constant is

$$k = -A^2 \frac{dp}{dV}.$$

(b) Assuming the sound waves are adiabatic, use this to conclude that

$$v^2 = \frac{\gamma p}{\rho}.$$

If each gas molecule has mass m, rewrite the result in terms of γ, T, and m.

Next, we consider some limitations of this result.

(c) In an ideal gas, we assume the particles are noninteracting: they pass right through each other. But for sound waves to propagate, adjacent packets of ideal gas must exert pressure on each other. How is this possible? Use this observation to estimate the maximum possible frequency of sound in a gas in terms of the number density $n = N/V = p/k_B T$, the radius r of a gas molecule, and the speed of sound v.

(d) Our analysis also breaks down if the pressure variations are no longer adiabatic. The rate of heat conduction in a gas with thermal conductivity k_t across a surface of area A is

$$\frac{dQ}{dt} = -Ak_t \frac{dT}{dx}$$

For a sound wave with angular frequency ω, show that the adiabatic approximation holds when $\omega \ll pk_B/mk_t$. Does this hold for audible sound in air, where $k_t \approx 25\,\mathrm{mW/(m \cdot K)}$?

Solution. (a) By definition, we have $k = -dF/dx$, where F is the force experienced by a piston at the end of the cylinder as it moves a distance x. But we also have $F = Ap$ and $dV = A\,dx$, and combining these gives the result.

(b) Plugging the result of part (a) in, we have

$$v^2 = -A^2 \frac{dp}{dV} \frac{L^2}{m} = -V \frac{dp}{dV} \frac{V}{m}.$$

For adiabatic sound waves, $V\,dp/dV = -\rho\,dp/d\rho = -\gamma p$, so that $v^2 = \gamma p/\rho$.

Alternatively, using the ideal gas law $p = \rho k_B T/m$, we can rewrite this as $v^2 = \gamma k_B T/m$.

(c) There have to be enough gas molecules so that each individual gas molecule undergoes many collisions per wave period, or else the wave still simply fall apart. We saw in part (b) that the typical thermal speed of the air molecules is comparable to the speed of sound, so the mean time between collisions is $\sim 1/(nr^2 v)$. Then we must have

$$f \ll nr^2 v.$$

One easy trick to evaluate this is to note that air molecules are separated by roughly 10 times their radius, $n \sim 1/(10\,r)^3$, and $r \sim 10^{-10}\,\text{m}$. Then we have $f \ll 10^{-3} v/r \sim \text{GHz}$.

(d) A sound wave consists of regions of higher and lower temperature. For the adiabatic approximation to be valid, the heat transfer rate has to be negligible compared to the rate at which a parcel's energy changes due to the propagation of the sound wave.

Let the sound wave have temperature amplitude ΔT, wavelength λ, and period $\tau \sim \lambda/v$. Then a half-wavelength of warm air loses heat to its surroundings at rate

$$\frac{dQ}{dt} \sim \frac{Ak_t\,\Delta T}{\lambda} \sim Ak_t\,\Delta T\,\frac{\omega}{v}.$$

On the other hand, its internal energy varies as the sound wave passes by, at the typical rate

$$\frac{dE}{dt} \sim nA\lambda k_B\,\frac{\Delta T}{\tau} \sim nAk_B\,\Delta T\,v.$$

Comparing these expressions, we must have

$$\omega \ll \frac{nk_B v^2}{k_t} \sim \frac{nk_B^2 T/m}{k_t} \sim \frac{pk_B}{mk_t} \sim \text{GHz}.$$

So sound waves in air are adiabatic, in all the frequencies they can even exist, and certainly at audible frequencies.

Note that you might have guessed, based on the intuition that adiabatic processes are "fast", that sound is adiabatic for *high* frequencies. Instead, it's the opposite. The reason is that heat transfer is enhanced at higher frequencies because λ gets shorter, so dT/dx gets higher.

Remark

Phase shifts upon reflection for sound waves can be a bit tricky. Recall from **W1** that a hard boundary for a transverse string wave $y(x,t)$ sets y to zero. As a result, upon reflection, y flips sign, but $v_y = \partial y/\partial t$ stays the same.

When a sound wave hits a hard wall, the wall sets the displacement $\xi(x,t)$ to zero. Then upon reflection, the displacement flips sign, while the pressure variation $\delta P(x,t) \propto \partial\xi/\partial x$

stays the same. In standing waves, a hard wall is thus a node for ξ and an antinode for δP. Similarly, when sound waves in a tube reflect off an open end, the end sets δP to zero (since everything outside the tube has atmospheric pressure), so it flips sign. An open end is thus a node for δP and an antinode for ξ.

The rule is always the same: whatever quantity gets fixed to zero by the boundary gets flipped in sign upon reflection, and for a standing wave, that quantity has a node at the boundary. But it's confusing enough that several common high school textbooks get it wrong. Some even state, in their confusion, that "hard boundaries flip transverse waves but not longitudinal ones", which is definitely not true in general.

[2] **Problem 2** (HRK)**.** Some conceptual questions about sound waves.

(a) What is larger for a sound wave, the relative density variations $\Delta\rho/\rho$ or the relative pressure variations $\Delta P/P$? Or does it depend on the situation?

(b) What is larger, the velocity of a sound wave v or the amplitude of the velocity variations Δu of the underlying particles? Or does it depend on the situation?

Solution. (a) For an adiabatic sound wave, $P \propto \rho^\gamma$, which implies $\Delta P/P = \gamma\,\Delta\rho/\rho$. Then $\Delta P/P$ is larger.

(b) For a sound wave of wavelength λ, during a time $t \sim \lambda/v$ the particles will move a distance $\Delta x \sim \Delta u \lambda/v$. Then the maximum relative compression will be $\Delta x/\lambda \sim \Delta u/v$. So as long as the density variations are small, which would be true for a typical sound wave, we have $\Delta u \ll v$. (If we had $\Delta u \sim v$, then we would instead have a strong shock wave, which can't be described by the results above.)

[2] **Problem 3.** A rubber rope with unstretched length L_0 is stretched to length $L > L_0$.

(a) Find the ratio of the speeds of transverse and longitudinal waves.

(b) Experimentally, it is found that the longitudinal waves are much more strongly damped. (You can check this at home, by making such a rope by tying together cut rubber bands.) Can you explain why, by considering the molecular structure of rubber?

Solution. (a) Using the result of problem 1, the longitudinal wave speed is

$$v_l = \sqrt{\frac{kL^2}{m}}$$

where k and m are the spring constant and mass. The transverse wave speed is

$$v_t = \sqrt{\frac{T}{m/L}} = \sqrt{\frac{k(L-L_0)L}{m}}$$

from which we conclude that

$$\frac{v_t}{v_l} = \sqrt{\frac{L-L_0}{L}}.$$

Note that they become approximately equal in the limit where the rope is highly stretched.

(b) As was discussed in **T2**, the molecules of rubber are long chains, which are curled up in the rubber's unstretched state, and get straightened as it stretches. Longitudinal waves thus involve crumpling and straightening the chains, which dissipates a lot of energy, while transverse waves only involve the chains bending from side to side.

Idea 1: Doppler Effect

Working in one dimension with speed of sound c, if a source of sound at frequency f_0 travels at velocity v_s while an observer to their right travels at velocity v_o, the observed frequency is

$$f = \frac{c - v_o}{c - v_s} f_0.$$

Example 1

A speaker is between two perfectly reflective walls and emits a sound of frequency f_0. If you carry the speaker and walk with small speed v towards one of the walls, what do you hear?

Solution

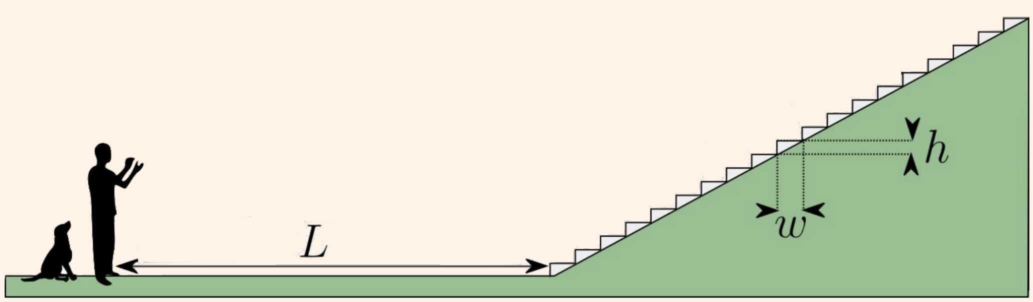
Let's work to lowest order in v/c everywhere. The wall you're walking toward experiences a sound of approximate frequency $f_0(1 + v/c)$ by the Doppler effect, and this is the frequency it reflects. Since you're walking towards the wall, a second Doppler effect occurs, causing you to hear frequency $f_0(1 + 2v/c)$. (We also saw this "double Doppler shift" back in **R1**.) By similar reasoning, you hear sound of frequency $f_0(1 - 2v/c)$ from the wall behind you. Since v/c is small, what this actually means is that you'll perceive a single tone of frequency f_0, but with "beats" of frequency $4v/c$.

Example 2

In my former college at Oxford, there is a long staircase that is said to "quack" when one claps at it. What is the explanation of this phenomenon?

Solution

A diagram of the staircase is given below, courtesy of Felix Flicker, fellow of New College.



The key is that each clap reflects off a stair individually. When the echoes arrive back at the listener, they arrive quickly enough to be heard as a pitch.

The width and height of the steps are $w = 30\,\text{cm}$ and $h = 16\,\text{cm}$. Suppose one claps at a distance $L \gg w, h$. The path length differences for reflections off the bottom few steps are approximately $2w$, giving the frequency

$$f = \frac{v}{2w} = 570\,\text{Hz}$$

where we used $v = 343\,\text{m/s}$. The quack then continues, due to reflections off higher and higher stairs. Once the stairs are much further away than L, path length differences for subsequent reflections are approximately $2\sqrt{w^2 + h^2}$, giving frequency

$$f = \frac{v}{2\sqrt{w^2 + h^2}} = 500\,\text{Hz}.$$

Hence the quack consists of a pitch that starts high and then falls slightly lower as it fades away. For further discussion, see the article *How the Mound got its Quack*.

[3] **Problem 4.** USAPhO 1998, problem B1. (The official solution has a qualitatively incorrect answer for the final part of the problem; see Stefan Ivanov's errata for the correct answer.)

[3] **Problem 5.** USAPhO 2016, problem A1.

[2] **Problem 6.** Some problems about sound waves in everyday life.

(a) Get a coffee cup with a handle and tap on the rim with a spoon. You will hear two distinct pitches, e.g. if you tap directly above the handle, or 45° away from this point. Investigate what happens for different angles. Can you explain why this happens?

(b) According to introductory textbooks, the fundamental mode for a pipe of length L and radius $r \ll L$, closed at one end and open at the other, has wavelength $4L$. In reality, it's a little bit different because the radius is nonzero. Is the wavelength actually longer or shorter than $4L$?

(c) Find a way to produce beats in real life.

Solution. (a) You should find that the pitch can be slightly lower or higher, and that it's lower if you tap direction above the handle, or 90°, 180°, or 270° away from it. See this nice video for the explanation.

The basic idea is that the dominant vibrational mode deforms the circular rim into an ellipse, whose major and minor axes swap places during the oscillation. There are two possible "polarizations" for this deformation. In one of them, the handle is on a major/minor axis, so it moves; in the other, it is 45° away from these axes, so it doesn't move. In the former case, there is more inertia, so the frequency is lower.

(b) It's a little bit longer. In the introductory textbook, we model the end of the pipe as an ideal pressure node, fixed to atmospheric pressure. But in reality, the sound wave propagates a bit out of the pipe and spreads out radially. (It won't even "know" the pipe ended until it's done this.) This effect is known as the end correction, and the added length is of order r.

(c) There are a lot of ways of doing this. You could pluck strings, or use your own voice. It's even possible with some kinds of dinner fork.

2 Polarization

Now we'll introduce polarization for light waves, putting the results of **E7** to work.

Idea 2

The polarization of a light wave refers to the direction of its electric field; the light waves we saw in **E7** were linearly polarized. For example, a light wave traveling along $\hat{\mathbf{z}}$ with its polarization an angle θ from the x-axis has electric field

$$E_x(z,t) = (E_0 \cos\theta)\cos(kz-\omega t), \quad E_y(z,t) = (E_0 \sin\theta)\cos(kz-\omega t).$$

A polarizer lets only light of a certain linear polarization through; if light with a linear polarization an angle θ from this axis passes through it, then a fraction $\cos^2\theta$ of the energy is transmitted. Just as light can be incoherent, it can be unpolarized; unpolarized light hitting a polarizer loses half its energy.

[3] **Problem 7.** A simple polarizer contains many very thin, closely spaced wires. If the wires are vertical, they block vertical electric fields, allowing only horizontally polarized light to go through; this is a horizontal polarizer. One can similarly make diagonal and vertical polarizers.

(a) Suppose that perfectly monochromatic, but unpolarized light is incident on a double slit. (In this case, assume "unpolarized" means that at each instant in time, the polarization of the light passing through each slit is the same, but over longer timescales that polarization can vary.) What does the intensity pattern on the screen look like?

(b) Next, suppose a vertical polarizer is placed in front of one slit, and a horizontal polarizer is placed in front of the other slit. Now what does the intensity pattern look like?

(c) Finally, we further modify the setup of part (b) by placing many diagonal polarizers, at 45° to the vertical and horizontal, right in front of the *screen*. What does the intensity pattern on the screen look like now?

(d) On an unrelated note, suppose we wish to rotate the polarization of linearly polarized light by using $N \gg 1$ intermediate polarizers. What's the best way to do this, and what's the fraction of light that passes through the stack?

Solution. (a) It's just an ordinary double slit interference pattern. Since the light goes through the slits with the same polarization at each moment, it shows up at the screen with the same polarization at each moment. Then the amplitudes just add, so the intensities add in quadrature, yielding interference effects.

(b) Let's let the horizontal and vertical directions be x and y. Then at a given point on the screen, the light from one slit gives E_x and the light from the other slit gives E_y. The intensity is proportional to $E_x^2 + E_y^2$, so the intensities just add, with no interference term. So the interference pattern is completely destroyed. The general point is that stable interference patterns can only exist when the things you're interfering have all the same properties.

(c) The amplitude of the diagonal light is now $(E_x/\sqrt{2}) + (E_y/\sqrt{2})$, so the intensity at the screen is now proportional to $(E_x + E_y)^2$, so it's now possible to see some interference effects again.

However, it's not perfect. For example, if the light entering the slits was horizontal at some moment, then it would get totally blocked at one slit, so the result shows no interference pattern. The same is true if the light entering the slits is vertical. However, if the light entered the slits polarized at 45° to the vertical, then its amplitude would get penalized by the same factor of $1/\sqrt{2}$ at each slit, and at the screen we would get a perfect interference pattern. In practice, the incoming polarization of "unpolarized" light rapidly changes over time, so on average we get a partially visible interference pattern, i.e. there are still minima and maxima but the minima don't have zero intensity.

This is an interesting result because you might think, based on the result of part (b), that polarizers can only destroy interference patterns; however, adding more of them at the screen can actually bring them back!

Exactly the same logic can be used to explain the famous "delayed choice quantum eraser". In popular science articles, people often say this experiment proves that quantum mechanics can "rewrite the past". But it's really just a tricky interference effect that also shows up classically.

(d) For a small misalignment, the fraction of intensity lost goes up quadratically with the mismatch in angle, so we should have the polarizers uniformly spaced by angle θ/N. Then each filter multiplies the intensity by $\cos^2(\theta/N) \approx 1 - \theta^2/(2N^2)$, so the overall intensity is multiplied by

$$\left(1 - \frac{\theta^2}{2N^2}\right)^N \approx \exp\left(-\frac{\theta^2}{2N}\right).$$

In the limit $N \to \infty$, no intensity is lost. So in principle you can rotate polarization this way, though it's better to just use a half-wave plate, which will be described below.

Idea 3

For a plane wave propagating along the z-axis with general polarization, it's useful to write

$$\mathbf{E}(z,t) = \mathrm{Re}\left(\mathbf{E}_0\, e^{i(kz-\omega t)}\right)$$

where $\mathbf{E}_0$ is a complex two-component vector, describing both its amplitude and polarization. For example, if $\mathbf{E}_0 = E_0\,\hat{\mathbf{x}}$, the light wave is horizontally polarized, if $\mathbf{E}_0 = iE_0\,\hat{\mathbf{x}}$, it's horizontally polarized with a phase shifted by $\pi/2$, if $\mathbf{E}_0 = E_0\,\hat{\mathbf{y}}$ it's vertically polarized, and if $\mathbf{E}_0 = E_0\,(\hat{\mathbf{x}} + \hat{\mathbf{y}})/2$ it's diagonally polarized.

When linear polarizations are combined with a relative phase, the result is circular (or more generally, elliptical) polarization. For example, when $\mathbf{E}_0 = E_0(\hat{\mathbf{x}} + i\hat{\mathbf{y}})/\sqrt{2}$, we have

$$E_x(z,t) = \frac{E_0}{\sqrt{2}}\cos(kz-\omega t), \quad E_y(z,t) = \frac{E_0}{\sqrt{2}}\sin(kz-\omega t)$$

which is a circularly polarized light wave; the electric field at a fixed point rotates in a circle over time, and if one draws the electric field vectors in a line along $\hat{\mathbf{k}}$, they trace out a spiral. Birefringent materials, which have different indices of refraction in different directions, cause such phase shifts, and thus can convert linear polarizations into other polarizations.

Example 3

A plane wave with amplitude $\mathbf{E}_0 = E_0\,\hat{\mathbf{x}}$ enters a linear optical device, which does not absorb or reflect any energy. When the plane wave exits the device, it has circular polarization, $\mathbf{E}_0 = E_0\,(\hat{\mathbf{x}} + i\hat{\mathbf{y}})/\sqrt{2}$. What does the device do to light with vertical polarization?

Solution

Vertical polarized light has to exit with circular polarization of the other handedness, i.e. with $\mathbf{E}_0 = E_0(\hat{\mathbf{x}} - i\hat{\mathbf{y}})/\sqrt{2}$, because this is the only possibility consistent with energy conservation.

To see this, note that the energy of a light wave is proportional to the time-averaged value of $|\mathbf{E}|^2$, which is turn proportional to $|\mathbf{E}_0|^2$. Since horizontal and vertical polarizations are orthogonal, they don't interfere, so sending in both a horizontal and vertical light wave of amplitude E_0 at the same time just doubles the input energy. This must also double the output energy, and indeed, under the above ansatz we have

$$\hat{\mathbf{x}} + \hat{\mathbf{y}} \to \frac{\hat{\mathbf{x}} + i\hat{\mathbf{y}}}{\sqrt{2}} + \frac{\hat{\mathbf{x}} - i\hat{\mathbf{y}}}{\sqrt{2}} = \sqrt{2}\,\hat{\mathbf{x}}$$

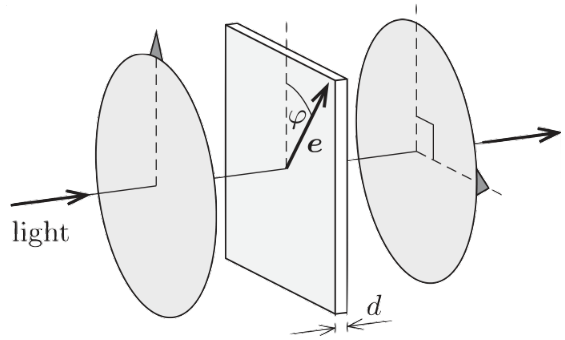
which indeed has double the energy of one wave by itself.

The more general principle here is that, since $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ were orthogonal to each other, they must be mapped to two other unit vectors which are still orthogonal, as complex vectors. That is indeed true, because

$$(\hat{\mathbf{x}} + i\hat{\mathbf{y}})^\dagger(\hat{\mathbf{x}} - i\hat{\mathbf{y}}) = \hat{\mathbf{x}} \cdot \hat{\mathbf{x}} + i^2\hat{\mathbf{y}} \cdot \hat{\mathbf{y}} = 0.$$

By the way, now that we know what the device does to both horizontal and vertically polarized light, we also know what it does to any polarization of light, by superposition.

[2] **Problem 8** (MPPP 127). A birefringent material is placed between two orthogonal polarizers. The material has thickness d, and has an index of refraction of n_1 for light linearly polarized along the axis $\mathbf{e}$, and n_2 for light polarized about an orthogonal axis.



If the system is illuminated with light of wavelength λ, give a value for d and orientation of $\mathbf{e}$ that maximizes the transmitted light.

Solution. We want to turn vertical polarization into horizontal polarization. Note that the horizontal polarization $\hat{\mathbf{x}}$ and vertical polarization $\hat{\mathbf{y}}$ can both be viewed as equal superpositions of the diagonal polarizations $(\hat{\mathbf{x}} \pm i\hat{\mathbf{y}})/\sqrt{2}$, but with opposite relative sign, i.e. a relative phase shift of π.

Thus, we want $\hat{\mathbf{e}}$ to point diagonally, $\varphi = 45^\circ$, and set d so that it flips the relative sign between these diagonal polarizations. This occurs if $d = (k + 1/2)\lambda/(|n_1 - n_2|)$ for a whole number k. This system is called a half-wave plate.

[2] **Problem 9** (MPPP 128)**.** In the first 3D movies, spectators would wear glasses with one eye tinted blue and the other tinted red. This was quickly abandoned in favor of a system that used the polarization of light.

(a) If you wear an old pair of 3D movie glasses, close one eye, and look in the mirror, then you can only see the open eye. Explain how these glasses employ light polarization. What disadvantages might this system have?

(b) If you wear a new pair of 3D movie glasses and do the same, then you can only see the closed eye. Explain why.

Solution. (a) One eye only lets vertically polarized light through, and the other only lets horizontally polarized light through, so that you can see different images with each eye. The disadvantage if that if you tilt your head, the images for each eye will get mixed together.

(b) One eye only lets clockwise polarized light through, and the other only lets counterclockwise polarized light through; the mirror flips the direction of circular polarization.

[2] **Problem 10** (HRK)**.** A quarter-wave plate is a birefringent plate that causes a $\pi/2$ phase shift between light polarized along $\mathbf{e}$ and perpendicular to $\mathbf{e}$. Similarly, a half-wave plate causes a π phase shift. Suppose you are given an object, which may be a quarter-wave plate, a half-wave plate, a linear polarizer, or just a semi-opaque disk of glass. How can you identify the object? You can use an unpolarized light source, and any number of polarizers and quarter-wave and half-wave plates.

Solution. There are many ways to approach this, but here's a way that just uses an unpolarized light source and up to two polarizers. First, a quarter-wave plate and a half-wave plate don't change the intensity of light, so if the intensity is reduced, it's either a polarizer or a semi-opaque disk. To distinguish between the latter two, you can check if the output is linearly polarized, by applying a polarizer. (For example, there should be an orientation of the polarizer where the output light is blocked entirely.)

To distinguish between a quarter-wave and half-wave plate, one could use a polarizer to send in linearly polarized light. A half-wave plate transforms linearly polarized light into linearly polarized light, with possibly a different polarization axis. That means that the output must be linearly polarized, which you can check by seeing if it can be totally blocked by a polarizer.

On the other hand, a quarter-wave plate can transform linearly polarized light into an arbitrary elliptical polarization, and if you orient it correctly, the output will be circularly polarized. When circular polarized light enters a polarizer of any orientation, exactly half of the intensity is blocked. So if the object is a quarter-wave plate, there will exist an orientation of the object so that when you when you put a polarizer after it, the output intensity is independent of that polarizer's orientation.

[2] **Problem 11** (HRK)**.** A polarizer and a quarter-wave plate are glued together so that, if the combination is placed with face A against a shiny coin, the face of the coin can be seen when illuminated by light of appropriate wavelength. When the combination is placed with face A away from the coin, the coin cannot be seen. Which component is on face A and what is the relative orientation of the components?

Solution. Let the components of the object be A and B. One simple way to think about it is to "unfold" the reflection, i.e. to replace the coin with a inverted version of this system behind it. Then the problem tells us that some light can pass through BAAB, but none can pass through ABBA.

This is only possible if A is the polarizer and B is the quarter wave plate. Some fraction of the light will always pass through BAAB. As for ABBA, suppose we let A be a vertical polarizer, and let the optical axis of B be at 45° to the vertical. Then two copies of B is effectively a half-wave plate, which converts vertically polarized light to horizontally polarized light, which gets blocked by the second copy of A. So no light goes through ABBA.

[2] **Problem 12.** Most optical elements are time-reversal symmetric, but the "Faraday rotator" is not. (It contains magnetic fields, and the direction of a magnetic field flips under time reversal.) When viewed from the top, linearly polarized light passing down through it will have its polarization rotated clockwise by θ, but light passing up through it will *also* have its polarization rotated clockwise by θ. In other words, if linearly polarized light goes through a Faraday rotator, bounces off a mirror, and returns through it, it doesn't return with the same polarization, but rather is rotated by 2θ.

(a) Explain how to use a Faraday rotator, in combination with other optical elements mentioned above, to construct an optical isolator: a system which allows some light to pass through in one direction, but none to pass through in the other direction.

(b) In **T2**, we argued that a "one-way" light filter would violate the second law of thermodynamics. So why is the setup in part (a) allowed?

Solution. (a) The simplest setup involves a horizontal polarizing filter, a Faraday rotator with $\theta = 45°$, and then a diagonal polarizing filter. When light passes from left to right, it hits the horizontal polarizing filter, gets rotated to the diagonal polarization, then leaves. When light passes from right to left, it hits the diagonal polarizing filter, gets rotated to the *vertical* polarization, then gets blocked by the horizontal polarizing filter.

(b) Unlike the ideal reflection and transmission considered in **T2**, the isolator here necessarily absorbs energy, which increases its entropy. The entropy gained by the isolator, $\Delta S = \Delta Q/T$, overwhelms the decrease in entropy from organizing photons only on one side of the isolator. (Then you might ask, what about the limit $T \to \infty$? But in that case, we must also account for the blackbody radiation emitted by both sides of the isolator, so we don't really have a functional isolator at all.)

[4] **Problem 13.** IZhO 2021, problem 3. A problem on the propagation of light through a waveguide, unifying material from **E7** and **W1**.

Solution. See the official solutions here.

3 Water Waves

Water waves are the most familiar examples of waves in everyday life, but you won't find them mentioned often in introductory textbooks, because they're far more complicated than any other kind of wave we'll consider. In all the problems below, we will completely neglect viscosity, surface tension, and compressibility of the water. Despite this, our results will still only be approximate.

[4] **Problem 14.** Consider water with depth d and density ρ. A shallow water wave (i.e. one with wavelength much greater than d) travels along the x-direction with height $h(x,t) \ll d$ relative to the water level. It turns out that the water has a horizontal velocity $v(x,t)$ which is approximately independent of height, and negligible vertical velocity. Even though the water is moving, the hydrostatic pressure formula still works because the water's vertical acceleration is negligible.

(a) Find a relation between the derivatives of $h(x,t)$ and $v(x,t)$ using conservation of mass. Using this result, show that the phase velocity v_w of the water waves obeys $v \ll v_w$.

(b) Find a relation between the derivatives of $h(x,t)$ and $v(x,t)$ using force and momentum.

(c) Combining these results, find the phase velocity v_w of shallow water waves.

Now let's consider what happens when a shallow water wave created at sea approaches the shore, and the depth d slowly decreases.

(d) Explain why waves always arrive at the shore moving perpendicular to the shoreline.

(e) If the depth is gradually halved, by what factor is the height of the wave multiplied? This phenomenon is known as shoaling.

Solution. (a) Consider the region between x and $x + dx$. The rate that water flows into this region must equal the rate of change of its volume due to the change in height. Then

$$A\, d(v(x) - v(x + dx)) = A \frac{dh}{dt}\, dx.$$

That is, we have

$$-d\, \frac{\partial v}{\partial x} = \frac{\partial h}{\partial t}.$$

For a sinusoidal wave, a derivative with respect to x gives a factor of k and a derivative with respect to t gives a factor of ω. So we have $dkv \sim \omega h$, and since $v_w = \omega/k$, we have $v \sim v_w h/d \ll v_w$.

(b) Let A be the cross-sectional area of the water, and consider the horizontal forces on the fixed piece of water that, at some moment, is between x and $x + dx$. This water has atmospheric pressure at its upper surface, so the net horizontal force on it due to hydrostatic pressure is

$$F = \frac{1}{2}\rho g A \left((d + h(x))^2 - (d + h(x + dx))^2\right) \approx -\rho g d A \frac{\partial h}{\partial x}\, dx$$

where we used the fact that $h \ll d$. This must be equal to the rate of change of momentum of this piece of water, which is

$$\frac{dp}{dt} = \rho A d\, dx\, \frac{dv}{dt} = \rho A d\, dx \left(\frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x}\right).$$

In the second step, we used the idea of the "convective derivative". That is, the chunk of water originally at x at time t will move to $x' = x + v(x,t)\, dt$ after a time dt, at which point its velocity will be $v(x', t + dt) = v(x,t) + ((\partial v/\partial t) + v(\partial v/\partial x))\, dt$, so the acceleration of the water has two terms.

Now, the rough sizes of these two terms are ωv and kv^2, so the second term is smaller by a factor of v/v_w. We thus neglect it, and conclude that

$$-g\, \frac{\partial h}{\partial x} = \frac{\partial v}{\partial t}.$$

(c) Combining the results of the last two parts, we have

$$\frac{\partial^2 h}{\partial t^2} = -d\,\frac{\partial^2 v}{\partial x\,\partial t} = gd\,\frac{\partial^2 h}{\partial x^2}$$

which is just the ideal wave equation with wave velocity $v_w = \sqrt{gd}$. (This is both the phase velocity and the group velocity.)

(d) The wave speed is proportional to $\sqrt{d}$, so the waves slow down as they approach the shore. So by Snell's law, they refract towards the normal direction.

(e) Note that nothing is adding or subtracting energy from the wave, and that the number of periods of the wave stays the same. Thus, the amount of energy in each period of the wave stays the same. (If we wanted to be fancier, we could identify an adiabatic invariant as discussed in **M4**. In this case, it is called the wave action, but the result is just the same.)

The gravitational and kinetic energy are equal on average, so for simplicity we'll consider the former. For a wave of wavelength λ, height h, and width w, the gravitational potential energy of one period is

$$E \sim \rho g \lambda w h^2.$$

The frequency of the wave stays the same, so λ is proportional to the wave speed,

$$E \propto v h^2 \propto \sqrt{d}\, h^2.$$

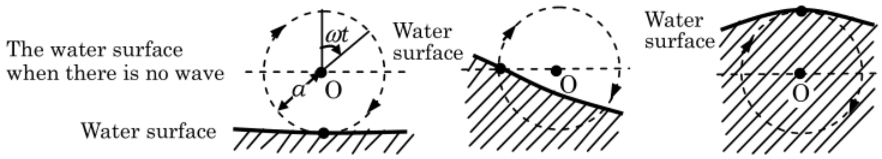
Thus, when d is halved, h increases by a factor of $2^{1/4}$. This is known as Green's law. Eventually, h becomes comparable to d and our expressions break down.

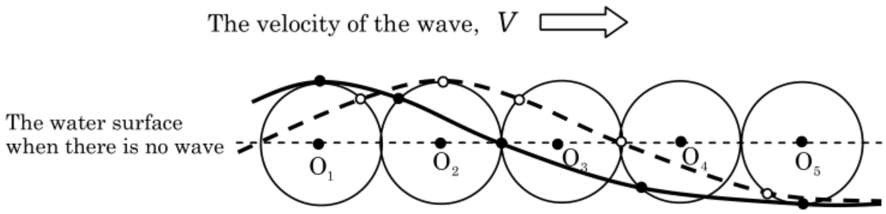
Remark

Textbooks commonly say that liquids can't support transverse waves, because they don't support shear stresses. But the waves considered in problem 14 are clearly transverse. This is possible because the textbook statement only applies to the internal forces of water *alone*. At the surface of the water, gravity provides the transverse restoring force; that's why these waves are also commonly called "gravity waves".

[2] **Problem 15** (Japan)**.** In the above problem, we considered shallow water waves, $d \ll \lambda$, in which case the motion of the water is approximately horizontal. But in general, it turns out that the motion of the water will depend on height, and also that the individual wave molecules move in ellipses. (This happens even when the wave has small amplitude, which we will always assume; the motion for large amplitude is even more complicated.) In the limit of deep water waves, $d \gg \lambda$, the water molecules move in circles. Using this fact, we can quickly derive the wave speed.

Assume the water molecules at the surface of the wave move in uniform circular motion with radius a and angular velocity ω, as shown.





(a) Consider the frame of reference moving to the right with velocity v. In this frame, the surface of the water is completely stationary, while molecules travel along the surface. Consider a small parcel of water which travels from a valley to a peak. By applying conservation of energy, derive a relationship between v, ω, and g.

(b) Find the phase and group velocity of the wave, in terms of g and the wavenumber k. In addition, find the condition on a and k for this derivation to make sense.

Showing that circular motion actually occurs takes more work, and involves solving partial differential equations; you can find a complete derivation here or here.

Solution. (a) At the bottom of the circle, the speed is $v + a\omega$, and at the top the speed is $v - a\omega$. However, since the parcel just moves along the surface of the water, no work is done on it (the water effectively just provides a normal force to keep the parcel on its surface), so this change in speed is balanced by a change in gravitational potential energy. That is,

$$\frac{1}{2}\left((v + a\omega)^2 - (v - a\omega)^2\right) = 2ga$$

which implies $v = g/\omega$. (Note that this argument only works for sufficiently small wave amplitudes, $a < v/\omega$.)

(b) The thought experiment in this problem considers an ideal plane wave, so it's computing the phase velocity. Thus, using $v_p = \omega/k$ and $v_p = g/\omega$, we have

$$v_p = \sqrt{g/k}.$$

In addition, we can solve for the dispersion relation to find

$$\omega(k) = \sqrt{gk}$$

from which we find the group velocity,

$$v_g = \frac{d\omega}{dk} = \frac{1}{2}\sqrt{g/k}.$$

Referring back to part (a), this derivation only makes sense if $v_p > a\omega$, which is equivalent to $ka < 1$. That makes sense, as when $ka \gtrsim 1$, the amplitude is large compared to the wavelength, indicating a nonlinear wave. (If we were doing a more rigorous derivation, we would find that the parcels of water only travel in circles in the small amplitude limit $ka \ll 1$.)

As you can see, water waves are quite complex. A diagram of the speeds of nine different limiting cases of water waves can be found in section 8.4 of The Art of Insight.

4 Reflection and Refraction

Now we'll introduce reflection and refraction with some real-world applications.

Idea 4

If a wave hits an interface, while traveling at an angle θ_1 to the normal to the interface, then it will generically both reflect and refract. The angle of the reflected ray is $\theta_2 = \theta_1$, and the angle of the refracted ray obeys $n_1 \sin\theta_1 = n_2 \sin\theta_2$. If there is no solution for θ_2 in the latter equation, then only reflection occurs.

These results follow directly from Huygens' principle, so they are very general, applying to light waves, sound waves, water waves, and so on, as long as the index of refraction n_i is always defined to be inversely proportional to the wave speed in each medium.

[1] **Problem 16.** Some conceptual questions about reflection and refraction.

(a) Does the index of refraction determine the phase velocity or the group velocity?

(b) Does a light beam of finite width get wider or narrower upon passing from air to water? Assume the light enters at an angle to the normal.

Solution. (a) It's the phase velocity. Recall from **E8** that Snell's law is derived by considering a plane wave incident on an interface, and demanding that the wave is appropriately continuous. This only depends on the plane wave's ω and k, which are related by the phase velocity.

(b) It gets wider because the beam bends towards the normal, as you check by drawing a sketch.

Example 4

Let the index of refraction at height h above the Earth's surface be $n(h)$. In terms of $n(0)$ and the Earth's radius R, what should dn/dh be at the surface so that light rays orbit in circles around the Earth, with constant height?

Solution

First, let's ignore the curvature of the Earth. Consider a light ray moving slightly upward, at a small angle θ to the horizontal, experiencing index of refraction n. Over a horizontal distance L, it goes up by a height $L\theta$. At this point, it will have a different angle θ' to the horizontal, and experience index of refraction $n + L\theta\, dn/dh$. Snell's law says

$$n\cos\theta = \left(n + L\theta\frac{dn}{dh}\right)\cos\theta'$$

and expanding to lowest order in the small angles θ and θ' gives

$$\frac{n}{2}\left(\theta'^2 - \theta^2\right) = L\theta\frac{dn}{dh}.$$

Approximating again to lowest order gives

$$\theta - \theta' \approx -\frac{L}{n}\frac{dn}{dh}.$$

Thus, the light ray turns through an angle of $(1/n)\,dn/dh$ per unit horizontal distance. For the light ray to stay at a constant height over the curved Earth, this must equal $1/R$, giving

$$\frac{dn}{dh} = -\frac{n(0)}{R}.$$

More generally, this calculation shows that light bends towards the direction with higher n. In the case of air, where $n - 1 \ll 1$, we can rewrite this as

$$\frac{d(n-1)}{dh} \approx -\frac{1}{R}$$

which can plausibly occur on Earth, due to the nice coincidence that $n - 1$ and H/R (where H is the typical scale height of the atmosphere) are both of order 10^{-3}.

Remark: Mirages

There are two classes of mirages.

- When $dn/dh < 0$, light rays bend down. If there is a distant object at the horizon, its image will appear *above* the horizon. This is called a "superior" mirage.
- When $dn/dh > 0$, light rays bend up. Then a distant object at the horizon will appear *below* the horizon, forming an "inferior" mirage. This also applies to the sky near the horizon, producing the illusion of water on the ground sometimes seen in deserts.

In air, the refractive index is close to 1, and $n - 1 \propto \rho \propto P/T$, where ρ is the air density and the second step used the ideal gas law. Usually we have $d\rho/dh < 0$, since $dP/dh < 0$ in hydrostatic equilibrium, but it depends on the value of dT/dh.

- In normal conditions, the Sun warms the ground and the hot air rises and adiabatically mixes the atmosphere (as discussed in **T1**), so that $dT/dh < 0$. This partially cancels the effect of the pressure variation, so that dn/dh is still negative but has small magnitude, so that mirage effects aren't apparent.
- In rare "thermal inversion" conditions, we have $dT/dh > 0$, so that dn/dh is negative with large magnitude, leading to strong superior mirage effects. If dn/dh is negative enough, it can match the value computed in example 4, allowing an observer to see arbitrarily far along the horizon despite the curvature of the Earth. This was the reason the famous Bedford Level experiment concluded the Earth was flat.
- In hot deserts, the air near the ground is very hot, so that $dT/dh < 0$ with a large magnitude. (A strongly negative dT/dh also occurs in cold days above water, since the water stays warmer than the air above it.) Here the temperature gradient overpowers the pressure gradient, so that $dn/dh > 0$ and inferior mirages can occur.

Proponents of the flat Earth hypothesis claim that the Earth only seems curved due to atmospheric refraction. But they have it backwards: in almost all conditions $dn/dh < 0$, which makes the Earth look *less* curved than it actually is.

[4] **Problem 17.** IPhO 1995, problem 2. Refraction in the presence of a linearly varying wave speed. (This is a classic setup with a neat solution, also featured in IPhO 1974, problem 2.)

[3] **Problem 18.** INPhO 2019, problem 1. Another exercise on refraction, with an uglier solution.

Solution. See the official solutions here.

[3] **Problem 19.** IPhO 2003, problem 3B. An exercise on refraction and radiation pressure.

[4] **Problem 20.** IPhO 1993, problem 2. Another exercise on the same theme.

5 Ray Tracing

Idea 5

A pointlike object emits light rays in all directions. When those light rays subsequently converge at some other point, that point is the object's real image. If they don't actually converge, but all propagate outward with a common center, that point is the object's virtual image. In general, if we're given that an image exists, we can find its location by following the paths of selected rays from the object and looking for intersections.

[2] **Problem 21.** A pinhole camera is a simplified camera with no lens. It simply consists of a box with a small hole (the "aperture"). An image of the outside appears on the inside of the box (the "screen"), opposite the hole.

(a) Explain how the pinhole camera works by ray tracing.

(b) What are the disadvantages of having a larger or smaller aperture?

(c) Assuming the object being photographed is very bright, estimate the optimal aperture size for taking a clear picture with a pinhole camera, for a box of side length L.

Solution. (a) Consider a point P on an object outside. Light is emitted from P in every direction. If the front of the box was just open, then light rays from P could hit the whole back of the box, brightening the whole screen. But if there's a pointlike hole, then only one ray from P can go through the hole, and that ray hits only one point on the back of the box, making a sharp image of P there.

This is very different from how images are formed by lenses. A lens tries to redirect the light rays from a source so that many of them hit the same point on the screen. A pinhole just removes the unwanted rays.

(b) Using a larger aperture would make the image blurrier, since more rays can get through. Using a smaller aperture would make the image dimmer; also, for very small holes, diffraction becomes more important, and makes the image blurrier again.

(c) If the aperture size is a, then two rays can enter the aperture and end up at the same point on the screen even if their directions are different by $\Delta\theta \sim a/L$. This is the "geometric optics" blurring effect, which is minimized by having a smaller hole. At the same time, diffraction causes light passing through the hole to spread out by $\Delta\theta \sim \lambda/a$, which is minimized by having a larger hole. The optimum occurs when the two are comparable, so $a \sim \sqrt{\lambda L}$. For visible light and a camera-sized box, this is a fraction of a millimeter, which you can easily achieve using a needle or pencil.

Pinhole cameras are extremely common in everyday life. They can form in the gaps between leaves; the resulting dappled light on the ground is just many images of the disk of the sun.

The next three problems will exercise your intuition with real-world examples.

[2] **Problem 22.** AuPhO 2020, section C.

Solution. See the official solutions here.

[2] **Problem 23.** AuPhO 2013, problem 11. Write your answers on the official answer booklet.

Solution. See the official solutions here.

[3] **Problem 24.** AuPhO 2019, problem 12. Write your answers on the official answer booklet.

Solution. See the official solutions here.

[3] **Problem 25.** IZhO 2020, problem 1.3. A test of your intuition for 3D ray tracing.

Solution. See the official solutions here. But note that, as pointed out by Stefan Ivanov here, the official solution gets the thicknesses of the borders wrong. In the first part, the thickness of the border of the triangle should be $2r_1 = 2\,\text{mm}$. In the second part, the thickness of the border of the star should be $3r_2 = 0.3\,\text{mm}$.

By the way, this isn't some random question cooked up for an Olympiad; it's a real effect in pinhole cameras that puzzled physicists of the past. As you can read here, this effect distorts the apparent sizes of the Sun and Moon, which puzzled Brahe. The problem was solved by Kepler in 1600, who developed basically the exact same ideas you did when solving this problem.

Also, there's an analogous phenomenon with shadows which you've seen many times in real life: the shadow of an object lit by an extended light source contains an umbra and penumbra.

Idea 6

Conic sections have some simple properties under reflection.

- Light rays emitted from one focus of an ellipse will all be reflected to its other focus.
- Light rays emitted from one focus on a hyperbola will all be reflected so that the resulting rays all travel radially outward from the other focus.
- Parallel light rays entering a parabola along its symmetry axis (i.e. the axis perpendicular to the directrix) will all be reflected to its focus.

In the language of idea 5, if the foci of an ellipse/hyperbola are called F_1 and F_2, then an object at F_1 produces a real/virtual image at F_2. Note that a parabola is simply an ellipse in the limit where F_1 becomes very far away, so that rays coming in from F_1 become approximately parallel.

[2] **Problem 26** (Povey)**.** The mirascope is a toy consisting of two parabolic mirrors, pointing toward each other, so that the focus of each one is at the vertex of the other.

(a) When an object is placed at the bottom vertex, a real image appears at the top vertex. Why?

(b) How is the image oriented relative to the object?

The real image made by this setup is very convincing. There's a Michelin starred restaurant that uses it in a course: when you reach for what looks like the food, your hand just passes through air.

Solution. (a) Rays departing from the bottom vertex reflect off the top mirror and end up going vertically downward. They then reflect off the bottom mirror and end up focused at the top vertex, which is where the image appears.

(b) To figure this out, you need to trace some rays starting from points near the bottom vertex. The result is that the image is flipped in the horizontal directions but not the vertical directions. For example, if the object is a little pig standing up and facing to the right, the image is a little pig standing up and facing to the left.

Idea 7: Paraxial Approximation

If a light ray hits a thin lens of focal length f at a shallow angle, and at a distance $y \ll f$ above the lens's center, then it will exit the lens bent vertically by an angle $\pm y/f$, where the sign depends on whether the lens is converging or diverging. (For example, any light ray going straight through the lens's center isn't bent at all.) This is the paraxial approximation, which only holds for light rays incident at shallow angles near the center of the lens.

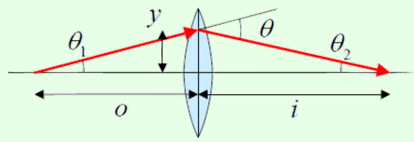
Conversely, if you don't know the focal length of a system, you can use this idea to find it. For example, the lensmaker's equation, giving the focal length of a lens of radii of curvature R_1 and R_2 and thickness d, is

$$\frac{1}{f} = (n-1)\left(\frac{1}{R_1} - \frac{1}{R_2} + \frac{(n-1)\,d}{nR_1R_2}\right)$$

and can be derived by computing the bending of the light ray at each interface.

Example 5

An object is placed a distance o behind a thin converging lens with focal length f.



An image is formed a distance i in front of the lens. How are o, i, and f related?

Solution

The horizontal light ray goes straight through, so let's consider another light ray which emerges at a small angle θ_1 to the horizontal. Then we read off

$$\theta_1 \approx \frac{y}{o}, \quad \theta_2 \approx \frac{y}{i}$$

but their sum is the deflection y/f, from which we conclude

$$\frac{1}{o} + \frac{1}{i} = \frac{1}{f}.$$

This is the familiar thin lens equation.

[2] **Problem 27.** Some basic questions about ray tracing.

(a) Show that if you have height h, you can see your entire body in a vertical mirror of length $h/2$ whose top is at eye level, no matter how close or far you stand from it. Nonetheless, when people want to look at their whole body in the mirror, they typically stand back. Why?

(b) A candle is placed behind a converging lens. An image is formed on a screen on the other side of the lens. Now suppose that the *top* half of the lens is covered with a black cloth. Describe how the image changes.

Solution. (a) This is clear from placing your image an equal distance behind the mirror and drawing similar triangles. It is a popular "gotcha" question in the physics education literature, which supposedly proves that many people don't know how their own eyes work. But there are good reasons to stand back. If you're right next to the mirror, the image of your head will be right next to you, while the image of your feet will be several feet away, so you won't be able to focus the light from both at once. Also, your sharpest vision only occupies a small part of the center of your field of view. So to look at yourself, your eyes have to constantly dart all the way up and down while adjusting their focus. This is fixed by standing back.

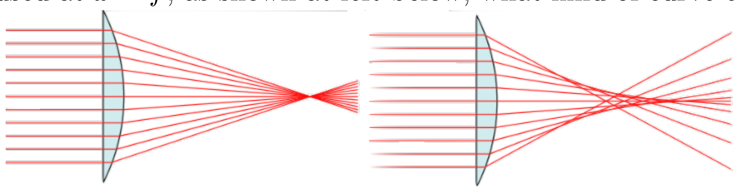
(b) It is tempting to say that half of the candle's image disappears, but that's not right. Ray tracing shows that you can get a complete image of the candle, since there are always rays that pass through the bottom half of the lens. Instead, by blocking half the lens, the image gets half as bright.

[3] **Problem 28.** USAPhO 2024, problem A3. A series of optics exercises relevant for real cameras.

Idea 8: Fermat's Principle

For fixed starting and ending points, light always takes the path of least time. This implies that if light from point P is all focused at point P', then all the relevant paths from P to P' take the same time. This principle is completely equivalent to the laws of reflection and refraction above, but may be more useful in certain situations.

[2] **Problem 29.** Parallel light rays coming in along the $+\hat{\mathbf{x}}$ direction enter a lens of index of refraction n, whose left edge is at $x = 0$ and whose right edge is described by the function $x(y)$. If all the light beams are to be focused at $x = f$, as shown at left below, what kind of curve does $x(y)$ have to be?



You should find that $x(y)$ is not an arc of a circle, which implies that a spherical lens will fail to focus all incoming horizontal light to a point. Instead, we will get spherical aberration, as shown at right above. However, most lenses are spherical because it's easier to make them that way.

Solution. The easiest method is to directly use Fermat's principle. For the ray coming in at height y, the total travel time is independent of y, so that

$$nx(y) + \sqrt{y^2 + (f - x(y))^2} = a$$

for a constant a. Solving for y shows that this is part of a hyperbola.

For more practice with geometric optics, I recommend Stefan Ivanov's collection of Russian problems.