UNIVERSITY OF AGRICULTURE, ABEOKUTA, NIGERIA

DEPARTMENT OF PHYSICS

PHS 342: ELECTROMAGNETIC WAVES AND OPTICS (3 UNITS)

By

Prof. J. A. Olowofela and Dr. R. Bello

Acknowledgement: This lecture note was part of lecture note written by Professor Richard Fitzpatrick, a Professor of Physics at the University of Texas at Austin. Many thanks to him for making the lecture note freely available.

1 Electromagnetic Waves

1.1 Maxwell's Equations

In the latter half of the nineteenth century, the Scottish physicist James Clerk Maxwell demonstrated that all previously established experimental facts regarding electric and magnetic fields could be summed up in just four equations. Nowadays, these equations are generally known as Maxwell's equations.

The first equation is simply Gauss' law. This equation describes how electric charges generate electric fields. Gauss' law states that:

> The electric flux through any closed surface is equal to the total charge enclosed by the surface, divided by ϵ_0 .

This can be written mathematically as

$$\oint s \ [\![E.dS]\!] = Q/\varepsilon_0 \tag{1.1}$$

where S is a closed surface enclosing the charge Q. The above expression can also be written $\oint s \ [\![E.dS]\!] = 1/\varepsilon_0 \oint s(\rho dV)$ (1.2)

where V is a volume bounded by the surface S, and ρ is the charge density: i.e., the electric charge per unit volume.

The second equation is the magnetic equivalent of Gauss' law. This equation describes how the non-existence of magnetic monopoles causes magnetic field-lines to form closed loops. Gauss' law for magnetic fields states that:

> The magnetic flux through any closed surface is equal to zero. This can

be written mathematically as

$$\oint_S B.dS = 0, \tag{1.3}$$

where S is a closed surface.

The third equation is Faraday's law. This equation describes how changing magnetic fields generate electric fields. Faraday's law states that:

> The line integral of the electric field around any closed loop is equal to minus the time rate of change of the magnetic flux through the loop.

This can be written mathematically as

$$\oint_c E.ds = -\frac{d}{dt}\oint_{s'} \mathrm{B}\,.\,\mathrm{ds}' \qquad (1.4)$$

where S' is a surface attached to the loop C.

The fourth, and final, equation is Ampère's circuital law. This equation describes how electric currents generates magnetic fields. Ampère's circuital law states that:

> The line integral of the magnetic field around any closed loop is equal to μ_0 times the algebraic sum of the currents which pass through the loop.

This can be written mathematically as

$$\oint s(B.d\boldsymbol{r}) = \mu_0 I \qquad (1.5)$$

where I is the net current flowing through loop C. This equation can also be written

$$\oint_C \mathrm{B}\cdot \mathrm{dr} = \mu_0 \int_{S'} \boldsymbol{j}.d\boldsymbol{S} \qquad (1.6)$$

where S' is a surface attached to the loop C, and j is the current density: i.e., the electrical current per unit area.

When Maxwell first wrote Eqs. (1.2), (1.3), (1.4), and (1.6) he was basically trying to summarize everything which was known at the time about electric and magnetic fields in mathematical form. However, the more Maxwell looked at his equations, the more convinced he became that they were incomplete. Eventually, he proposed adding a new term, called the displacement current, to the right-hand side of his fourth equation. In fact, Maxwell was able to show that (1.2), (1.3), (1.4), and (1.6) are mathematically inconsistent unless the displacement current term is added to Eq. (1.6). Unfortunately, Maxwell's demonstration of this fact requires some advanced mathematical techniques which lie well beyond the scope of this course. In the following, we shall give a highly simplified version of his derivation of the missing term.

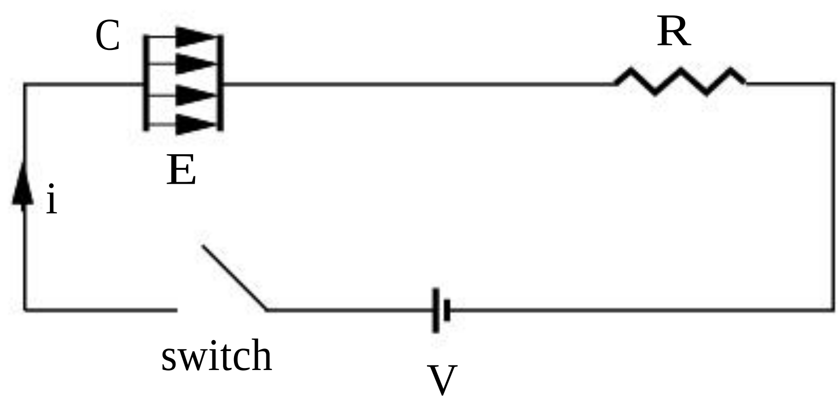


Figure 1.1: Circuit containing a charging capacitor.

Consider a circuit consisting of a parallel plate capacitor of capacitance C in series with a resistance R and an steady emf V , as shown in Fig. 1.1. Let A be the area of the capacitor plates, and let d be their separation. Suppose that the switch is closed at $t = 0$. The current i flowing around the circuit starts from an initial value of $I = V/R$, and gradually decays to zero on the RC time of the circuit. Simultaneously, the charge q on the positive plates of the capacitor starts from zero, and gradually increases to a final value of $Q = C\,V$. As the charge q varies, so does the potential difference v between the capacitor plates, since $v = q/C$.

The electric field in the region between the plates is approximately uniform, directed perpendicular to the plates (running from the positively charged plate to the negatively charged plate), and is of magnitude $E = v/d$. It follows that

$$q = C\,v = C\,d\,E. \tag{1.7}$$

In a time interval dt, the charge on the positive plate of the capacitor increases by an amount dq = C d dE, where dE is the corresponding increase in the electric field-strength between the plates. Note that both C and d are time-independent quantities. It follows that

$$\frac{dq}{dt} = Cd\frac{dE}{dt}, \tag{1.8}$$

Now, dq/dt is numerically equal to the instantaneous current i flowing around the circuit (since all of the charge which flows around the circuit must accumulate on the plates of the capacitor). Also, C = ϵ_0 A/d for a parallel plate capacitor. Hence, we can write

$$i = \frac{dq}{dt} = Cd\frac{dE}{dt} = \varepsilon_0 A\frac{dE}{dt} \qquad (1.9$$

Since the electric field E is normal to the area A, ~~we c~~an also write

$$i = \epsilon_0\, A\frac{dE_\perp}{dt} \tag{1.10}$$

Equation (1.10) relates the instantaneous current flowing around the circuit to the time rate of change of the electric field between the capacitor plates. According to Eq. (1.6), the current flowing around the circuit generates a magnetic field. This field circulates around the current carrying wires connecting the various components of the circuit. However, since there is no actual current flowing between the plates of the capacitor, no magnetic field is generated in this region, according to Eq. (1.6). Maxwell demonstrated that for reasons of mathematical self-consistency there must, in fact, be a magnetic field generated in the region between the plates of the capacitor. Furthermore, this magnetic field must be the same as that which would be generated if the current i (i.e., the same current as that which flows around the rest of the circuit) flowed between the plates. Of course, there is no actual current flowing between the plates. However, there is a changing electric field. Maxwell argued that a changing electric field constitutes an effective current (i.e., it generates a magnetic field in just the same manner as an actual current). For historical reasons (which do not particularly interest us at the moment), Maxwell called this type of current a displacement current. Since the displacement current I_D flowing between the plates of the capacitor must equal the current i flowing around the rest of the circuit, it follows from

Eq. (1.10) that

$$I_D = \varepsilon_0 A \frac{dE_\perp}{dt} \tag{1.11}$$

Equation (1.11) was derived for the special case of the changing electric field generated in the region between the plates of a charging parallel plate capacitor. Nevertheless, this equation turns out to be completely general. Note that $A\ E_\perp$ is equal to the electric flux Φ_E between the plates of the capacitor. Thus, the most general expression for the displacement current passing through some closed loop is

$$I_D = \epsilon_0 \frac{d\Phi_E}{dt}, \tag{1.12}$$

where Φ_E is the electric flux through the loop.

According to Maxwell's argument, a displacement current is just as effective at generating a magnetic field as a real current. Thus, we need to modify Ampère's circuital law to take displacement currents into account. The modified law, which is known as the Ampère-Maxwell law, is written

> The line integral of the electric field around any closed loop is equal to μ_0 times the algebraic sum of the actual currents and which pass through the loop plus μ_0 times the displacement current passing through the loop.

This can be written mathematically as

$$\oint_C \mathbf{B} \cdot d\mathbf{r} = \mu_0 \, (I + I_D), \tag{1.13}$$

where C is a loop through which the electric current I and the displacement current I_D pass. This equation can also be written

$$\oint_C \mathbf{B} \cdot d\mathbf{r} = \mu_0 \int_{S'} j \cdot dS' + \mu_0 \, \epsilon_0 \, \frac{d}{dt} \int_{S'} \mathbf{E} \cdot d\mathbf{S}', \tag{1.14}$$

where S' is a surface attached to the loop C.

Equations (1.2), (1.3), (1.4), and (1.14) are known collectively as Maxwell's equations. They constitute a complete and mathematically self-consistent description of the behaviour of electric and magnetic fields.

1.2 Electromagnetic Waves

One of the first things that Maxwell did with his four equations, once he had obtained them, was to look for wave-like solutions. Maxwell knew that the wave-like solutions of the equations of gas dynamics correspond to sound waves, and the wave-like solutions of the equations of fluid dynamics correspond to gravity waves in water, so he reasoned that if his equations possessed wave-like solutions then these would correspond to a completely new type of wave, which he called an electromagnetic wave.

Maxwell was primarily interested in electromagnetic waves which can propagate through a vacuum (i.e., a region containing no charges or currents). Now, in a vacuum, Maxwell's equations reduce to

$$\oint_S \mathbf{E}\cdot d\mathbf{S}, \quad = \quad 0, \tag{1.15}$$

$$\oint_S \mathbf{B}\cdot d\mathbf{S} \quad = \quad 0, \tag{1.16}$$

$$\oint_C \mathbf{E}\cdot d\mathbf{r} \quad = \quad -\frac{d}{dt}\int_{S'} \boldsymbol{B}.\,d\boldsymbol{S}' \tag{1.17}$$

$$\oint_C \mathbf{B}\cdot d\mathbf{r} \quad = \quad \mu_0\,\epsilon_0\frac{d}{dt}\int_{S'} \boldsymbol{E}.\,d\boldsymbol{S}' \tag{1.18}$$

where S is a closed surface, and S' a surface attached to some loop C. Note that, with the addition of the displacement current term on the right-hand side of Eq. (1.18), these equations exhibit a nice symmetry between electric and magnetic fields. Unfortunately, Maxwell's mathematical proof that the above equations possess wave-like solutions lies well beyond the scope of this course. We can, nevertheless, still write down these solutions, and comment on them.

Consider a plane electromagnetic wave propagating along the z-axis. According to Maxwell's calculations, the electric and magnetic fields associated with such a wave take the form

$$E_x \;=\; E_0\,\cos[2\pi\,(z/\lambda - f\,t)], \tag{1.19}$$

$$B_y \;=\; B_0\,\cos[2\pi\,(z/\lambda - f\,t)]. \tag{1.20}$$

Note that the fields are periodic in both time and space. The oscillation frequency (in hertz) of the fields at a given point in space is f. The equation of a wave crest is

$$\frac{z}{\lambda} - ft = N \qquad (1.21$$

where N is an integer. It can be seen that the distance along the z-axis between successive wave crests is given by λ. This distance is conventionally termed the wavelength. Note that each wave crest propagates along the z-axis. In a time interval dt, the Nth wave crest moves a distance dz = λ f dt, according to Eq. (1.21). Hence, the velocity c = dz/dt with which the wave propagates along the z-axis is given by

$$c = f\lambda. \qquad (1.22)$$

Maxwell was able to establish that electromagnetic waves possess the following properties:

1. The magnetic field oscillates in phase with the electric field. In other words, a wave maximum of the magnetic field always coincides with a wave maximum of the electric field in both time and space.

2. The electric field is always perpendicular to the magnetic field, and both fields are directed at right-angles to the direction of propagation of the wave. In fact, the wave propagates in the direction E × B. Electromagnetic waves are clearly a type of transverse wave.

3. For a z-directed wave, the electric field is free to oscillate in any direction which lies in the x-y plane. The direction in which the electric field oscillates is conventionally termed the direction of polarization of the wave. Thus, Eqs. (1.19) represent a plane electromagnetic wave which propagates along the z-axis, and is polarized in the x-direction.

4. The maximum amplitudes of the electric and the magnetic fields are related via

$$E_0 = c\,B_0\,. \qquad (1.23)$$

5. There is no constraint on the possible frequency or wavelength of electromagnetic waves. However, the propagation velocity of electromagnetic waves is fixed, and takes the value

$$c = \frac{1}{\sqrt{\mu_0\,\epsilon_0}}. \tag{1.24}$$

According to Eqs. (1.17) and (1.18), a changing magnetic field generates an electric field, and a changing electric field generates a magnetic field. Thus, we can think of the propagation of an electromagnetic field through a vacuum as due to a kind of "leap-frog" effect, in which a changing electric field generates a magnetic field, which, in turn, generates an electric field, and so on. Note that the displacement current term in Eq. (1.18) plays a crucial role in the propagation of electromagnetic waves. Indeed, without this term, a changing electric field is incapable of generating a magnetic field, and so there can be no leap-frog effect. Electromagnetic waves have many properties in common with other types of wave (e.g., sound waves). However, they are unique in one respect: i.e., they are able to propagate through a vacuum. All other types of waves require some sort of medium through which to propagate.

Maxwell deduced that the speed of propagation of an electromagnetic wave through a vacuum is entirely determined by the constants μ_0 and ϵ_0 [see Eq. (1.24)]. The former constant is related to the strength of the magnetic field generated by a steady current, whereas the latter constant is related to the strength of the electric field generated by a stationary charge. The values of both constants were well known in Maxwell's day. In modern units, $\mu_0 = 4\pi \times 10^{-7}\,\mathrm{N\,s^2\,C^{-2}}$ and $\epsilon_0 = 8.854 \times 10^{-12}\,\mathrm{C^2\,N^{-1}\,m^{-2}}$. Thus, when Maxwell calculated the velocity of electromagnetic waves he obtained

$$c = \frac{1}{\sqrt{(4\pi \times 10^{-7})\,(8.854 \times 10^{-12})}} = 2.998 \times 10^{8}\,m/s. \tag{1.25}$$

Now, Maxwell knew [from the work of Fizeau (1849) and Foucault (1850)] that the velocity of light was about $3 \times 10^{8}\ \mathrm{m\ s^{-1}}$. The remarkable agreement between this experimentally determined velocity and his theoretical prediction for the velocity of electromagnetic waves immediately lead Maxwell to hypothesize that light is a form of electromagnetic wave. Of course, this hypothesis turned out to be correct. We can still appreciate that Maxwell's achievement in identifying light as a form of electromagnetic wave was quite remarkable. After all, his equations were derived from the results of bench-top laboratory experiments involving charges, batteries, coils, and currents, etc., which apparently had nothing whatsoever to do with light.

Maxwell was able to make another remarkable prediction. The wavelength of light was well known in the late nineteenth century from studies of diffraction through slits, etc. Visible light actually occupies a surprisingly narrow range of wavelengths. The shortest wavelength blue light which is visible has a wavelength of $\lambda = 0.40$ microns (one micron is 10^{-6} meters). The longest wavelength red light which is visible has a wavelength of $\lambda = 0.76$ microns. However, there is nothing in Maxwell's analysis which suggested that this particular range of wavelengths is special. In principle, electromagnetic waves can have any wavelength. Maxwell concluded that visible light forms a small element of a vast spectrum of previously undiscovered types of electromagnetic radiation.

Since Maxwell's time, virtually all of the non-visible parts of the electromagnetic spectrum have been observed. Table 1.1 gives a brief guide to the electromagnetic spectrum. Electromagnetic waves are of particular importance because they are our only source of information regarding the Universe around us. Radio waves and microwaves (which are comparatively hard to scatter) have provided much of our knowledge about the centre of the Galaxy. This is completely unobservable in visible light, which is strongly scattered by interstellar gas and dust lying in the galactic plane. For the same reason, the spiral arms of the Galaxy can only be mapped out using radio waves. Infrared radiation is useful for detecting proto-stars which are not yet hot enough to emit visible radiation. Of course, visible radiation is still the mainstay of astronomy. Satellite based ultraviolet observations have yielded invaluable insights into the structure and distribution of distant galaxies. Finally, X-ray and γ-ray astronomy usually concentrates on exotic objects in the Galaxy such as pulsars and supernova remnants.

Radiation Type	Wavelength Range (m)
Gamma Rays	$< 10^{-11}$
X-Rays	10^{-11} - 10^{-9}
Ultraviolet	10^{-9} - 10^{-7}
Visible	10^{-7} - 10^{-6}
Infrared	10^{-6} - 10^{-4}
Microwave	10^{-4} - 10^{-1}
TV-FM	10^{-1} - 10^{1}
Radio	$> 10^{1}$

Table 1.1: The electromagnetic spectrum.

1.3 Effect of Dielectric Materials

It turns out that electromagnetic waves cannot propagate very far through a conducting medium before they are either absorbed or reflected. However, electromagnetic waves are able to propagate through transparent dielectric media without difficultly. The speed of electromagnetic waves propagating through a dielectric medium is given by

$$c' = \frac{c}{\sqrt{K}}, \qquad (1.26)$$

where K is the dielectric constant of the medium in question, and c the velocity of light in a vacuum. Since K > 1 for dielectric materials, we conclude that:

> The velocity with which electromagnetic waves propagate through a dielectric medium is always less than the velocity with which they propagate through a vacuum.

1.4 Energy in Electromagnetic Waves

The energy stored per unit volume in an electromagnetic wave is given by

$$w = \varepsilon_0 \frac{E^2}{2} + \frac{B^2}{2\mu_0}. \qquad (1.27)$$

Since, B = E/c, for an electromagnetic wave, and c = $1/\sqrt{\mu_0 \epsilon_0}$, the above expression yields

$$w = \frac{\varepsilon_0 E^2}{2} + \frac{E^2}{2\mu_0 c^2} = \frac{\varepsilon_0 E^2}{2} + \frac{\varepsilon_0 E^2}{2}, \quad (1.28)$$

or

$$w = \epsilon_0 E^2 . \quad (1.29)$$

It is clear, from the above, that half the energy in an electromagnetic wave is carried by the electric field, and the other half is carried by the magnetic field.

As an electromagnetic field propagates it transports energy. Let P be the power per unit area carried by an electromagnetic wave: i.e., P is the energy transported per unit time across a unit cross-sectional area perpendicular to the direction in which the wave is traveling. Consider a plane electromagnetic wave propagating along the z-axis. The wave propagates a distance c dt along the z-axis in a time interval dt. If we consider a cross-sectional area A at right-angles to the z-axis, then in a time dt the wave sweeps through a volume dV of space, where dV = A c dt. The amount of energy filling this volume is

$$dW = w\, dV = \epsilon_0 E^2 A\, c\, dt. \quad (1.30)$$

It follows, from the definition of P, that the power per unit area carried by the wave is given by

$$P = \frac{dW}{Adt} = \frac{\varepsilon_0 E^2 Acdt}{Adt} \quad (1.31)$$

so that

$$P = \epsilon_0 E^2 c. \quad (1.32)$$

Since half the energy in an electromagnetic wave is carried by the electric field, and the other half is carried by the magnetic field, it is conventional to convert the above expression into a form involving both the electric and magnetic field strengths. Since, E = c B, we have

$$P = \epsilon_0 c E (c B) = \epsilon_0 c^2 E B = \frac{EB}{\mu_o} \quad (1.33)$$

Thus, $$P = \frac{E B}{\mu_0} \quad (1.34)$$

Equation (1.34) specifies the power per unit area transported by an electromagnetic wave at any given instant of time. The peak power is given by

$$P_0 = \frac{E_0 B_0}{\mu_0}, \tag{1.35}$$

where E_0 and B_0 are the peak amplitudes of the oscillatory electric and magnetic fields, respectively. It is easily demonstrated that the average power per unit area transported by an electromagnetic wave is half the peak power, so that

$$S = \overline{P} = \frac{E_0 B_0}{2\mu_0} = \frac{\varepsilon_0 c E_0}{2} = \frac{c B_0^{\,2}}{2\mu_0} \tag{1.36}$$

The quantity S is conventionally termed the intensity of the wave.

1.5 Worked Examples

Example 1.1: Electromagnetic waves

Question: Consider electromagnetic waves of wavelength λ = 30 cm in air. What is the frequency of such waves? If such waves pass from air into a block of quartz, for which K = 4.3, what is their new speed, frequency, and wavelength?

Answer: Since, f λ = c, assuming that the dielectric constant of air is approximately unity, it follows that

$$f = \frac{c}{\lambda} = \frac{(3x10^8)}{(0.3)} = 1\ x\ 10^9 Hz.$$

The new speed of the waves as they pass propagate through the quartz is

$$c' = \frac{c}{\sqrt{K}} = \frac{(3\ x\ 10^{8)}}{\sqrt{4.3}} = 1.4\ x\ 10^8 m/s$$

The frequency of electromagnetic waves does not change when the medium through which the waves are propagating changes. Since c ′ = f λ for electro-magnetic waves propagating through a dielectric medium, we have

$$\lambda_{quartz} = \frac{c'}{f} = \frac{(1.4\ x\ 10^8)}{(1\ x\ 10^9)} = 14cm$$

Example 1.2: Intensity of electromagnetic radiation

Question: Suppose that the intensity of the sunlight falling on the ground on a particular day is 140 W m^{-2} . What are the peak values of the electric and magnetic fields associated with the incident radiation?

Answer: According to Eq. (1.36), the peak electric field is given by

$$E_0 = \sqrt{\frac{2\mu_0 S}{c\varepsilon_0}} = \sqrt{\frac{(2)(140)}{(3x10^8)(8.85x10^{-12})}} = 324.7V / m.$$

Likewise, the peak magnetic field is given by

$$B_0 = \sqrt{\frac{2\mu_0 S}{c}} = \sqrt{\frac{(2)(4\pi x10^{-7})(140)}{(3x10^8)}} = 1.083x10^{-6}T.$$

2 Geometric Optics

2.1 Introduction

Optics deals with the propagation of light through transparent media, and its interaction with mirrors, lenses, slits, etc. Optical effects can be divided into two broad classes. Firstly, those which can be explained without reference to the fact that light is fundamentally a wave phenomenon, and, secondly, those which can only be explained on the basis that light is a wave phenomenon. Let us, for the moment, consider the former class of effects. It might seem somewhat surprising that any optical effects at all can be accounted for without reference to waves. After all, as we saw in Sect. 1, light really is a wave phenomenon. It turns out, however, that wave effects are only crucially important when the wavelength of the wave is either comparable to, or much larger than, the size of the objects with which it interacts. When the wavelength of the wave becomes much smaller than the size of the objects with which it interacts then the interactions can be accounted for in a very simple geometric manner, as explained in this section. Since the wavelength of visible light is only of order a micron, it is very easy to find situations in which its wavelength is very much smaller than the size of the objects with which it interacts. Thus, “wave-less” optics, which is usually called geometric optics, has a very wide range of applications.

In geometric optics, light is treated as a set of rays, emanating from a source, which propagate through transparent media according to a set of three simple laws. The first law is the law of rectilinear propagation, which states that light rays propagating through a homogeneous transparent medium do so in straight-lines. The second law is the law of reflection, which governs the interaction of light rays with conducting surfaces (e.g., metallic mirrors). The third law is the law of refraction, which governs the behaviour of light rays as they traverse a sharp boundary between two different transparent media (e.g., air and glass).

2.2 History of Geometric Optics

Let us first consider the law of rectilinear propagation. The earliest surviving optical treatise, Euclid's Catoptrics[1] (280 BC), recognized that light travels in straightlines in homogeneous media. However, following the teachings of Plato, Euclid (and all other ancient Greeks) thought that light rays emanate from the eye, and intercept external objects, which are thereby "seen" by the observer. The ancient Greeks also thought that the speed with which light rays emerge from the eye is very high, if not infinite. After all, they argued, an observer with his eyes closed can open them and immediately see the distant stars.

Hero of Alexandria, in his Catoptrics (first century BC), also maintained that light travels with infinite speed. His argument was by analogy with the free fall of objects. If we throw an object horizontally with a relatively small velocity then it manifestly does not move in a straight-line. However, if we throw an object horizontally with a relatively large velocity then it appears to move in a straight-line to begin with, but eventually deviates from this path. The larger the velocity with which the object is thrown, the longer the initial period of apparent rectilinear motion. Hero reasoned that if an object were thrown with an infinite velocity then it would move in a straight-line forever. Thus, light, which travels in a straight-line, must move with an infinite velocity. The erroneous idea that light travels with an infinite velocity persisted until 1676, when the Danish astronomer Olaf Römer demonstrated that light must have a finite velocity, using his timings of the successive eclipses of the satellites of Jupiter, as they passed into the shadow of the planet.

The first person to realize that light actually travels from the object seen to the eye was the Arab philosopher "Alhazan" (whose real name was Abu'ali al-hasan ibn al-haytham), who published a book on optics in about 1000 AD.

The law of reflection was correctly formulated in Euclid's book. Hero of Alexandria demonstrated that, by adopting the rule that light rays always travel between two points by the shortest path (or, more rigorously, the extremal path), it is possible to derive the law of reflection using geometry.

---

[1] Catoptrics is the ancient Greek word for reflection.

The law of refraction was studied experimentally by Claudius Ptolemy (100-170 AD), and is reported in Book V of his Catoptrics. Ptolemy formulated a very inaccurate version of the law of refraction, which only works when the light rays are almost normally incident on the interface in question. Despite its obvious inaccuracy, Ptolemy's theory of refraction persisted for nearly 1500 years. The true law of refraction was discovered empirically by the Dutch mathematician Willebrord Snell in 1621. However, the French philosopher René Descartes was the first to publish, in his La Dioptrique (1637), the now familiar formulation of the law of refraction in terms of sines. Although there was much controversy at the time regarding plagiarism, Descartes was apparently unaware of Snell's work. Thus, in English speaking countries the law of refraction is called "Snell's law", but in French speaking countries it is called "Descartes' law".

In 1658, the French mathematician Pierre de Fermat demonstrated that all three of the laws of geometric optics can be accounted for on the assumption that light always travels between two points on the path which takes the least time (or, more rigorously, the extremal time). Fermat's ideas were an extension of those of Hero of Alexandria. Fermat's (correct) derivation of the law of refraction depended crucially on his (correct) assumption that light travels more slowly in dense media than it does in air. Unfortunately, many famous scientists, including Newton, maintained that light travels faster in dense media than it does in air. This erroneous idea held up progress in optics for over one hundred years, and was not conclusively disproved until the mid-nineteenth century. Incidentally, Fermat's principle of least time can only be justified using wave theory.

2.3 Law of Geometric Propagation

According to geometric optics, an opaque object illuminated by a point source of light casts a sharp shadow whose dimensions can be calculated using geometry. The method of calculation is very straightforward. The source emits light-rays uniformly in all directions. These rays can be represented as straight lines radiating from the source. The light-rays propagate away from the source until they encounter an opaque object, at which point they stop. This is illustrated in

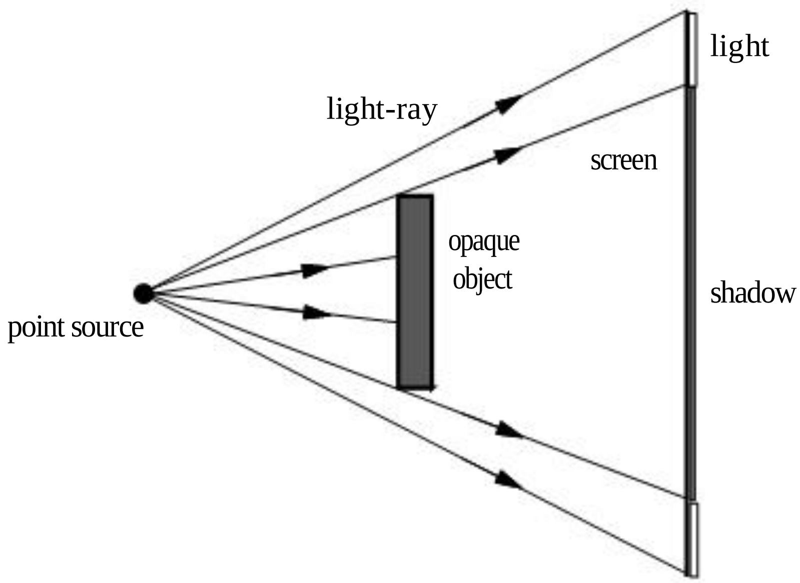


Figure 2.1: An opaque object illuminated by a point light source.

For an extended light source, each element of the source emits light-rays, just like a point source. Rays emanating from different elements of the source are assumed not to interfere with one another. Figure 2.2 shows how the shadow cast by an opaque sphere illuminated by a spherical light source is calculated using a small number of critical light-rays. The shadow consists of a perfectly black disk called the umbra, surrounded by a ring of gradually diminishing darkness called the penumbra. In the umbra, all of the light-rays emitted by the source are blocked by the opaque sphere, whereas in the penumbra only some of the rays emitted by the source are blocked by the sphere. As was well-known to the ancient Greeks, if the light-source represents the Sun, and the opaque sphere the Moon, then at a point on the Earth's surface which is situated inside the umbra the Sun is totally eclipsed, whereas at a point on the Earth's surface which is situated in the penumbra the Sun is only partially eclipsed.

In the wave picture of light, a wave-front is defined as a surface joining all adjacent points on a wave that have the same phase (e.g., all maxima, or minima, of the electric field). A light-ray is simply a line which runs perpendicular to the wave-fronts at all points along the path of the wave. This is illustrated in

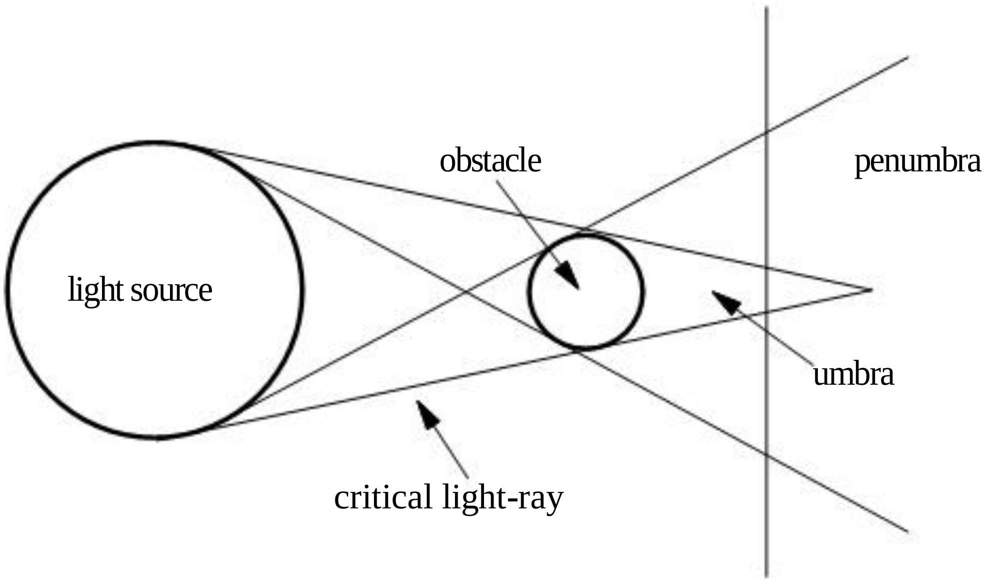


Figure 2.2: An opaque object illuminated by an extended light source.

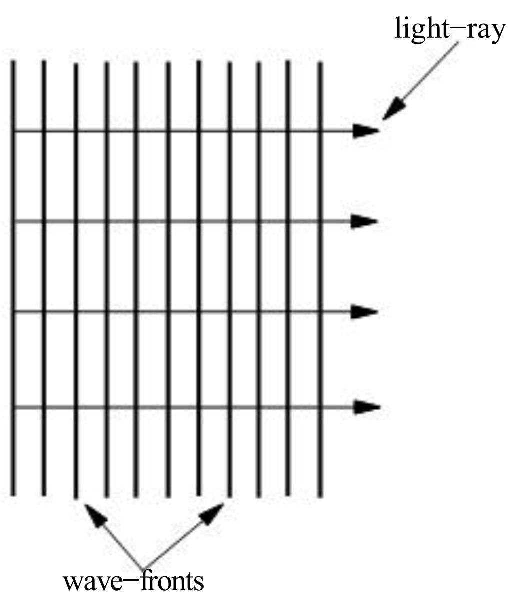


Figure 2.3: Relationship between wave-fronts and light-rays.

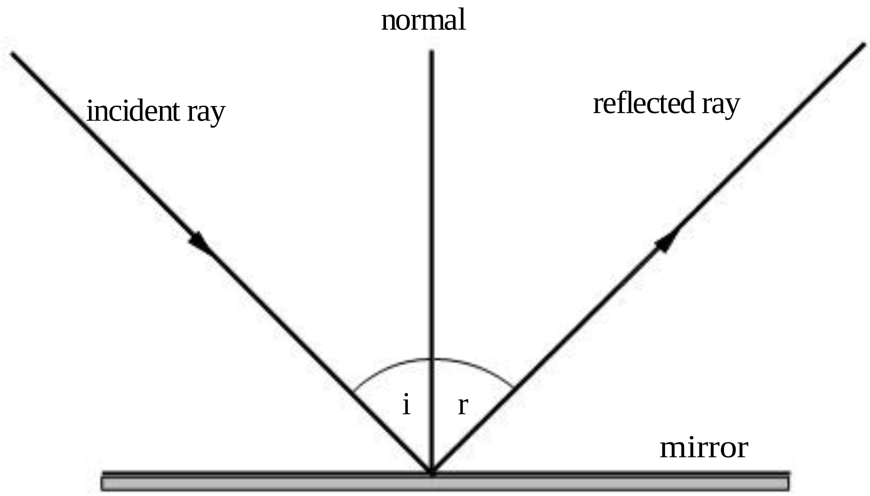


Figure 2.4: The law of reflection

Fig. 2.3. Thus, the law of rectilinear propagation of light-rays also specifies how wave-fronts propagate through homogeneous media. Of course, this law is only valid in the limit where the wavelength of the wave is much smaller than the dimensions of any obstacles which it encounters.

2.4 Law of Reflection

The law of reflection governs the reflection of light-rays off smooth conducting surfaces, such as polished metal or metal-coated glass mirrors.

Consider a light-ray incident on a plane mirror, as shown in Fig. 2.4. The law of reflection states that the incident ray, the reflected ray, and the normal to the surface of the mirror all lie in the same plane. Furthermore, the angle of reflection r is equal to the angle of incidence i. Both angles are measured with respect to the normal to the mirror.

The law of reflection also holds for non-plane mirrors, provided that the normal at any point on the mirror is understood to be the outward pointing normal to the local tangent plane of the mirror at that point. For rough surfaces, the law of reflection remains valid. It predicts that rays incident at slightly different points on the surface are reflected in completely different directions, because the normal to a rough surface varies in direction very strongly from point to point on the surface. This type of reflection is called diffuse reflection, and is what enables us to see non-shiny objects.

2.5 Law of Refraction

The law of refraction, which is generally known as Snell's law, governs the behaviour of light-rays as they propagate across a sharp interface between two transparent dielectric media.

Consider a light-ray incident on a plane interface between two transparent dielectric media, labelled 1 and 2, as shown in Fig. 2.5. The law of refraction states that the incident ray, the refracted ray, and the normal to the interface, all lie in the same plane. Furthermore,

$$n_1 \sin\theta_1 = n_2 \sin\theta_2, \tag{2.1}$$

where θ_1 is the angle subtended between the incident ray and the normal to the interface, and θ_2 is the angle subtended between the refracted ray and the normal to the interface. The quantities n_1 and n_2 are termed the refractive indices of media 1 and 2, respectively. Thus, the law of refraction predicts that a light-ray always deviates more towards the normal in the optically denser medium: i.e., the medium with the higher refractive index. Note that $n_2 > n_1$ in the figure. The law of refraction also holds for non-planar interfaces, provided that the normal to the interface at any given point is understood to be the normal to the local tangent plane of the interface at that point.

By definition, the refractive index n of a dielectric medium of dielectric constant K is given by

$$n = \sqrt{K}. \tag{2.2}$$

Table 2.1 shows the refractive indices of some common materials (for yellow light of wavelength $\lambda = 589$ nm).

The law of refraction follows directly from the fact that the speed v with which light propagates through a dielectric medium is inversely proportional to the re-

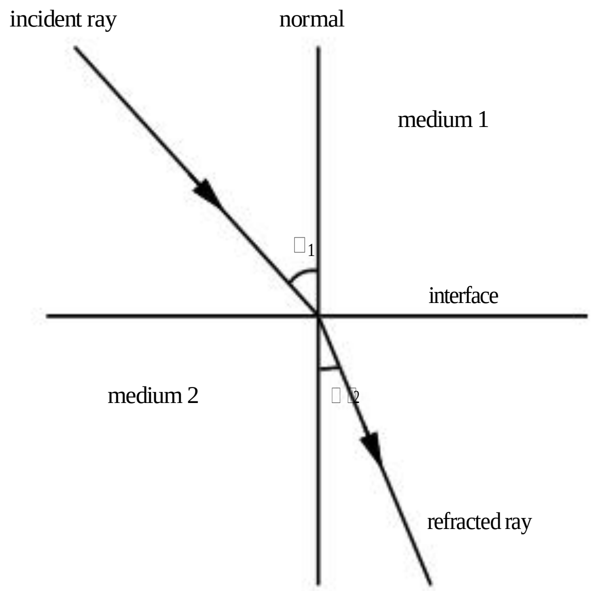


Figure 2.5: The law of refraction.

Material	n
Air (STP)	1.00029
Water	1.33
Ice	1.31
Glass:	
Light flint	1.58
Heavy flint	1.65
Heaviest flint	1.89
Diamond	2.42

Table 2.1: Refractive indices of some common materials at $\lambda = 589$ nm.

fractive index of the medium (see Sect. 1.3). In fact,

$$v = \frac{c}{n} \tag{2.3}$$

where c is the speed of light in a vacuum. Consider two parallel light-rays, a and b, incident at an angle θ_1 with respect to the normal to the interface between two dielectric media, 1 and 2. Let the refractive indices of the two media be n_1 and n_2 respectively, with $n_2 > n_1$. It is clear from Fig. 2.6 that ray b must move from point B to point Q, in medium 1, in the same time interval, Δt, in which ray a moves between points A and P, in medium 2. Now, the speed of light in medium 1 is $v_1 = c/n_1$, whereas the speed of light in medium 2 is $v_2 = c/n_2$. It follows that the length BQ is given by $v_1 \Delta t$, whereas the length AP is given by $v_2 \Delta t$. By trigonometry,

$$\sin\theta_1 = \frac{BQ}{AQ} = \frac{v_1 \Delta t}{AQ} \qquad 2.4$$

$$and \;\; \sin\theta_2 = \frac{AP}{AQ} = \frac{v_2 \Delta t}{AQ} \qquad 2.5$$

$$Hence, \qquad \frac{\sin\theta_1}{\sin\theta_2} = \frac{v_1}{v_2} = \frac{n_1}{n_2} \qquad 2.6$$

which can be rearranged to give Snell's law. Note that the lines AB and PQ represent wave-fronts in media 1 and 2, respectively, and, therefore, cross rays a and b at right-angles.

When light passes from one dielectric medium to another its velocity v changes, but its frequency f remains unchanged. Since, v = f λ for all waves, where λ is the wavelength, it follows that the wavelength of light must also change as it crosses an interface between two different media. Suppose that light propagates from medium 1 to medium 2. Let n_1 and n_2 be the refractive indices of the two media, respectively. The ratio of the wave-lengths in the two media is given by

$$\frac{\lambda_1}{\lambda_2} = \frac{v_2/f}{v_1/f} = \frac{v_2}{v_1} = \frac{n_1}{n_2} \tag{2.7}$$

$$\frac{\lambda_1}{\lambda_2} = \frac{v_2/f}{v_1/f} = \frac{v_2}{v_1} = \frac{n_1}{n_2} \tag{2.7}$$

Thus, as light moves from air to glass its wavelength decreases.

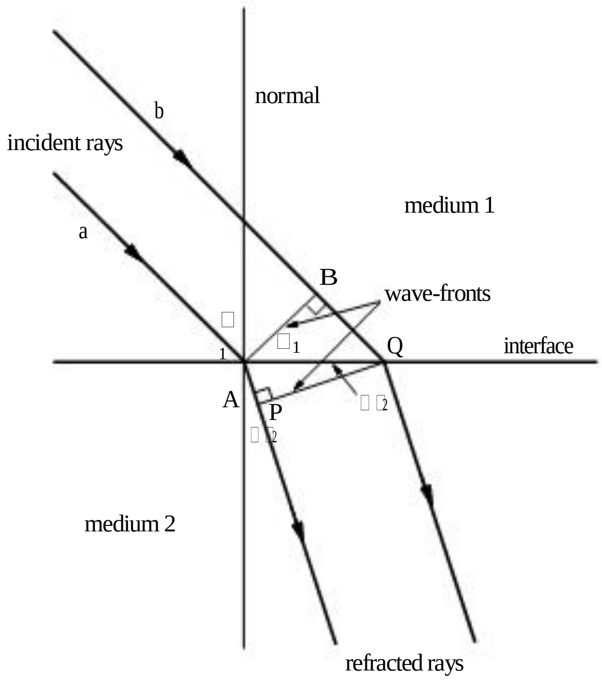


Figure 2.6: Derivation of Snell's law.

2.6 Total Internal Reflection

An interesting effect known as total internal reflection can occur when light attempts to move from a medium having a given refractive index to a medium having a lower refractive index. Suppose that light crosses an interface from medium 1 to medium 2, where $n_2 < n_1$. According to Snell's law,

$$\sin\theta_2 = \frac{n_1}{n_2}\sin\theta_1 . \tag{2.8}$$

Since $n_1/n_2 > 1$, it follows that $\theta_2 > \theta_1$. For relatively small angles of incidence, part of the light is refracted into the less optically dense medium, and part is reflected (there is always some reflection at an interface). When the angle of incidence θ_1 is such that the angle of refraction $\theta_2 = 90^\circ$, the refracted ray runs along the interface between the two media. This particular angle of incidence is called the critical angle, θ_c. For $\theta_1 > \theta_c$, there is no refracted ray. Instead, all of the light incident on the interface is reflected—see Fig. 2.7. This effect is called total internal reflection, and occurs whenever the angle of incidence exceeds the critical angle. Now when $\theta_1 = \theta_c$, we have $\theta_2 = 90^\circ$, and so $\sin\theta_2 = 1$. It follows

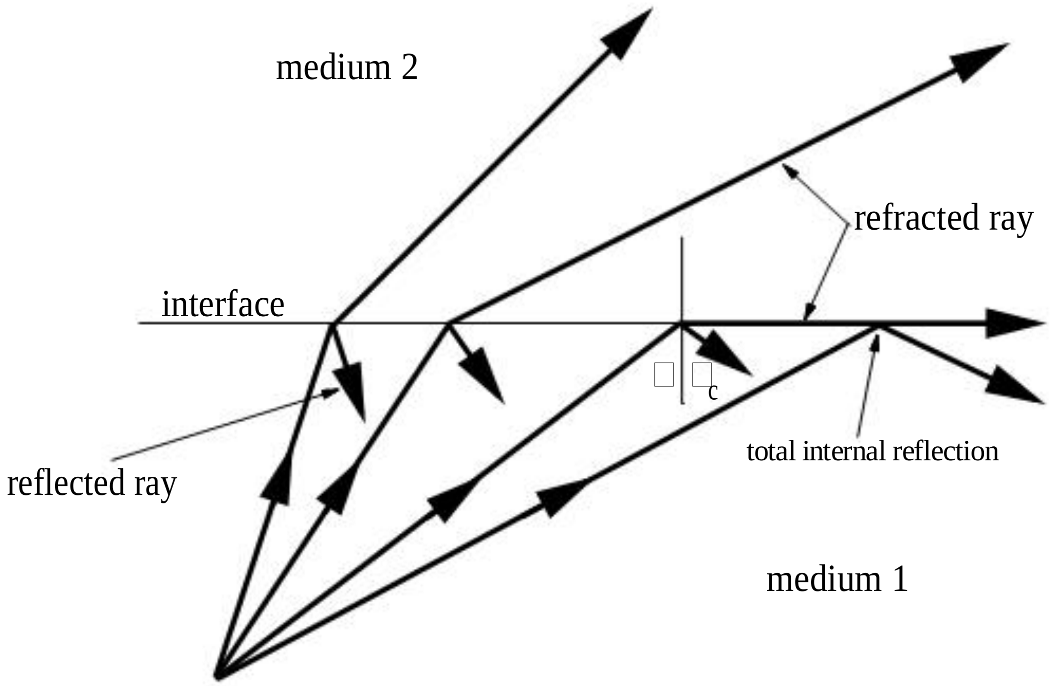


Figure 2.7: Total internal reflection.

from Eq. (2.8) that

$$\sin\theta_c = \frac{n_2}{n_1}. \tag{2.9}$$

Consider a fish (or a diver) swimming in a clear pond. As Fig. 2.8 makes clear, if the fish looks upwards it sees the sky, but if it looks at too large an angle to the vertical it sees the bottom of the pond reflected on the surface of the water. The critical angle to the vertical at which the fish first sees the reflection of the bottom of the pond is, of course, equal to the critical angle θ_c for total internal reflection at an air-water interface. From Eq. (2.9), this critical angle is given by

$$\theta_c = \sin^{-1}(1.00/1.33) = 48.8^\circ, \tag{2.10}$$

since the refractive index of air is approximately unity, and the refractive index of water is 1.33.

When total internal reflection occurs at an interface the interface in question acts as a perfect reflector. This allows 45° crown glass prisms to be used, in place of mirrors, to reflect light in binoculars. This is illustrated in Fig. 2.9. The angles

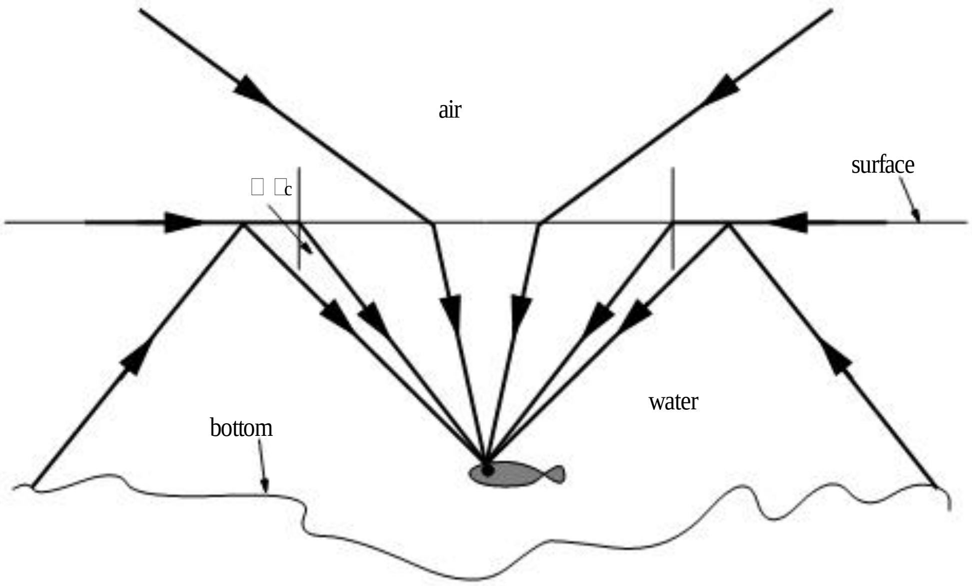


Figure 2.8: A fish's eye view.

of incidence on the sides of the prism are all 45° , which is greater than the critical angle 41° for crown glass (at an air-glass interface).

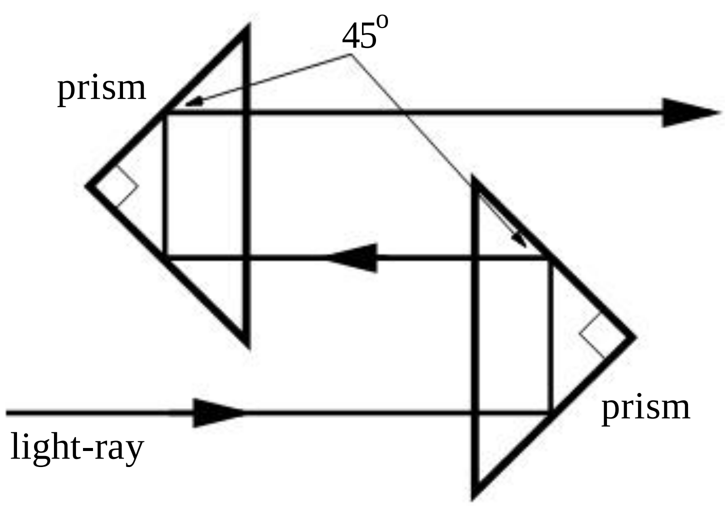


Figure 2.9: Arrangement of prisms used in binoculars.

Diamonds, for which n = 2.42, have a critical angle θ_c which is only 24° . The facets on a diamond are cut in such a manner that much of the incident light on the diamond is reflected many times by successive total internal reflections before it escapes. This effect gives rise to the characteristic sparkling of cut diamonds.

Total internal reflection enables light to be transmitted inside thin glass fibers. The light is internally reflected off the sides of the fiber, and, therefore, follows the path of the fiber. Light can actually be transmitted around corners using a glass fiber, provided that the bends in the fiber are not too sharp, so that the light always strikes the sides of the fiber at angles greater than the critical angle. The whole field of fiber optics, with its many useful applications, is based on this effect.

2.7 Dispersion

When a wave is refracted into a dielectric medium whose refractive index varies with wavelength then the angle of refraction also varies with wavelength. If the incident wave is not monochromatic, but is, instead, composed of a mixture of waves of different wavelengths, then each component wave is refracted through a different angle. This phenomenon is called dispersion.

Figure 2.10 shows the refractive indices of some common materials as functions of wavelength in the visible range. It can be seen that the refractive index always decreases with increasing wavelength in the visible range. In other words, violet light is always refracted more strongly than red light.

Suppose that a parallel-sided glass slab is placed in a beam of white light. Dispersion takes place inside the slab, but, since the rays which emerge from the slab all run parallel to one another, the dispersed colours recombine to form white light again, and no dispersion is observed except at the very edges of the beam. This is illustrated in Fig. 2.11. It follows that the dispersion of white light through a parallel-sided glass slab is not generally a noticeable effect.

Suppose that a glass prism is placed in a beam of white light. Dispersion takes place inside the prism, and, since the emerging rays are not parallel for different colours, the dispersion is clearly noticeable, especially if the emerging rays are projected onto a screen which is placed a long way from the prism. This is illustrated in Fig. 2.12. It is clear that a glass prism is far more effective at separating white light into its component colours than a parallel-sided glass slab

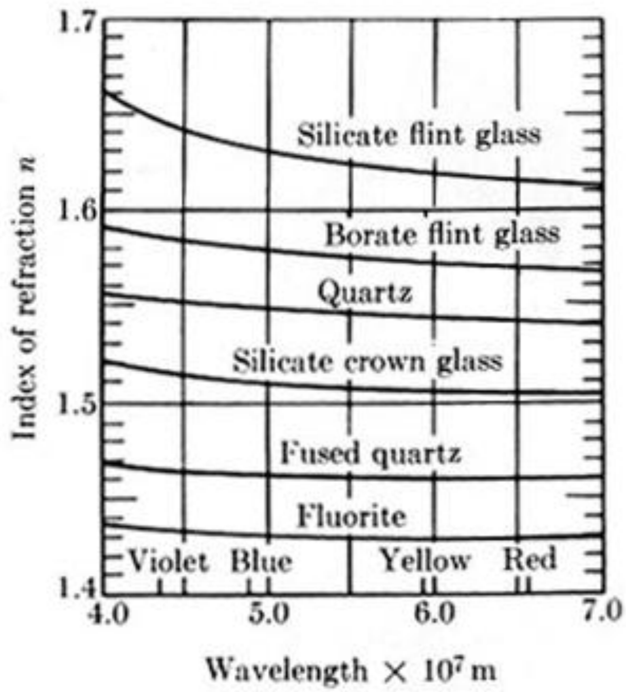


Figure 2.10: Refractive indices of some common materials as functions of wavelength.

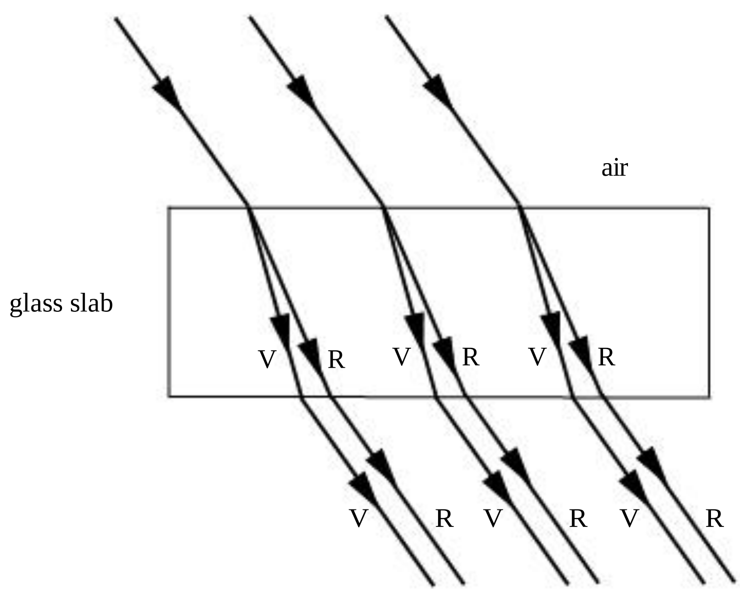


Figure 2.11: Dispersion of light by a parallel-sided glass slab.

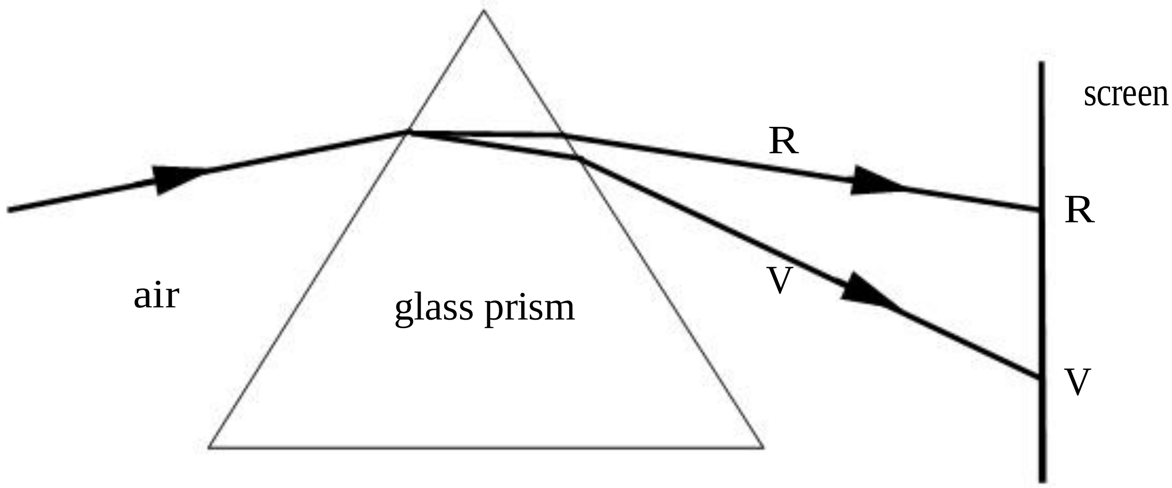


Figure 2.12: Dispersion of light by a glass prism.

(which explains why prisms are generally employed to perform this task).

2.8 Rainbows

The most well-known, naturally occurring phenomenon which involves the dispersion of light is a rainbow. A rainbow is an arc of light, with an angular radius of 42°, centred on a direction which is opposite to that of the Sun in the sky (i.e., it is centred on the direction of propagation of the Sun's rays)—see Fig. 2.13. Thus, if the Sun is low in the sky (i.e., close to the horizon) we see almost a full semi-circle. If the Sun is higher in the sky we see a smaller arc, and if the Sun is more than 42° above the horizon then there is no rainbow (for viewers on the Earth's surface). Observers on a hill may see parts of the rainbow below the horizontal: i.e., an arc greater than a semi-circle. Passengers on an airplane can sometimes see a full circle.

The colours of a rainbow vary smoothly from red on the outside of the arc to violet on the inside. A rainbow has a diffuse inner edge, and a sharp outer edge. Sometimes a secondary arc is observed. This is fainter and larger (with an angular radius of 50°) than the primary arc, and the order of the colours is reversed (i.e., red is on the inside, and violet on the outside). The secondary arc has a diffuse outer edge, and a sharp inner edge. The sky between the two arcs sometimes

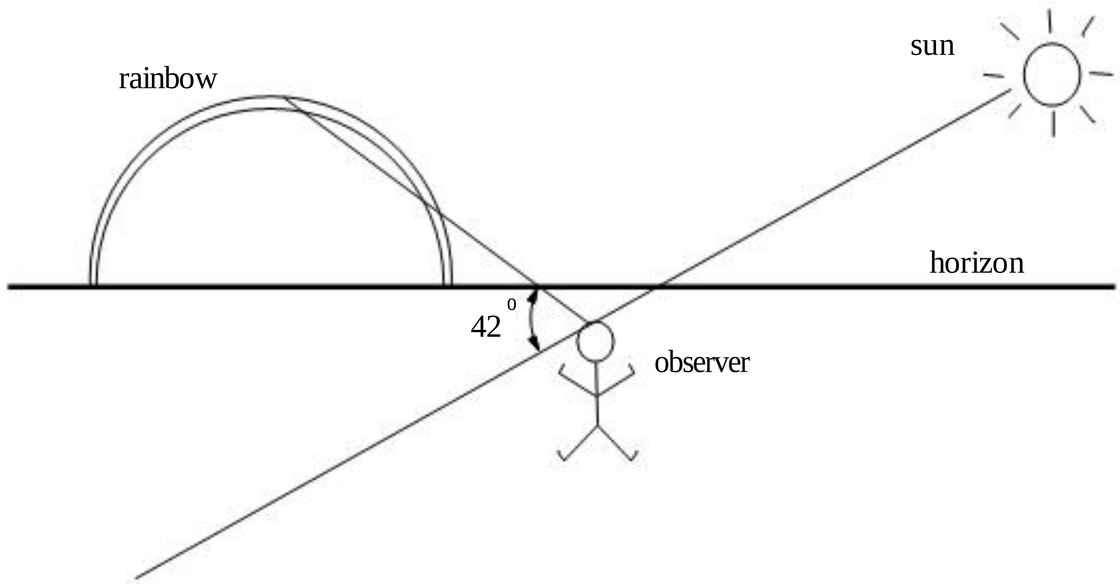


Figure 2.13: A rainbow.

appears to be less bright than the sky elsewhere. This region is called Alexander's dark band, in honour of Alexander of Aphrodisias who described it some 1800 years ago.

Rainbows have been studied since ancient times. Aristotle wrote extensively on rainbows in his De Meteorologica,[2] and even speculated that a rainbow is caused by the reflection of sunlight from the drops of water in a cloud.

The first scientific study of rainbows was performed by Theodoric, professor of theology at Freiburg, in the fourteenth century. He studied the path of a light-ray through a spherical globe of water in his laboratory, and suggested that the globe be thought of as a model of a single falling raindrop. A ray, from the Sun, entering the drop, is refracted at the air-water interface, undergoes internal reflection from the inside surface of the drop, and then leaves the drop in a backward direction, after being again refracted at the surface. Thus, looking away from the Sun, towards a cloud of raindrops, one sees an enhancement of light due to these rays. Theodoric did not explain why this enhancement is concentrated at a particular angle from the direction of the Sun's rays, or why the light is split into different colours.

[2] "On Weather".

The first person to give a full explanation of how a rainbow is formed was René Descartes. He showed mathematically that if one traces the path through a spherical raindrop of parallel light-rays entering the drop at different points on its surface, each emerges in a different direction, but there is a concentration of emerging rays at an angle of 42° from the reverse direction to the incident rays, in exact agreement with the observed angular size of rainbows. Furthermore, since some colours are refracted more than others in a raindrop, the "rainbow angle" is slightly different for each colour, so a raindrop disperses the Sun's light into a set of nearly overlapping coloured arcs.

Figure 2.13 illustrated Descartes' theory in more detail. It shows parallel light-rays entering a spherical raindrop. Only rays entering the upper half contribute to the rainbow effect. Let us follow the rays, one by one, from the top down to the middle of the drop. We observe the following pattern. Rays which enter near the top of the drop emerge going in almost the reverse direction, but a few degrees below the horizontal. Rays entering a little further below the top emerge at a greater angle below the horizontal. Eventually, we reach a critical ray, called the rainbow ray, which emerges in an angle 42° below the horizontal. Rays entering the drop lower than the rainbow ray emerge at an angle less than 42°. Thus, the rainbow ray is the one which deviates most from the reverse direction to the incident rays. This variation, with 42° being the maximum angle of deviation from the reverse direction, leads to a bunching of rays at that angle, and, hence, to an unusually bright arc of reflected light centred around 42° from the reverse direction. The arc has a sharp outer edge, since reflected light cannot deviate by more than 42° from the reverse direction, and a diffuse inner edge, since light can deviate by less than 42° from the reverse direction: 42° is just
the most likely angle of deviation. Finally, since the rainbow angle varies slightly with wavelength (because the refractive index of water varies slightly with wavelength), the arcs corresponding to each colour appear at slightly different angles relative to the reverse direction to the incident rays. We expect violet light to be refracted more strongly than red light in a raindrop. It is, therefore, clear, from Fig. 2.14, that the red arc deviates slightly more from the reverse direction to the incident rays than the violet arc. In other words, violet is concentrated on the inside of the rainbow, and red is concentrated on the outside.

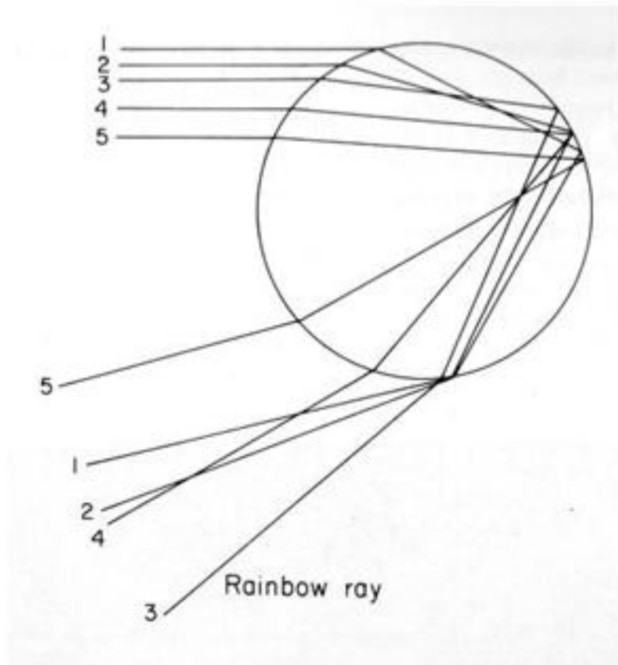


Figure 2.14: Descarte's theory of the rainbow.

Descartes was also able to show that light-rays which are internally reflected twice inside a raindrop emerge concentrated at an angle of 50° from the reverse direction to the incident rays. Of course, this angle corresponds exactly to the angular size of the secondary rainbow sometimes seen outside the first. This rainbow is naturally less intense than the primary rainbow, since a light-ray loses some of its intensity at each reflection or refraction event. Note that 50° represents the angle of maximum deviation of doubly reflected light from the reverse direction (i.e., doubly reflected light can deviate by more than this angle, but not by less). Thus, we expect the secondary rainbow to have a diffuse outer edge, and a sharp inner edge. We also expect doubly reflected violet light to be refracted more strongly in a raindrop than doubly reflected red light. It follows, from Fig. 2.15, that the red secondary arc deviates slightly less from the reverse direction to the incident rays than the violet secondary arc. In other words, red is concentrated on the inside of the secondary rainbow, and violet on the outside. Since no reflected light emerges between the primary and secondary rainbows (i.e., in the angular range 42° to 50°, relative to the reverse direction), we naturally expect this region of the sky to look slightly less bright than the other

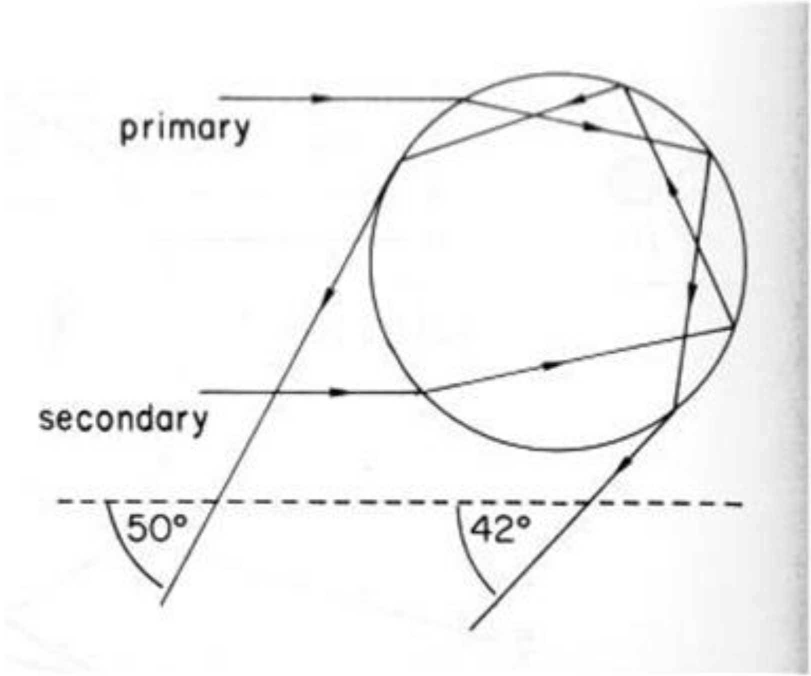


Figure 2.15: Rainbow rays for the primary and secondary arcs of a rainbow.

surrounding regions of the sky, which explains Alexander's dark band.

2.9 Worked Examples

Example 2.1: The corner-cube reflector

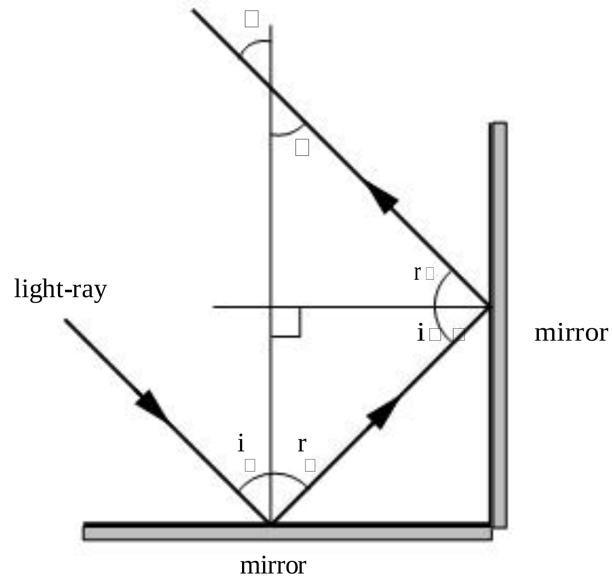
Question: Two mirrors are placed at right-angles to one another. Show that a light-ray incident from any direction in the plane perpendicular to both mirrors is reflected through 180°.

Answer: Consider the diagram. We are effectively being asked to prove that $\alpha = i_1$, for any value of i_1. Now, from trigonometry,

$$i_2 = 90^\circ - r_1 .$$

But, from the law of reflection, $r_1 = i_1$ and $i_2 = r_2$, so

$$r_2 = 90^\circ - i_1 .$$



Trigonometry also yields

$$\alpha = 90^\circ - r_2 .$$

It follows from the previous two equations that

$$\alpha = 90^\circ - (90^\circ - i_1) = i_1 .$$

Hence, $\alpha = i_1$, for all values of i_1.

It can easily be appreciated that a combination of three mutually perpendicular mirrors would reflect a light-ray incident from any direction through 180°. Such a combination of mirrors is called a corner-cube reflector. Astronauts on the Apollo 11 mission (1969) left a panel of corner-cube reflectors on the surface of the Moon. These reflectors have been used ever since to measure the Earth-Moon distance via laser range finding (basically, a laser beam is fired from the Earth, reflects off the corner-cube reflectors on the Moon, and then returns to the Earth. The time of travel of the beam can easily be converted into the Earth-Moon distance). The Earth-Moon distance can be measured to within an accuracy of 3 cm using this method.

Example 2,2: Refraction

Question: A light-ray of wavelength $\lambda_1 = 589$ nm traveling through air is incident on a smooth, flat slab of crown glass (refractive index 1.52) at an angle of $\theta_1 = 30.0^\circ$ to the normal. What is the angle of refraction? What is the wavelength λ_2 of the light inside the glass? What is the frequency f of the light inside the glass?

Answer: Snell's law can be written

$$\sin\theta_2 = \frac{n_1}{n_2}\sin\theta_1 .$$

In this case, $\theta_1 = 30^\circ$, $n_1 \simeq 1.00$ (here, we neglect the slight deviation of the refractive index of air from that of a vacuum), and $n_2 = 1.52$. Thus,

$$\sin\theta_2 = \frac{(1.00)}{(1.52)}(0.5) = 0.329,$$

giving

$$\theta_2 = 19.2^\circ$$

as the angle of refraction (measured with respect to the normal).

The wavelength λ_2 of the light inside the glass is given by

$$\lambda_2 = \frac{n_1}{n_2}\lambda_1 = \frac{(1.00)}{(1.52)}(589) = 387.5 nm$$

The frequency f of the light inside the glass is exactly the same as the frequency outside the glass, and is given by

$$f = \frac{c}{n_1 \lambda_1} = \frac{(3\times 10^8)}{(1.00)(589\times 10^{-9})} = 5.09\times 10^{14}\text{ Hz}.$$

3 Paraxial Optics

3.1 Spherical Mirrors

A spherical mirror is a mirror which has the shape of a piece cut out of a spherical surface. There are two types of spherical mirrors: concave, and convex. These are illustrated in Fig. 3.1. The most commonly occurring examples of concave mirrors are shaving mirrors and makeup mirrors. As is well-known, these types of mirrors magnify objects placed close to them. The most commonly occurring examples of convex mirrors are the passenger-side wing mirrors of cars. These type of mirrors have wider fields of view than equivalent flat mirrors, but objects which appear in them generally look smaller (and, therefore, farther away) than they actually are.

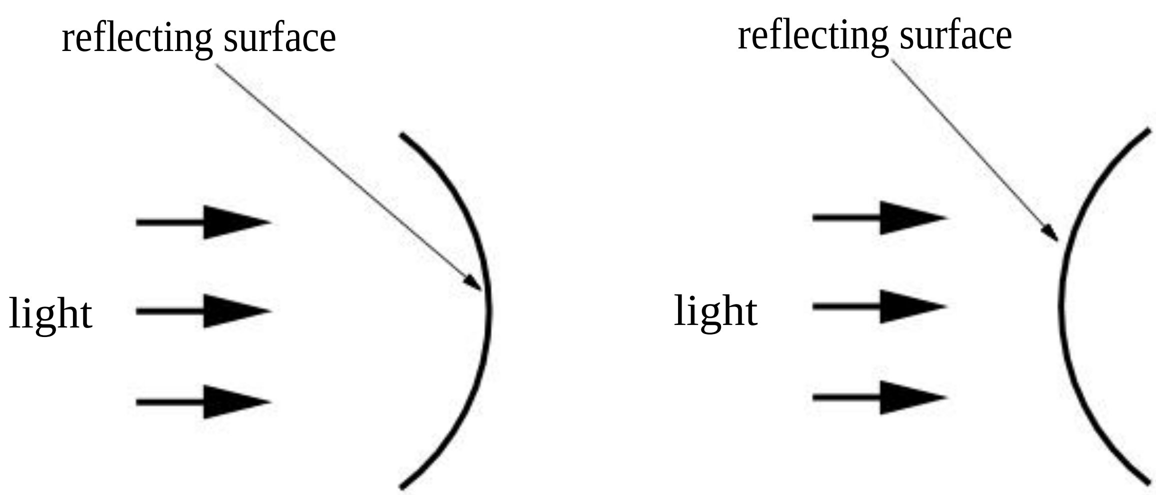


Figure 3.1: A concave (left) and a convex (right) mirror

Let us now introduce a few key concepts which are needed to study image formation by a concave spherical mirror. As illustrated in Fig. 3.2, the normal to the centre of the mirror is called the principal axis. The mirror is assumed to be rotationally symmetric about this axis. Hence, we can represent a three-dimensional mirror in a two-dimensional diagram, without loss of generality. The point V at which the principal axis touches the surface of the mirror is called the vertex. The point C, on the principal axis, which is equidistant from all points on the reflecting surface of the mirror is called the centre of curvature. The distance along the principal axis from point C to point V is called the radius of curvature of the mirror, and is denoted R. It is found experimentally that rays striking a concave mirror parallel to its principal axis, and not too far away from this axis, are reflected by the mirror such that they all pass through the same point F on the principal axis. This point, which is lies between the centre of curvature and the vertex, is called the focal point, or focus, of the mirror. The distance along the principal axis from the focus to the vertex is called the focal length of the mirror, and is denoted f.

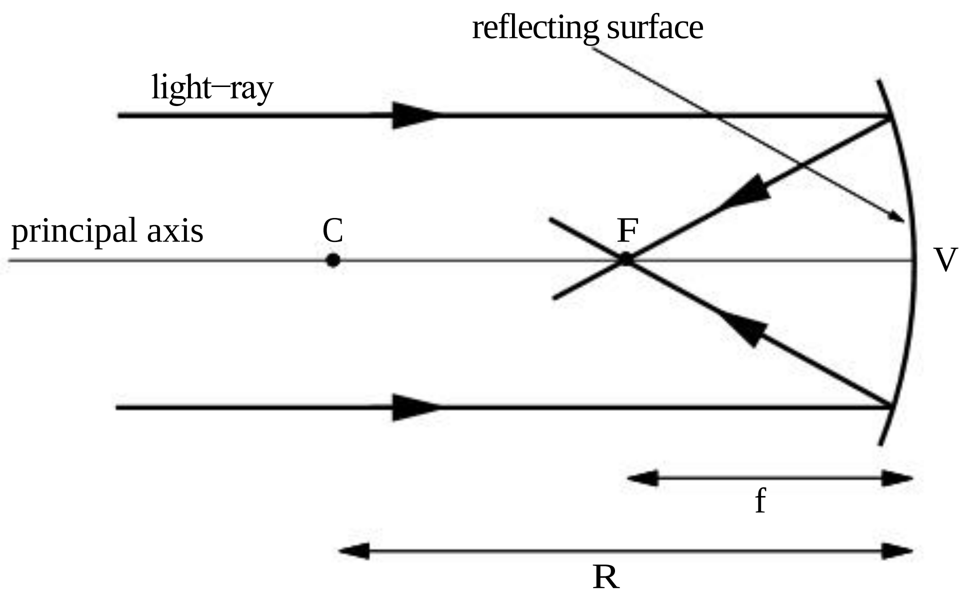


Figure 3.2: Image formation by a concave mirror.

In our study of concave mirrors, we are going to assume that all light-rays which strike a mirror parallel to its principal axis (e.g., all rays emanating from a distant object) are brought to a focus at the same point F. Of course, as mentioned above, this is only an approximation. It turns out that as rays from a distant object depart further from the principal axis of a concave mirror they are brought to a focus ever closer to the mirror, as shown in Fig. 3.3. This lack of perfect focusing of a spherical mirror is called spherical aberration. The approximation in which we neglect spherical aberration is called the paraxial approximation.[3]

[3] "Paraxial" is derived from ancient Greek roots, and means "close to the axis".

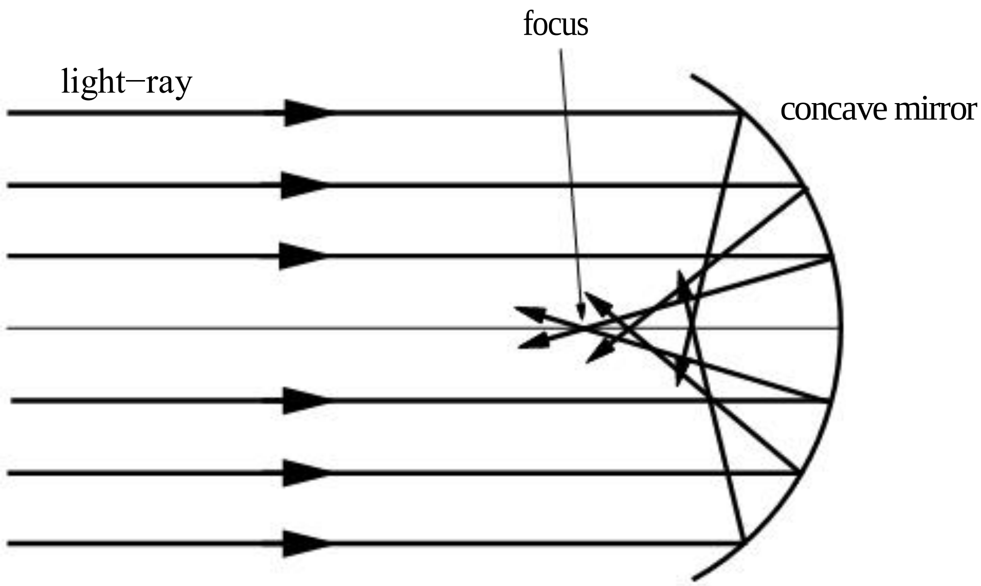


Figure 3.3: Spherical aberration in a concave mirror.

Likewise, the study of image formation under this approximation is known as paraxial optics. This field of optics was first investigated systematically by the famous German mathematician Karl Friedrich Gauss in 1841.

It can be demonstrated, by geometry, that the only type of mirror which does not suffer from spherical aberration is a parabolic mirror (i.e., a mirror whose reflecting surface is the surface of revolution of a parabola). Thus, a ray traveling parallel to the principal axis of a parabolic mirror is brought to a focus at the same point F, no matter how far the ray is from the axis. Since the path of a light-ray is completely reversible, it follows that a light source placed at the focus F of a parabolic mirror yields a perfectly parallel beam of light, after the light has reflected off the surface of the mirror. Parabolic mirrors are more difficult, and, therefore, more expensive, to make than spherical mirrors. Thus, parabolic mirrors are only used in situations where the spherical aberration of a conventional spherical mirror would be a serious problem. The receiving dishes of radio telescopes are generally parabolic. They reflect the incoming radio waves from (very) distant astronomical sources, and bring them to a focus at a single point, where a detector is placed. In this case, since the sources are extremely faint, it is imperative to avoid the signal losses which would be associated with spherical aberration. A car headlight consists of a light-bulb placed at the focus of a parabolic reflector. The use of a parabolic reflector enables the headlight to cast a very straight beam of light ahead of the car. The beam would be nowhere near as well-focused were a spherical reflector used instead.

3.2 Image Formation by Concave Mirrors

There are two alternative methods of locating the image formed by a concave mirror. The first is purely graphical, and the second uses simple algebraic analysis.

The graphical method of locating the image produced by a concave mirror consists of drawing light-rays emanating from key points on the object, and finding where these rays are brought to a focus by the mirror. This task can be accomplished using just four simple rules:

1. An incident ray which is parallel to the principal axis is reflected through the focus F of the mirror.
2. An incident ray which passes through the focus F of the mirror is reflected parallel to the principal axis.
3. An incident ray which passes through the centre of curvature C of the mirror is reflected back along its own path (since it is normally incident on the mirror).
4. An incident ray which strikes the mirror at its vertex V is reflected such that its angle of incidence with respect to the principal axis is equal to its angle of reflection.

The validity of these rules in the paraxial approximation is fairly self-evident.

Consider an object ST which is placed a distance p from a concave spherical mirror, as shown in Fig. 3.4. For the sake of definiteness, let us suppose that the object distance p is greater than the focal length f of the mirror. Each point on the object is assumed to radiate light-rays in all directions. Consider four light-rays emanating from the tip T of the object which strike the mirror, as shown

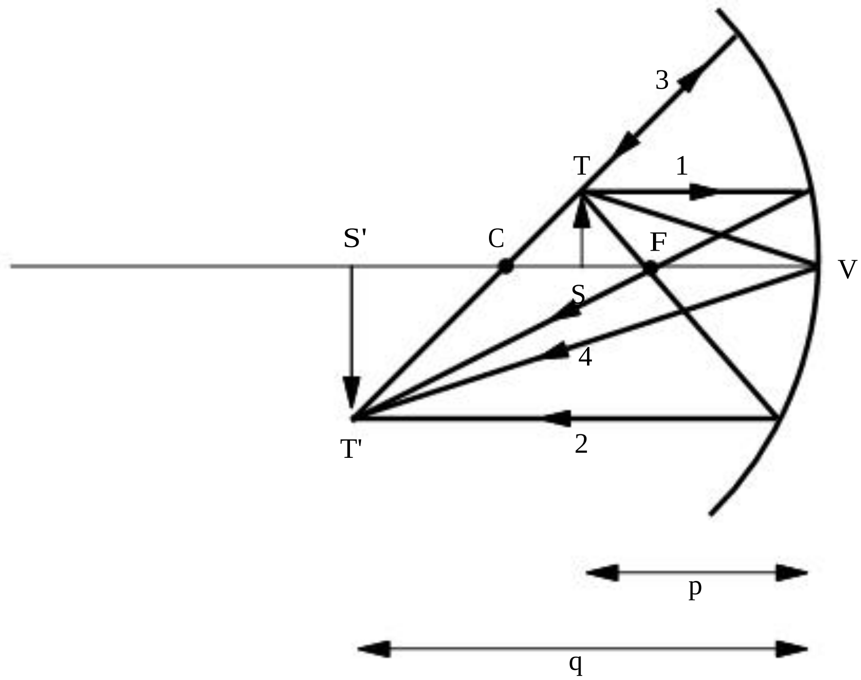


Figure 3.4: Formation of a real image by a concave mirror.

in the figure. The reflected rays are constructed using rules 1-4 above, and the rays are labelled accordingly. It can be seen that the reflected rays all come together at some point T'. Thus, T' is the image of T (i.e., if we were to place a small projection screen at T' then we would see an image of the tip on the screen). As is easily demonstrated, rays emanating from other parts of the object are brought into focus in the vicinity of T' such that a complete image of the object is produced between S' and T' (obviously, point S' is the image of point S). This image could be viewed by projecting it onto a screen placed between points S' and T'. Such an image is termed a real image. Note that the image $S'T'$ would also be directly visible to an observer looking straight at the mirror from a distance greater than the image distance q (since the observer's eyes could not tell that the light-rays diverging from the image were in anyway different from those which would emanate from a real object). According to the figure, the image is inverted with respect to the object, and is also magnified.

Figure 3.5 shows what happens when the object distance p is less than the focal length f. In this case, the image appears to an observer looking straight at the mirror to be located behind the mirror. For instance, rays emanating from

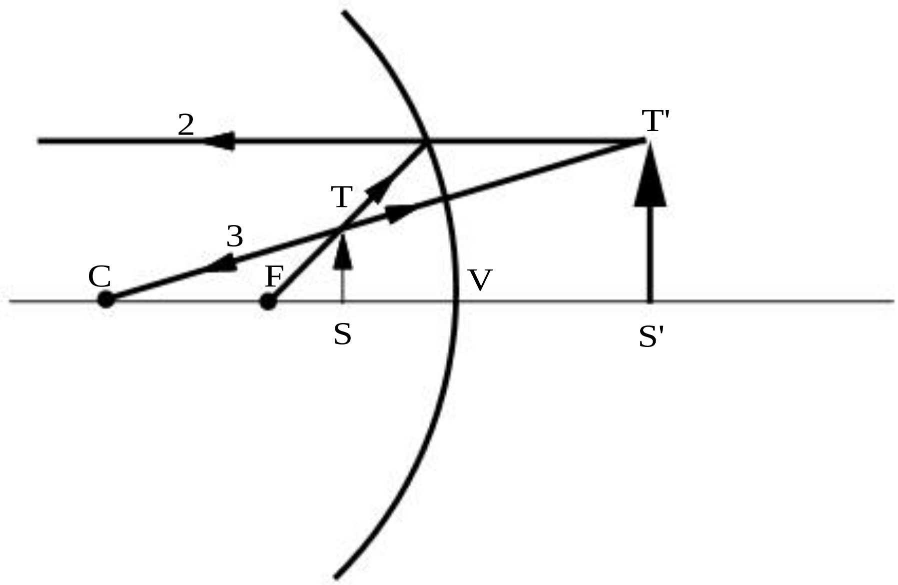


Figure 3.5: Formation of a virtual image by a concave mirror.

the tip T of the object appear, after reflection from the mirror, to come from a point T' which is behind the mirror. Note that only two rays are used to locate T', for the sake of clarity. In fact, two is the minimum number of rays needed to locate a point image. Of course, the image behind the mirror cannot be viewed by projecting it onto a screen, because there are no real light-rays behind the mirror. This type of image is termed a virtual image. The characteristic difference between a real image and a virtual image is that, immediately after reflection from the mirror, light-rays emitted by the object converge on a real image, but diverge from a virtual image. According to Fig. 3.5, the image is upright with respect to the object, and is also magnified.

The graphical method described above is fine for developing an intuitive understanding of image formation by concave mirrors, or for checking a calculation, but is a bit too cumbersome for everyday use. The analytic method described below is far more flexible.

Consider an object ST placed a distance p in front of a concave mirror of radius of curvature R. In order to find the image $S'T'$ produced by the mirror, we draw two rays from T to the mirror—see Fig. 3.6. The first, labelled 1, travels from T to the vertex V and is reflected such that its angle of incidence θ equals its angle

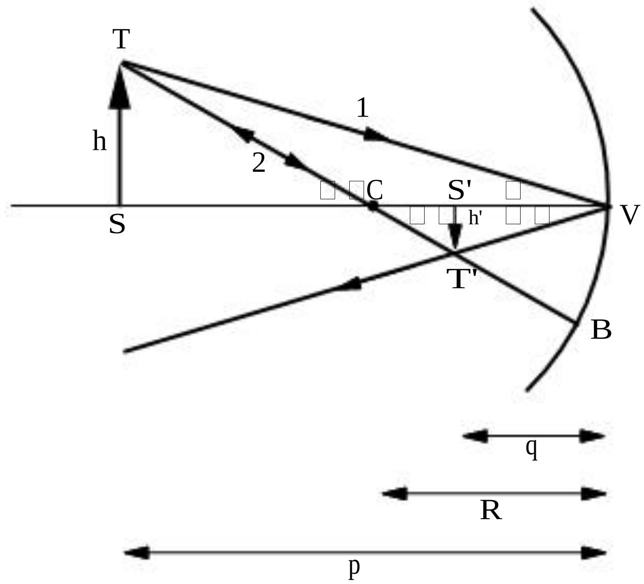


Figure 3.6: Image formation by a concave mirror.

of reflection. The second ray, labelled 2, passes through the centre of curvature C of the mirror, strikes the mirror at point B, and is reflected back along its own path. The two rays meet at point T'. Thus, $S'T'$ is the image of ST, since point S' must lie on the principal axis.

In the triangle STV, we have $\tan\theta = h/p$, and in the triangle $S'T'V$ we have $\tan\theta = -h'/q$, where p is the object distance, and q is the image distance. Here, h is the height of the object, and h' is the height of the image. By convention, h' is a negative number, since the image is inverted (if the image were upright then h' would be a positive number). It follows that

$$\tan\theta = \frac{h}{p} = \frac{-h'}{q} \qquad (3.1)$$

Thus, the magnification M of the image with respect to the object is given by

$$M = \frac{h'}{h} = -\frac{q}{p} \qquad (3.2)$$

By convention, M is negative if the image is inverted with respect to the object, and positive if the image is upright. It is clear that the magnification of the image is just determined by the ratio of the image and object distances from the vertex.

From triangles STC and $S'T'C$, we have $\tan\alpha = h/(p-R)$ and $\tan\alpha = -h'/(R-q)$, respectively. These expressions yield

$$\tan\alpha = \frac{h}{p-R} = \frac{h'}{R-q} \qquad (3.3)$$

Equations (3.2) and (3.3) can be combined to give

$$\frac{-h'}{h} = \frac{R-q}{p-R} = \frac{q}{p}, \qquad (3.4)$$

which easily reduces to

This expression relates the object distance, the image distance, and the radius of curvature of the mirror.

For an object which is very far away from the mirror (i.e., $p \to \infty$), so that light-rays from the object are parallel to the principal axis, we expect the image to form at the focal point F of the mirror. Thus, in this case, $q = f$, where f is the focal length of the mirror, and Eq. (3.5) reduces to

$$0 + \frac{1}{f} = \frac{2}{R} \qquad (3.6)$$

The above expression yields

$$f = \frac{R}{2} \qquad (3.7)$$

In other words, in the paraxial approximation, the focal length of a concave spherical mirror is half of its radius of curvature. Equations (3.5) and (3.7) can be combined to give

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{f} \qquad (3.8)$$

The above expression was derived for the case of a real image. However, as is easily demonstrated, it also applies to virtual images provided that the following sign convention is adopted. For real images, which always form in front of the mirror, the image distance q is positive. For virtual images, which always form

Position of object	Position of image	Character of image
At ∞	At F	Real, zero size
Between ∞ and C	Between F and C	Real, inverted, diminished
At C	At C	Real, inverted, same size
Between C and F	Between C and ∞	Real, inverted, magnified
At F	At ∞	
Between F and V	From $-\infty$ to V	Virtual, upright, magnified
At V	At V	Virtual, upright, same size

Table 3.1: Rules for image formation by concave mirrors.

behind the mirror, the image distance q is negative. It immediately follows, from Eq. (3.2), that real images are always inverted, and virtual images are always upright. Table 3.1 shows how the location and character of the image formed in a concave spherical mirror depend on the location of the object, according to Eqs. (3.2) and (3.8). It is clear that the modus operandi of a shaving mirror, or a makeup mirror, is to place the object (i.e., a face) between the mirror and the focus of the mirror. The image is upright, (apparently) located behind the mirror, and magnified.

3.3 Image Formation by Convex Mirrors

The definitions of the principal axis, centre of curvature C, radius of curvature R, and the vertex V , of a convex mirror are analogous to the corresponding definitions for a concave mirror. When parallel light-rays strike a convex mirror they are reflected such that they appear to emanate from a single point F located behind the mirror, as shown in Fig. 3.7. This point is called the virtual focus of the mirror. The focal length f of the mirror is simply the distance between V and F. As is easily demonstrated, in the paraxial approximation, the focal length of a convex mirror is half of its radius of curvature.

There are, again, two alternative methods of locating the image formed by a convex mirror. The first is graphical, and the second analytical.

According to the graphical method, the image produced by a convex mirror can always be located by drawing a ray diagram according to four simple rules:

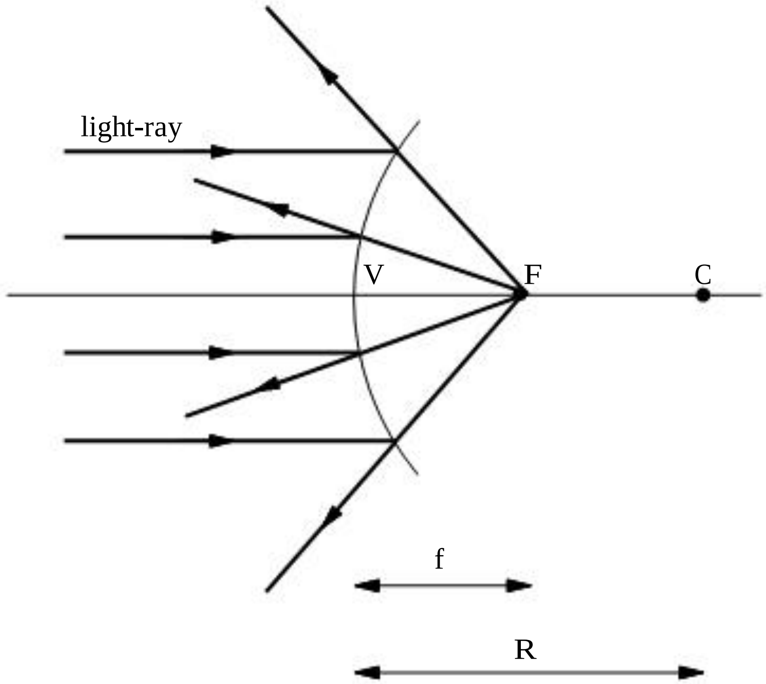


Figure 3.7: The virtual focus of a convex mirror.

1. An incident ray which is parallel to the principal axis is reflected as if it came from the virtual focus F of the mirror.

2. An incident ray which is directed towards the virtual focus F of the mirror is reflected parallel to the principal axis.

3. An incident ray which is directed towards the centre of curvature C of the mirror is reflected back along its own path (since it is normally incident on the mirror).

4. An incident ray which strikes the mirror at its vertex V is reflected such that its angle of incidence with respect to the principal axis is equal to its angle of reflection.

The validity of these rules in the paraxial approximation is, again, fairly selfevident.

In the example shown in Fig. 3.8, two rays are used to locate the image $S'T'$ of an object ST placed in front of the mirror. It can be seen that the image is virtual, upright, and diminished.

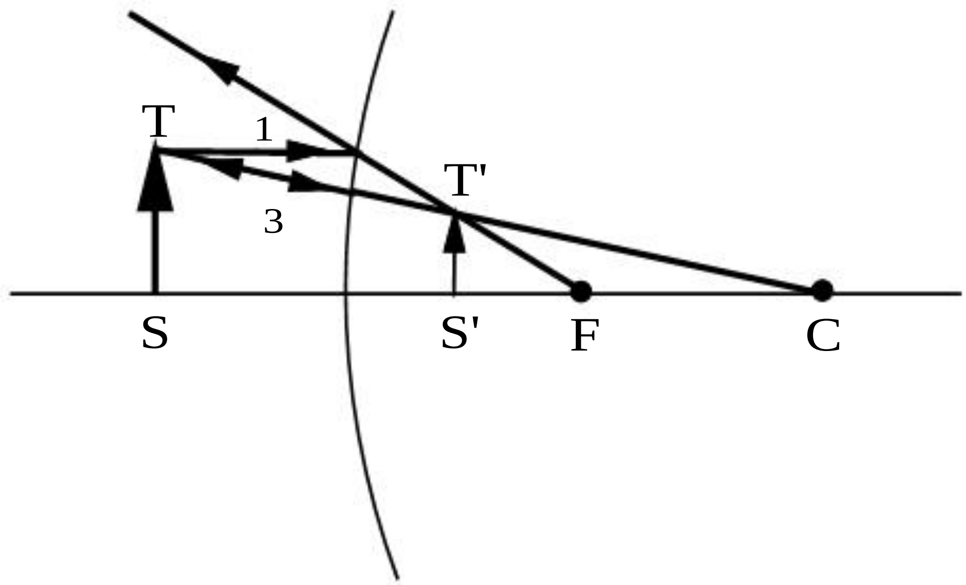


Figure 3.8: Image formation by a convex mirror.

Position of object	Position of image	Character of image
At ∞	At F	Virtual, zero size
Between ∞ and V	Between F and V	Virtual, upright, diminished
At V	At V	Virtual, upright, same size

Table 3.2: Rules for image formation by convex mirrors.

As is easily demonstrated, application of the analytical method to image formation by convex mirrors again yields Eq. (3.2) for the magnification of the image, and Eq. (3.8) for the location of the image, provided that we adopt the following sign convention. The focal length f of a convex mirror is redefined to be minus the distance between points V and F. In other words, the focal length of a concave mirror, with a real focus, is always positive, and the focal length of a convex mirror, with a virtual focus, is always negative. Table 3.2 shows how the location and character of the image formed in a convex spherical mirror depend on the location of the object, according to Eqs. (3.2) and (3.8) (with $f < 0$).

In summary, the formation of an image by a spherical mirror involves the crossing of light-rays emitted by the object and reflected off the mirror. If the light-rays actually cross in front of the mirror then the image is real. If the light-rays do not actually cross, but appear to cross when projected backwards behind the mirror, then the image is virtual. A real image can be projected onto a screen, a virtual image cannot. However, both types of images can be seen by an observer, and photographed by a camera. The magnification of the image is specified by Eq. (3.2), and the location of the image is determined by Eq. (3.8). These two formulae can be used to characterize both real and virtual images formed by either concave or convex mirrors, provided that the following sign conventions are observed:

1. The height h' of the image is positive if the image is upright, with respect to the object, and negative if the image is inverted.
2. The magnification M of the image is positive if the image is upright, with respect to the object, and negative if the image is inverted.
3. The image distance q is positive if the image is real, and, therefore, located in front of the mirror, and negative if the image is virtual, and, therefore, located behind the mirror.
4. The focal length f of the mirror is positive if the mirror is concave, so that the focal point F is located in front of the mirror, and negative if the mirror is convex, so that the focal point F is located behind the mirror.

Note that the front side of the mirror is defined to be the side from which the light is incident.

3.4 Image Formation by Plane Mirrors

Both concave and convex spherical mirrors asymptote to plane mirrors in the limit in which their radii of curvature R tend to infinity. In other words, a plane mirror can be treated as either a concave or a convex mirror for which $R \to \infty$. Now, if $R \to \infty$, then $f = \pm R/2 \to \infty$, so $1/f \to 0$, and Eq. (3.8) yields

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{f} = 0 \tag{3.9}$$

or

$$q = -p. \tag{3.10}$$

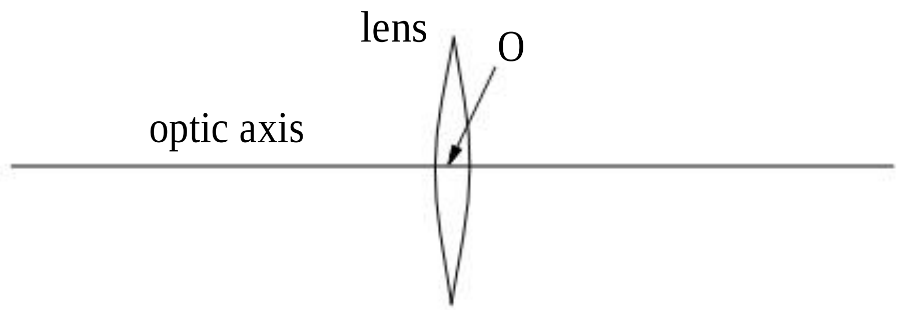


Figure 3.9: The optic axis of a lens.

Thus, for a plane mirror the image is virtual, and is located as far behind the mirror as the object is in front of the mirror. According to Eq. (3.2), the magnification of the image is given by

$$M = \frac{-q}{p} = 1 \tag{3.11}$$

Clearly, the image is upright, since M > 0, and is the same size as the object, since |M| = 1. However, an image seen in a plane mirror does differ from the original object in one important respect: i.e., left and right are swapped over. In other words, a right-hand looks like a left-hand in a plane mirror, and vice versa.

3.5 Thin Lenses

A lens is a transparent medium (usually glass) bounded by two curved surfaces (generally either spherical, cylindrical, or plane surfaces). As illustrated in Fig. 3.9, the line which passes normally through both bounding surfaces of a lens is called the optic axis. The point O on the optic axis which lies midway between the two bounding surfaces is called the optic centre.

There are two basic kinds of lenses: converging, and diverging. A converging lens brings all incident light-rays parallel to its optic axis together at a point F, behind the lens, called the focal point, or focus, of the lens. A diverging lens spreads out all incident light-rays parallel to its optic axis so that they appear to diverge from a virtual focal point F in front of the lens. Here, the front side of the lens is conventionally defined to be the side from which the light is incident.

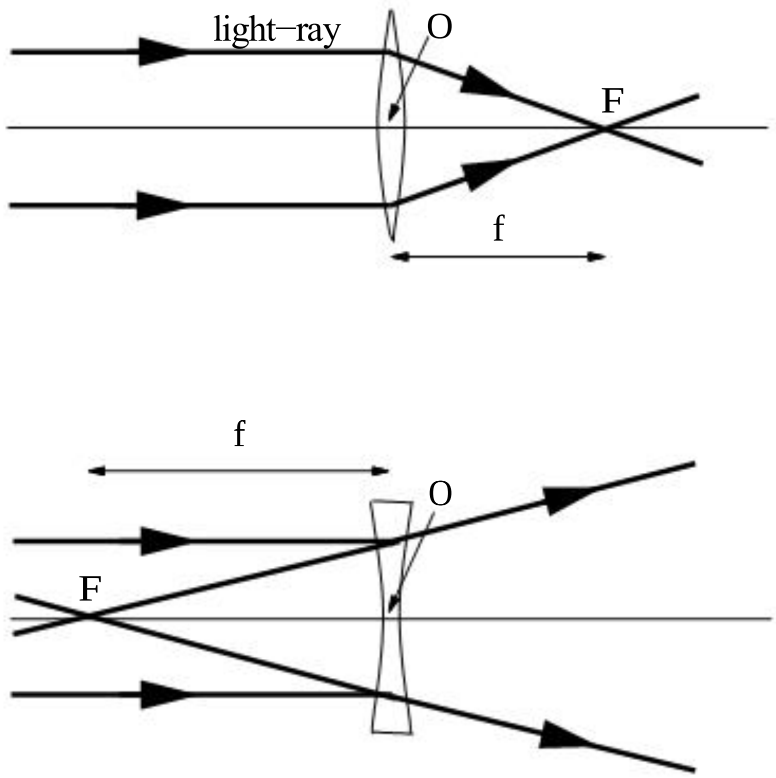


Figure 3.10: The focii of converging (top) and diverging (bottom) lens.

The differing effects of a converging and a diverging lens on incident light-rays parallel to the optic axis (i.e., emanating from a distant object) are illustrated in Fig. 3.10.

Lenses, like mirrors, suffer from spherical aberration, which causes light-rays parallel to the optic axis, but a relatively long way from the axis, to be brought to a focus, or a virtual focus, closer to the lens than light-rays which are relatively close to the axis. It turns out that spherical aberration in lenses can be completely cured by using lenses whose bounding surfaces are non-spherical. However, such lenses are more difficult, and, therefore, more expensive, to manufacture than conventional lenses whose bounding surfaces are spherical. Thus, the former sort of lens is only employed in situations where the spherical aberration of a conventional lens would be a serious problem. The usual method of curing spherical aberration is to use combinations of conventional lenses (i.e., compound lenses). In the following, we shall make use of the paraxial approximation, in which spherical aberration is completely ignored, and all light-rays parallel to the optic axis

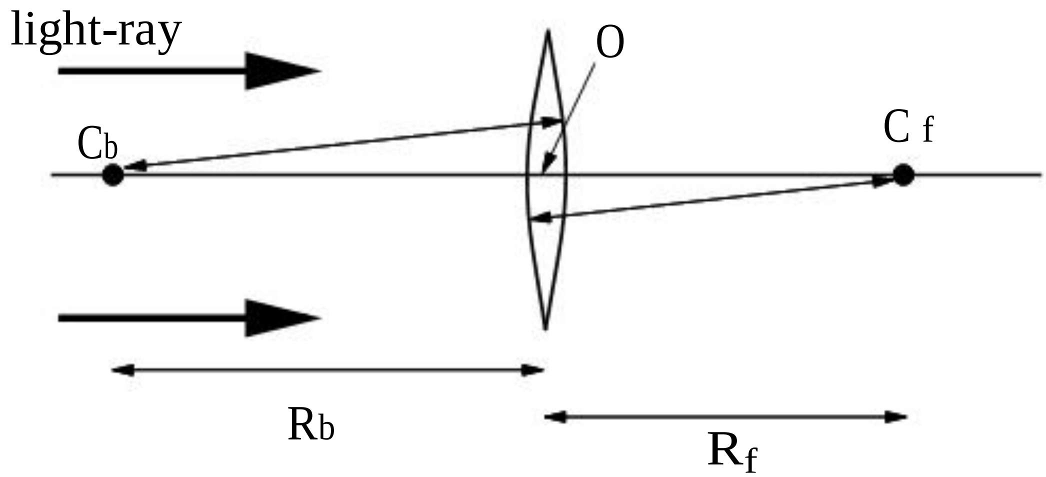


Figure 3.11: A thin lens.

are assumed to be brought to a focus, or a virtual focus, at the same point F. This approximation is valid as long as the radius of the lens is small compared to the object distance and the image distance.

The focal length of a lens, which is usually denoted f, is defined as the distance between the optic centre O and the focal point F, as shown in Fig. 3.10. However, by convention, converging lenses have positive focal lengths, and diverging lenses have negative focal lengths. In other words, if the focal point lies behind the lens then the focal length is positive, and if the focal point lies in front of the lens then the focal length is negative.

Consider a conventional lens whose bounding surfaces are spherical. Let C_f be the centre of curvature of the front surface, and C_b the centre of curvature of the back surface. The radius of curvature R_f of the front surface is the distance between the optic centre O and the point C_f. Likewise, the radius of curvature R_b of the back surface is the distance between points O and C_b. However, by convention, the radius of curvature of a bounding surface is positive if its centre of curvature lies behind the lens, and negative if its centre of curvature lies in front of the lens. Thus, in Fig. 3.11, R_f is positive and R_b is negative.

In the paraxial approximation, it is possible to find a simple formula relating the focal length f of a lens to the radii of curvature, R_f and R_b, of its front and back bounding surfaces. This formula is written

$$\frac{1}{f}=(n-1)\left(\frac{1}{R_f}-\frac{1}{R_b}\right), \qquad (3.12)$$

where n is the refractive index of the lens. The above formula is usually called the lens-maker's formula, and was discovered by Descartes. Note that the lens-maker's formula is only valid for a thin lens whose thickness is small compared to its focal length. What Eq. (3.12) is basically telling us is that light-rays which pass from air to glass through a convex surface are focused, whereas light-rays which pass from air to glass through a concave surface are defocused. Furthermore, since light-rays are reversible, it follows that rays which pass from glass to air through a convex surface are defocused, whereas rays which pass from air to glass through a concave surface are focused. Note that the net focusing or defocusing action of a lens is due to the difference in the radii of curvature of its two bounding surfaces.

Suppose that a certain lens has a focal length f. What happens to the focal length if we turn the lens around, so that its front bounding surface becomes its back bounding surface, and vice versa? It is easily seen that when the lens is turned around $R_f \rightarrow -R_b$ and $R_b \rightarrow -R_f$. However, the focal length f of the lens is invariant under this transformation, according to Eq. (3.12). Thus, the focal length of a lens is the same for light incident from either side. In particular, a converging lens remains a converging lens when it is turned around, and likewise for a diverging lens.

The most commonly occurring type of converging lens is a bi-convex, or doubleconvex, lens, for which $R_f > 0$ and $R_b < 0$. In this type of lens, both bounding surfaces have a focusing effect on light-rays passing through the lens. Another fairly common type of converging lens is a plano-convex lens, for which $R_f > 0$ and $R_b = \infty$. In this type of lens, only the curved bounding surface has a focusing effect on light-rays. The plane surface has no focusing or defocusing effect. A less common type of converging lens is a convex-meniscus lens, for which $R_f > 0$ and $R_b > 0$, with $R_f < R_b$. In this type of lens, the front bounding surface has a focusing effect on light-rays, whereas the back bounding surface has a defocusing effect, but the focusing effect of the front surface wins out.

The most commonly occurring type of diverging lens is a bi-concave, or doubleconcave, lens, for which $R_f < 0$ and $R_b > 0$. In this type of lens, both bounding surfaces have a defocusing effect on light-rays passing through the lens. Another fairly common type of converging lens is a plano-concave lens, for which $R_f < 0$ and $R_b = \infty$. In this type of lens, only the curved bounding surface has a defocusing effect on light-rays. The plane surface has no focusing or defocusing effect. A less common type of converging lens is a concave-meniscus lens, for which $R_f < 0$ and $R_b < 0$, with $R_f < |R_b|$. In this type of lens, the front bounding surface has a defocusing effect on light-rays, whereas the back bounding surface has a focusing effect, but the defocusing effect of the front surface wins out.

Figure 3.12 shows the various types of lenses mentioned above. Note that, as a general rule, converging lenses are thicker at the centre than at the edges, whereas diverging lenses are thicker at the edges than at the centre.

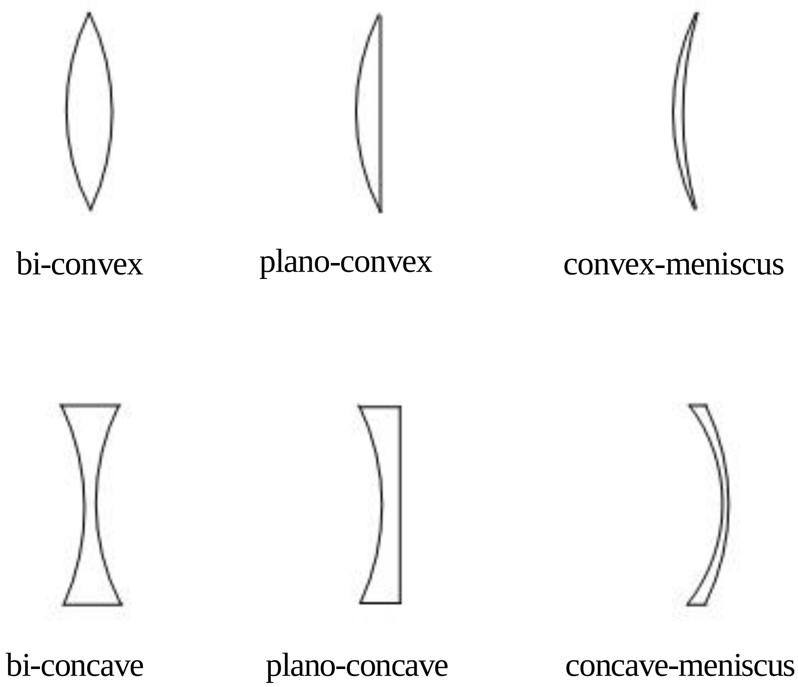


Figure 3.12: Various different types of thin lens.

3.6 Image Formation by Thin Lenses

There are two alternative methods of locating the image formed by a thin lens. Just as for spherical mirrors, the first method is graphical, and the second analytical.

The graphical method of locating the image formed by a thin lens involves drawing light-rays emanating from key points on the object, and finding where these rays are brought to a focus by the lens. This task can be accomplished using a small number of simple rules.

Consider a converging lens. It is helpful to define two focal points for such a lens. The first, the so-called image focus, denoted F_i , is defined as the point behind the lens to which all incident light-rays parallel to the optic axis converge after passing through the lens. This is the same as the focal point F defined previously. The second, the so-called object focus, denoted F_o , is defined as the position in front of the lens for which rays emitted from a point source of light placed at that position would be refracted parallel to the optic axis after passing through the lens. It is easily demonstrated that the object focus F_o is as far in front of the optic centre O of the lens as the image focus F_i is behind O. The distance from the optic centre to either focus is, of course, equal to the focal length f of the lens. The image produced by a converging lens can be located using just three simple rules:

1. An incident ray which is parallel to the optic axis is refracted through the image focus F_i of the lens.
2. An incident ray which passes through the object focus F_o of the lens is refracted parallel to the optic axis.
3. An incident ray which passes through the optic centre O of the lens is not refracted at all.

The last rule is only an approximation. It turns out that although a light-ray which passes through the optic centre of the lens does not change direction, it is displaced slightly to one side. However, this displacement is negligible for a thin lens.

Figure 3.13 illustrates how the image $S'T'$ of an object ST placed in front of a converging lens is located using the above rules. In fact, the three rays, 1-3, emanating from the tip T of the object, are constructed using rules 1-3, respectively. Note that the image is real (since light-rays actually cross), inverted, and diminished.

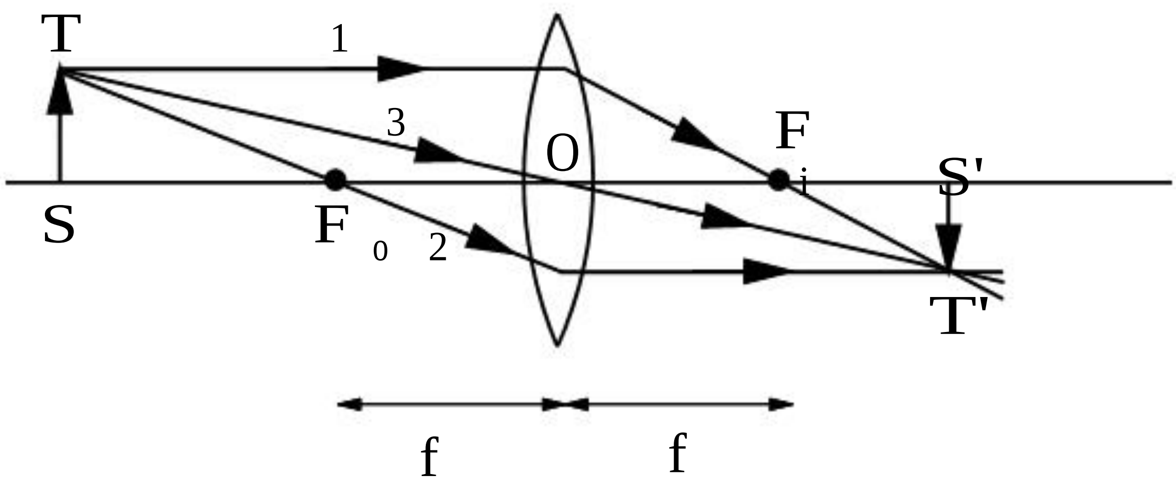


Figure 3.13: Image formation by a converging lens.

Consider a diverging lens. It is again helpful to define two focal points for such a lens. The image focus F_i is defined as the point in front of the lens from which all incident light-rays parallel to the optic axis appear to diverge after passing through the lens. This is the same as the focal point F defined earlier. The object focus F_o is defined as the point behind the lens to which all incident light-rays which are refracted parallel to the optic axis after passing through the lens appear to converge. Both foci are located a distance f from the optic centre, where f is the focal length of the lens. The image produced by a diverging lens can be located using the following three rules:

1. An incident ray which is parallel to the optic axis is refracted as if it came from the image focus F_i of the lens.

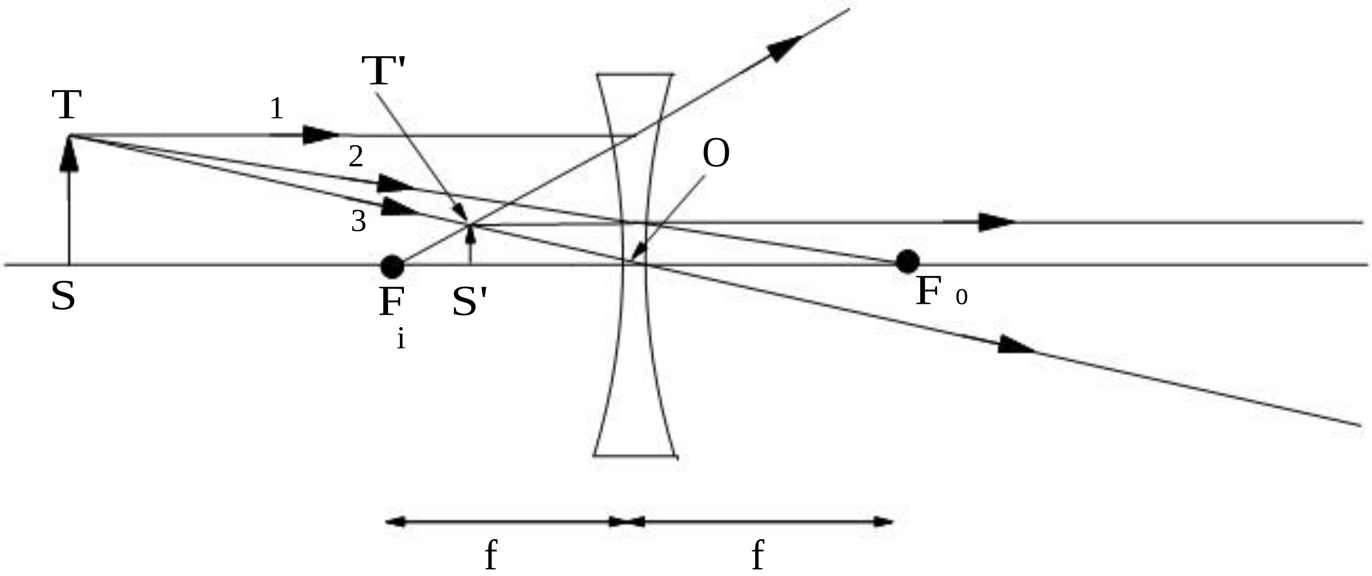


Figure 3.14: Image formation by a diverging lens.

2. An incident ray which is directed towards the object focus F_o of the lens is refracted parallel to the optic axis.

3. An incident ray which passes through the optic centre O of the lens is not refracted at all.

Figure 3.14 illustrates how the image $S'T'$ of an object ST placed in front of a diverging lens is located using the above rules. In fact, the three rays, 1-3, emanating from the tip T of the object, are constructed using rules 1-3, respectively. Note that the image is virtual (since light-rays do not actually cross), upright, and diminished.

Let us now investigate the analytical method. Consider an object of height h placed a distance p in front of a converging lens. Suppose that a real image of height h' is formed a distance q behind the lens. As is illustrated in Fig. 3.15, the image can be located using rules 1 and 3, discussed above.

Now, the right-angled triangles SOT and $S'OT'$ are similar, so

$$\frac{-h'}{h} = \frac{OS'}{OS} = \frac{q}{p}, \tag{3.13}$$

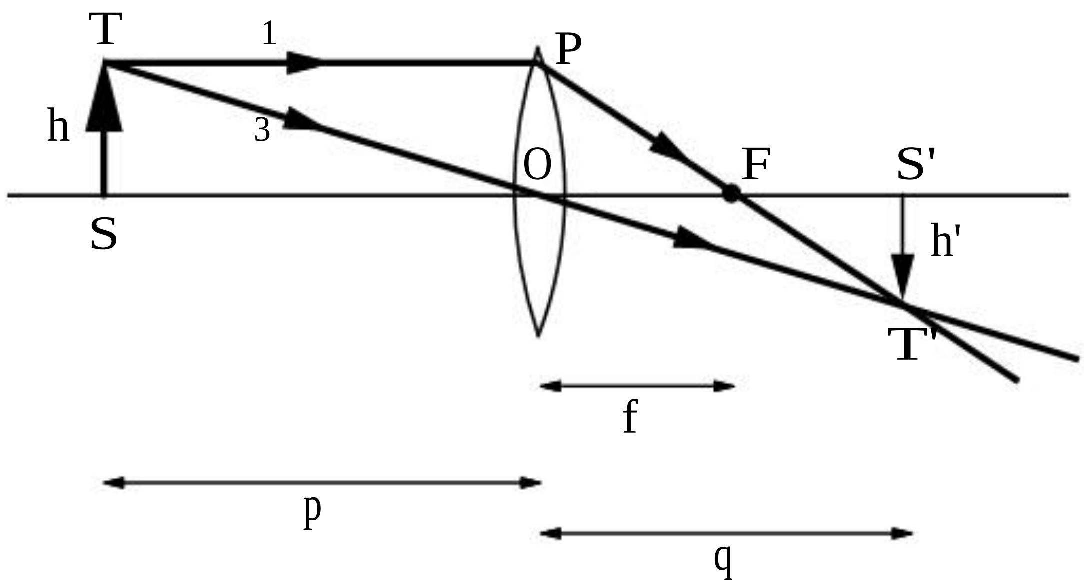


Figure 3.15: Image formation by a converging lens.

Here, we have adopted the convention that the image height h' is negative if the image is inverted. The magnification of a thin converging lens is given by

$$M = \frac{h'}{h} = -\frac{q}{p} \tag{3.14}$$

This is the same as the expression (3.2) for the magnification of a spherical mirror. Note that we are again adopting the convention that the magnification is negative if the image is inverted.

The right-angled triangles OPF and S′T′F are also similar, and so

$$\frac{S'T'}{OP} = \frac{FS'}{OF}, \tag{3.15}$$

or

$$\frac{-h'}{h} = \frac{q}{p} = \frac{q-f}{f}. \tag{3.16}$$

The above expression can be rearranged to give

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{f} \tag{3.17}$$

Note that this is exactly the same as the formula (3.8) relating the image and object distances in a spherical mirror.

Position of object	Position of image	Character of image
At $+\infty$	At F	Real, zero size
Between $+\infty$ and V_o	Between F and V_i	Real, inverted, diminished
At V_o	At V_i	Real, inverted, same size
Between V_o and F	Between V_i and $-\infty$	Real, inverted, magnified
At F	At $-\infty$	
Between F and O	From $+\infty$ to O	Virtual, upright, magnified
At O	At O	Virtual, upright, same size

Table 13.3: Rules for image formation by converging lenses.

Position of object	Position of image	Character of image
At ∞	At F_i	Virtual, zero size
Between ∞ and O	Between F_i and O	Virtual, upright, diminished
At O	At O	Virtual, upright, same size

Table 3.4: Rules for image formation by diverging lenses.

Although formulae (3.14) and (3.17) were derived for the case of a real image formed by a converging lens, they also apply to virtual images, and to images formed by diverging lenses, provided that the following sign conventions are adopted. First of all, as we have already mentioned, the focal length f of a converging lens is positive, and the focal length of a diverging lens is negative. Secondly, the image distance q is positive if the image is real, and, therefore, located behind the lens, and negative if the image is virtual, and, therefore, located in front of the lens. It immediately follows, from Eq. (3.14), that real images are always inverted, and virtual images are always upright.

Table 3.3 shows how the location and character of the image formed by a converging lens depend on the location of the object. Here, the point V_o is located on the optic axis two focal lengths in front of the optic centre, and the point V_i is located on the optic axis two focal lengths behind the optic centre. Note the almost exact analogy between the image forming properties of a converging lens and those of a concave spherical mirror.

Table 3.4 shows how the location and character of the image formed by a diverging lens depend on the location of the object. Note the almost exact analogy between the image forming properties of a diverging lens and those of a convex spherical mirror.

Finally, let us reiterate the sign conventions used to determine the positions and characters of the images formed by thin lenses:

1. The height h' of the image is positive if the image is upright, with respect to the object, and negative if the image is inverted.
2. The magnification M of the image is positive if the image is upright, with respect to the object, and negative if the image is inverted.
3. The image distance q is positive if the image is real, and, therefore, located behind the lens, and negative if the image is virtual, and, therefore, located in front of the lens.
4. The focal length f of the lens is positive if the lens is converging, so that the image focus F_i is located behind the lens, and negative if the lens is diverging, so that the image focus F_i is located in front of the lens.

Note that the front side of the lens is defined to be the side from which the light is incident.

3.7 Chromatic aberration

We have seen that both mirrors and lenses suffer from spherical aberration, an effect which limits the clarity and sharpness of the images formed by such devices. However, lenses also suffer from another type of abberation called chromatic abberation. This occurs because the index of refraction of the glass in a lens is different for different wavelengths. We have seen that a prism refracts violet light more than red light. The same is true of lenses. As a result, a simple lens focuses violet light closer to the lens than it focuses red light. Hence, white light produces a slightly blurred image of an object, with coloured edges.

For many years, chromatic abberation was a sufficiently serious problem for lenses that scientists tried to find ways of reducing the number of lenses in scientific instruments, or even eliminating them all together. For instance, Isaac Newton developed a type of telescope, now called the Newtonian telescope, which uses a mirror instead of a lens to collect light. However, in 1758, John Dollond, an English optician, discovered a way to eliminate chromatic abberation. He combined two lenses, one converging, the other diverging, to make an achromatic doublet. The two lenses in an achromatic doublet are made of different type of glass with indices of refraction chosen such that the combination brings any two chosen colours to the same sharp focus.

Modern scientific instruments use compound lenses (i.e., combinations of simple lenses) to simultaneously eliminate both chromatic and spherical aberration.

3.8 Worked Examples

Example 3.1: Concave mirrors

Question: An object of height h = 4 cm is placed a distance p = 15 cm in front of a concave mirror of focal length f = 20 cm. What is the height, location, and nature of the image? Suppose that the object is moved to a new position a distance p = 25 cm in front of the mirror. What now is the height, location, and nature of the image?

Answer: According to Eq. (3.8), the image distance q is given by

$$q = \frac{1}{1/f - 1/p} = \frac{1}{(1/20 - 1/15)} = -60 \text{ cm.}$$

Thus, the image is virtual (since $q < 0$), and is located 60 cm behind the mirror. According to Eq. (3.2), the magnification M of the image is given by

$$M = -\frac{q}{p} = \frac{(-60)}{(15)} = 4.$$

Thus, the image is upright (since $M > 0$), and magnified by a factor of 4. It follows that the height h' of the image is given by

$$h' = M\,h = (4)\,(4) = 16 \text{ cm.}$$

If the object is moved such that p = 25 cm then the new image distance is given by

$$q = \frac{1}{1/f - 1/p} = \frac{1}{(1/20 - 1/25)} = 100 \text{ cm.}$$

Thus, the new image is real (since q > 0), and is located 100 cm in front of the mirror. The new magnification is given by

$$M = -\frac{q}{p} = \frac{(100)}{(15)} = -6.67.$$

Thus, the image is inverted (since M < 0), and magnified by a factor of 6.67. It follows that the new height of the image is

$$h' = M\,h = -(6.67)\,(4) = -26.67 \text{ cm.}$$

Note that the height is negative because the image is inverted.

Example 3.2: Convex mirrors

Question: How far must an object be placed in front of a convex mirror of radius of curvature R = 50 cm in order to ensure that the size of the image is ten times less than the size of the object? How far behind the mirror is the image located?

Answer: The focal length f of a convex mirror is minus half of its radius of curvature (taking the sign convention for the focal lengths of convex mirrors into account). Thus, f = −25 cm. If the image is ten times smaller than the object then the magnification is M = 0.1. We can be sure that M = +0.1, as opposed to −0.1, because we know that images formed in convex mirrors are always virtual and upright. According to Eq. (3.2), the image distance q is given by

$$q = -M\,p,$$

where p is the object distance. This can be combined with Eq. (3.8) to give

$$p = (1 - \frac{1}{M}) = -(25)\,(1 - 10) = 225 \text{ cm.}$$

Thus, the object must be placed 225 cm in front of the mirror. The image distance is given by

$$q = -M\,p = -(0.1)\,(225) = -22.5 \text{ cm}.$$

Thus, the image is located 22.5 cm behind the mirror.

Example 3.3: Converging lenses

Question: An object of height h = 7 cm is placed a distance p = 25 cm in front of a thin converging lens of focal length f = 35 cm. What is the height, location, and nature of the image? Suppose that the object is moved to a new location a distance p = 90 cm in front of the lens. What now is the height, location, and nature of the image?

Answer: According to Eq. (3.17), the image distance q is given by

$$q = \frac{1}{1/f - 1/p} = \frac{1}{(1/35 - 1/25)} = -87.5 \text{ cm}.$$

Thus, the image is virtual (since $q < 0$), and is located 87.5 cm in front of the lens. According to Eq. (10.24), the magnification M of the image is given by

$$M = -\frac{q}{p} = -\frac{(-87.5)}{(25)} = 3.5$$

Thus, the image is upright (since $M > 0$), and magnified by a factor of 3.5. It follows that the height h' of the image is given by

$$h' = M\,h = (3.5)\,(7) = 24.5 \text{ cm}.$$

If the object is moved such that p = 90 cm then the new image distance is given by

$$q = \frac{1}{1/f - 1/p} = \frac{1}{(1/35 - 1/90)} = 57.27 \text{ cm}.$$

Thus, the new image is real (since $q > 0$), and is located 57.27 cm behind the lens. The new magnification is given by

$$M = -\frac{q}{p} = -\frac{(57.27)}{(90)} = -0.636$$

Thus, the image is inverted (since $M < 0$), and diminished by a factor of 0.636. It follows that the new height of the image is

$$h' = M\,h = -(9.636)\,(7) = -4.45 \text{ cm}.$$

Note that the height is negative because the image is inverted.

Example 3.4: Diverging lenses

Question: How far must an object be placed in front of a diverging lens of focal length 45 cm in order to ensure that the size of the image is fifteen times less than the size of the object? How far in front of the lens is the image located?

Answer: The focal length f of a diverging lens is negative by convention, so $f = -45$ cm, in this case. If the image is fifteen times smaller than the object then the magnification is $M = 0.0667$. We can be sure that $M = +0.0667$, as opposed to -0.0667, because we know that images formed in diverging lenses are always virtual and upright. According to Eq. (3.14), the image distance q is given by

$$q = -M\,p,$$

where p is the object distance. This can be combined with Eq. (3.17) to give

$$p = f\,(1 - \frac{1}{M}) = -(45)(1-15) = 630 cm.$$

Thus, the object must be placed 630 cm in front of the lens. The image distance is given by

$$q = -M\,p = -(0.0667)\,(630) = -42 \text{ cm}.$$

Thus, the image is located 42 cm in front of the lens.