1 Introduction

Light can be described as a field $E(\vec{r},t)$ that varies in space and time. These variations can be decomposed into spatial frequencies $\vec{\kappa}$ and temporal frequencies ν respectively. That is, we write

$$E(\vec{r},t) = \iint \mathrm{d}^3\vec{\kappa}\,\mathrm{d}\nu\,\mathcal{E}(\vec{\kappa},\nu)e^{i2\pi\left(\vec{\kappa}\cdot\vec{r}-\nu t\right)} \tag{1.1}$$

$$\mathcal{E}(\vec{\kappa},\nu) = \iint \mathrm{d}^3\vec{r}\,\mathrm{d}t\,E(\vec{r},t)e^{-i2\pi\left(\vec{\kappa}\cdot\vec{r}-\nu t\right)} \tag{1.2}$$

where throughout this book $\mathcal{E}(\vec{\kappa},\nu)$ will be called a radiant field and $\vec{\kappa} = (\kappa_x,\kappa_y,\kappa_z)$ is the wavevector associated with the field. Note that a distinction will be made here between a wavevector $\vec{\kappa}$ and an angular wavevector $\vec{k} = 2\pi\vec{\kappa}$, the latter being commonly found in the literature. Because $\vec{\kappa}$ is the Fourier conjugate variable of the position vector $\vec{r} = (x,y,z)$, it can be thought of as associated with momentum.

To each temporal frequency ν is associated a wavenumber κ (as distinct from an angular wavenumber), defined by

$$\kappa = \frac{n}{c}\nu \tag{1.3}$$

where n is the index of refraction of the surrounding medium ($n = 1$ for free space) and c is the speed of light in free space ($c = 3.0 \times 10^8$ m/s). To each wavenumber is also associated a free-space wavelength, defined by

$$\lambda = \frac{n}{\kappa} \tag{1.4}$$

The use of the variables $\vec{\kappa}$ to denote wavevector and κ to denote wavenumber is not accidental. As we will see in Chapter 2, for propagating fields the two are linked by a fundamental law known as the energy–momentum relation, given by

$$\left|\vec{\kappa}\right| = \kappa \tag{1.5}$$

Written in this notation this relationship appears self-evident, but one must bear in mind that the left-hand side represents the integration variables in Eq. 1.1 (related to momentum) while the right-hand side comes from Eq. 1.3 (related to energy). In other words, the wavevector here can have arbitrary direction but its magnitude is constrained to lie on a spherical shell defined by the optical frequency ν. Off-shell components of $\vec{\kappa}$ are possible, but these do not propagate. Since we will only concern ourselves with propagating fields throughout this book, the notation in Eq. 1.5 will be maintained. Equation 1.5 is perhaps the most fundamental relation in imaging theory and stems directly from the wave equation, as will be seen in the next chapter.

1.1 COMPLEX FIELDS

Throughout this book, the field $E(\vec{r}, t)$ will be treated as complex. That is, $E(\vec{r}, t)$ should not be confused with an electric field, since, by definition, an electric field is a physically measurable quantity that must be real. Nevertheless, $E(t)$ will serve as a *representation* of an electric field $E_{\mathrm{elec}}(\vec{r}, t)$, such that

$$E_{\mathrm{elec}}(\vec{r}, t) \propto \mathrm{Re}\left[E(\vec{r}, t)\right] \tag{1.6}$$

While this relation places a constraint on $\mathrm{Re}\left[E(\vec{r}, t)\right]$, it allows $\mathrm{Im}\left[E(\vec{r}, t)\right]$ to be chosen arbitrarily. By general convention, this arbitrariness is removed when $\mathrm{Im}\left[E(\vec{r}, t)\right]$ is chosen to fulfill a second condition, applied now to the radiant field and given by

$$\mathcal{E}(\vec{\kappa}, \nu) = 0 \qquad \text{when} \qquad \nu < 0 \tag{1.7}$$

This second condition implies that $E(\vec{r}, t)$ is an analytic function of t. It imposes the constraint that $\mathrm{Re}\left[E(\vec{r}, t)\right]$ and $\mathrm{Im}\left[E(\vec{r}, t)\right]$ are related by what is known as a Hilbert transform (in t), thereby specifying $E(\vec{r}, t)$ completely. While the conditions imposed by Eqs. 1.6 and 1.7 will rarely be mentioned again throughout this book, they should nevertheless always be kept in mind. For more information on analytic functions and Hilbert transforms, the reader can consult [1] or [3].

The purpose of this introductory chapter is to motivate some general concepts in optical imaging theory. In the most basic imaging applications, light from a two-dimensional (2D) plane (called the object plane) is mapped onto another 2D plane (called the image plane) some distance away. The goal of imaging theory is to analyze this mapping process. Because 2D planes are the starting point of imaging theory, our coordinate systems will be tailored accordingly and we write $\vec{r} = (\vec{\rho}, z)$, where $\vec{\rho}$ lies in the 2D plane of interest. We also begin by considering light whose time dependence contains only a single harmonic frequency ν_0. Such light is called monochromatic. The field in a 2D plane is then written in a simplified form as $E(\vec{\rho})$, where the harmonic time dependence is implicit. Correspondingly, the general Fourier transform relations linking a 2D field and 2D radiant field reduce to

$$E(\vec{\rho}) = \int \mathrm{d}^2\vec{\kappa}_\perp\, \mathcal{E}(\vec{\kappa}_\perp) e^{i2\pi\vec{\rho}\cdot\vec{\kappa}_\perp} \tag{1.8}$$

$$\mathcal{E}(\vec{\kappa}_\perp) = \int \mathrm{d}^2\vec{\rho}\, E(\vec{\rho}) e^{-i2\pi\vec{\kappa}_\perp\cdot\vec{\rho}} \tag{1.9}$$

where $\vec{\kappa} = (\vec{\kappa}_\perp, \kappa_z)$.

Having established this basic formalism for complex fields, we turn now to some intuitive notions of beam directionality and ray optics.

1.2 INTENSITY AND RADIANCE

Typical optical frequencies are on the order of $\nu_0 \sim 10^{15}$ Hz, which is far too fast to be directly measurable with detectors based on current technology. Standard detectors do not directly measure the light field but rather the light intensity, defined by

$$I(\vec{\rho}) = \left\langle E(\vec{\rho})E^*(\vec{\rho})\right\rangle \tag{1.10}$$

where the brackets $\langle ... \rangle$ denote a temporal average over many temporal oscillations (this time average will be better defined in Chapter 7). Because $I(\vec{\rho})$ is a physically measurable parameter, it must be real. Throughout this book, $I(\vec{\rho})$ will provide a unit of reference. In particular, $I(\vec{\rho})$ has units of power per area, or $\mathrm{W/m^2}$. Correspondingly, $E(\vec{\rho})$, from the relationship defined above, must have units $\sqrt{\mathrm{W}}/\mathrm{m}$ and $\mathcal{E}(\vec{\kappa}_\perp)$ must have units $\sqrt{\mathrm{W}}\mathrm{m}$. Again, these units are not those associated with the physically measurable electric field $E_{\mathrm{elec}}(\vec{\rho})$ (units: V/m), and should only be interpreted as convenient units to verify dimensional consistency.

Technically, as defined by Eq. 1.10, $I(\vec{\rho})$ should be referred to as an irradiance rather than an intensity; however, in recent times the latter term seems to have come into favor, particularly amongst experimentalists. This book will yield to the new jargon.

To gain an intuitive picture of light propagation from one 2D plane to another, it is often convenient to think of light as composed of rays that transport power, as depicted in Fig. 1.1. The rigorous connection between the notion of light rays and the description of light as complex fields remains a difficult problem. One might be tempted to think of the direction of a light ray

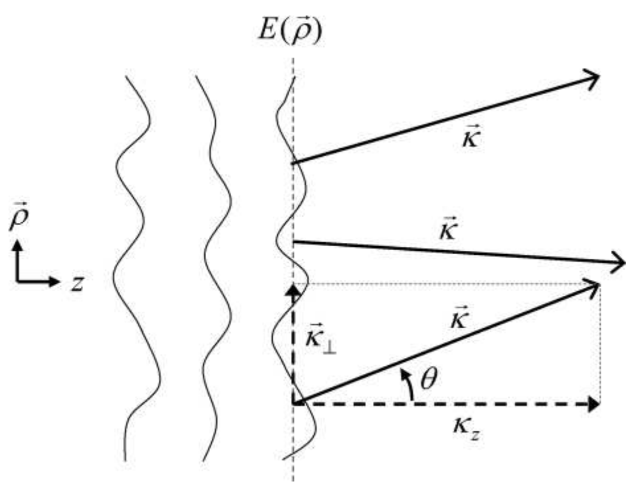

Figure 1.1. Connection between wave optics and ray optics.

as linked to a wavevector $\vec{\kappa}$. Following this line of reasoning, one might consider inferring this wavevector $\vec{\kappa}$ from a Fourier transform of $I(\vec{\rho}) = \langle E(\vec{\rho})E^*(\vec{\rho})\rangle$. While this might have been a good starting point, it is clearly problematic because a Fourier transform of $\langle E(\vec{\rho})E^*(\vec{\rho})\rangle$ is inherently non-local in space and does not fit our intuitive notion of light rays emanating from specific locations with specific directions. Instead, a commonly accepted connection between complex fields and ray optics was formulated by Walther [8] and Friberg [2], and is based on a parameter called the radiance [7], defined by

$$\mathcal{L}(\vec{\kappa}_\perp;\vec{\rho}) = \kappa^2 \int \mathrm{d}^2\vec{\rho}\,' \left\langle E(\vec{\rho} + \tfrac{1}{2}\vec{\rho}\,')E^*(\vec{\rho} - \tfrac{1}{2}\vec{\rho}\,')\right\rangle e^{-i2\pi\vec{\kappa}_\perp\cdot\vec{\rho}\,'} \tag{1.11}$$

By construction, the radiance is a *local* Fourier transform (often called a Wigner function), though not of the intensity but rather of the field autocorrelation function. The direction of the light ray emanating from point $\vec{\rho}$ is then prescribed by the value of $\vec{\kappa}_\perp$ about which $\mathcal{L}(\vec{\kappa}_\perp;\vec{\rho})$ is peaked. The local angle of propagation from point $\vec{\rho}$, accordingly, is defined by $(\theta_x, \theta_y) = \left(\sin^{-1}(\hat{\kappa}_x/\kappa), \sin^{-1}(\hat{\kappa}_y/\kappa)\right)$, where $(\hat{\kappa}_x, \hat{\kappa}_y)$ corresponds to this peak. This connection between radiance and ray direction is valid only for fields that are slowly spatially varying on the scale of a wavelength, meaning, in effect, it is valid only for small angles, a condition known as the paraxial limit, which will be invoked repeatedly. Indeed, the paraxial limit has been implicitly assumed in the definition of radiance provided above.

1.3 RAY OPTICS

The notion of light rays is very convenient in that it provides a simple and intuitive description of light propagation through basic elements of an optical imaging device. Since $\vec{\rho}$ and $\vec{\kappa}_\perp$ are Fourier conjugate coordinates, we may also think of (x, y) and $(n\sin\theta_x, n\sin\theta_y)$ as conjugate coordinates, or, in the paraxial limit, (x, y) and $(n\theta_x, n\theta_y)$. A light ray can then be described as a vector

$$\begin{pmatrix} x \\ n\theta_x \end{pmatrix} \tag{1.12}$$

where, for purposes of discussion, we consider rays in the x–z plane only. This vector indicates the position and direction of the ray at a given optical plane (plane 0). The effect of an optical system, imaging or otherwise, is to transfer this ray to a new plane (plane 1) leading, in general, to a change both in position and direction of the ray. A linear optical system can then be characterized by a general transfer matrix $\mathbf{M}$, such that,

$$\begin{pmatrix} x_1 \\ n_1\theta_{x1} \end{pmatrix} = \mathbf{M}\cdot\begin{pmatrix} x_0 \\ n_0\theta_{x0} \end{pmatrix} \tag{1.13}$$

as schematically depicted in Fig. 1.2, where allowances have been made for different indices of refraction on either side of the system. $\mathbf{M}$ is often referred to as an ABCD transfer matrix

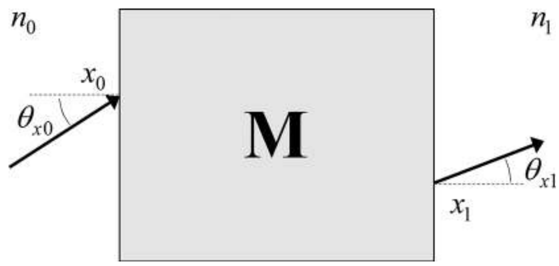

Figure 1.2. General ABCD matrix.

because it consists of four elements. Detailed discussions of ABCD transfer matrix formalism are provided in several optics textbooks, such as [4] and [6].

The following are a few basic results:

For propagation through an interface, then

$$\mathbf{M}_{\text{interface}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{1.14}$$

indicating that on either side of the interface we have $x_1 = x_0$ and $n_1\theta_1 = n_0\theta_0$, the latter being a statement of conservation of momentum along the x direction (also known as Snell's law).

For propagation an axial distance z through a medium of index n, then

$$\mathbf{M}_{\text{propagation}} = \begin{pmatrix} 1 & z/n \\ 0 & 1 \end{pmatrix} \tag{1.15}$$

For propagation through a thin lens of focal length f surrounded by a medium of index n, then

$$\mathbf{M}_{\text{lens}} = \begin{pmatrix} 1 & 0 \\ -n/f & 1 \end{pmatrix} \tag{1.16}$$

For a system that performs perfect imaging with magnification M, then

$$\mathbf{M}_{\text{image}} = \begin{pmatrix} M & 0 \\ 0 & 1/M \end{pmatrix} \tag{1.17}$$

Finally, for a system that exchanges the coordinates x and $n\theta_x$, then

$$\mathbf{M}_{\text{FT}} = \begin{pmatrix} 0 & L \\ -1/L & 0 \end{pmatrix} \tag{1.18}$$

where the length scale L is introduced for dimensional consistency. Since x and $n\theta_x$ can be thought of as Fourier conjugate coordinates, $\mathbf{M}_{\text{FT}}$ can be though of as performing a perfect Fourier transform where the coordinates are swapped, similar to Eqs. 1.8 and 1.9 (hence the subscript FT).

Two observations can be made. First, all of the above transfer matrices satisfy the condition $\det[\mathbf{M}] = 1$. This is in fact a general rule for lossless systems, resulting from the conservation of a fundamental quantity called the étendue of the light beam. Thus, for perfect imaging (Eq. 1.17), the magnification of the position by a factor M is necessarily accompanied by the de-magnification of the propagation angle by a factor $1/M$. A similar conclusion holds for a perfect Fourier transform (Eq. 1.18), though applied to the crossed terms. Much more will be said about étendue in Chapter 6.

A second observation relates to the effect of the index of refraction on propagation distances. A distinction can be made between physical distances, such as z and f, and their associated optical distances, such as nz and nf. For systems embedded in a medium, it is the latter that become important.

Optical imaging systems, in general, involve the transfer of light rays through a medium (or several media) and through lenses. We consider these transfers separately and then in combination. For simplicity, we begin with systems surrounded by free space ($n = 1$). A more general case involving unequal indices of refraction will be considered at the end of the chapter.

Propagation Through Free Space

In the simplest scenario, an arbitrary light ray starts at plane 0 and propagates an axial distance z, leading to

$$\begin{pmatrix} x_0 + z\theta_{x0} \\ \theta_{x0} \end{pmatrix} = \mathbf{M}_{\text{propagation}} \cdot \begin{pmatrix} x_0 \\ \theta_{x0} \end{pmatrix} \tag{1.19}$$

In the limit where z becomes very large, we can eventually neglect the starting position of the ray and write

$$\begin{pmatrix} x_1 \\ \theta_{x1} \end{pmatrix} \rightarrow \begin{pmatrix} \approx z\theta_{x0} \\ \theta_{x0} \end{pmatrix} \tag{1.20}$$

This limit is called the far-field or Fraunhofer limit. It is important because we find that in this limit the position of the ray scales directly with its initial starting angle, the scaling factor being simply z. In effect, far-field propagation can be thought of as performing a scaled Fourier transform of the initial field by exchanging x_0 with $z\theta_{x0}$. However, this Fourier transform is one-way since only x_0 is exchanged with θ_{x0} and not vice versa.

Propagation Through a Lens

Next, we examine the propagation of a light ray through a thin lens by taking the planes 0 and 1 to be on either side of the lens.

Referring to Eq. 1.16, we conclude that if a ray impinges upon the center of the lens at an arbitrary angle θ_{x0}, then it exits the center of the lens undeviated. That is,

$$\begin{pmatrix} 0 \\ \theta_{x0} \end{pmatrix} = \mathbf{M}_{\text{lens}} \cdot \begin{pmatrix} 0 \\ \theta_{x0} \end{pmatrix} \tag{1.21}$$

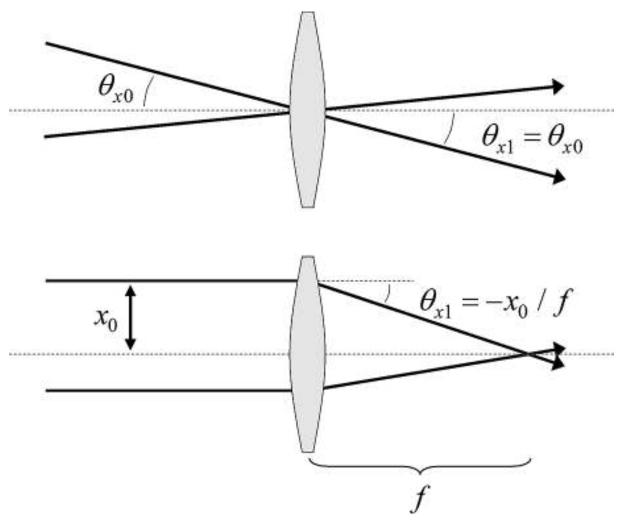

Figure 1.3. Propagation of rays through a lens of focal length *f*.

Similarly, if a ray travels parallel to the optical axis but impinges upon the lens at an arbitrary position, then it exits the lens at the same position, but deflected at an angle:

$$\begin{pmatrix} x_0 \\ -x_0/f \end{pmatrix} = \mathbf{M}_{\text{lens}} \cdot \begin{pmatrix} x_0 \\ 0 \end{pmatrix} \tag{1.22}$$

This deflection angle causes the outgoing ray to intersect the optical axis at a distance *f* from the lens, independently of the initial position of the ray at the lens. In other words, a lens causes *any* ray parallel to the optical axis to converge to the same focal point. Both the above properties are illustrated in Fig. 1.3.

1.4 BASIC TRANSFER PROPERTIES OF A LENS

Finally, we derive some basic transfer properties of a lens in free space by again considering planes 0 and 1 on either side of the lens, but this time generalizing the position of these planes to be arbitrary distances s_0 and s_1, respectively, from the lens. The transfer of a light ray from plane 0 to plane 1 is then governed by the composite matrix

$$\mathbf{M}_{\text{T}} = \mathbf{M}_{\text{free}}(s_1) \cdot \mathbf{M}_{\text{lens}}(f) \cdot \mathbf{M}_{\text{free}}(s_0) = \begin{pmatrix} 1 - \frac{s_1}{f} & s_0 + s_1 - \frac{s_0 s_1}{f} \\ -\frac{1}{f} & 1 - \frac{s_0}{f} \end{pmatrix} \tag{1.23}$$

Two specific cases are of particular interest:

1.4.1 Fourier Transform with a Lens

In the event that $s_0 = s_1 = f$, Eq. 1.23 simplifies to

$$\mathbf{M}_\mathrm{T} = \begin{pmatrix} 0 & f \\ -1/f & 0 \end{pmatrix} \tag{1.24}$$

and we recognize from Eq. 1.18 that in this specific configuration a lens performs a perfect Fourier transform accompanied by a scaling factor f. It should be emphasized that no Fraunhofer approximation was required to obtain this Fourier transform analogy, in contrast to the far-field propagation result derived above (the small angle approximation, however, remains valid). Moreover, the Fourier transform operation derived here is now two-way.

1.4.2 Imaging with a Lens

In the event that s_0 and s_1 satisfy a relation known as the thin-lens formula, given by

$$\frac{1}{s_0} + \frac{1}{s_1} = \frac{1}{f} \tag{1.25}$$

then Eq. 1.23 simplifies to

$$\mathbf{M}_\mathrm{T} = \begin{pmatrix} M & 0 \\ -1/f & 1/M \end{pmatrix} \tag{1.26}$$

where we have introduced the magnification factor

$$M = -\frac{s_1}{s_0} \tag{1.27}$$

A lens performs near-perfect imaging in this specific configuration, as illustrated in Fig. 1.4. The only defect in the imaging arises from the off-axis element $-1/f$ that imparts an extraneous dependence of the propagation angle at the image plane (plane 1) on the ray position at the object plane (plane 0), as is observed by carrying out the matrix multiplication

$$\begin{pmatrix} x_1 \\ \theta_{x1} \end{pmatrix} = \mathbf{M}_\mathrm{T} \cdot \begin{pmatrix} x_0 \\ \theta_{x0} \end{pmatrix} = \begin{pmatrix} Mx_0 \\ -x_0/f + \theta_{x0}/M \end{pmatrix} \tag{1.28}$$

This extraneous dependence corresponds to a coupling between the ray direction and the ray position at the image plane, which would be absent in the case of a perfect imaging system described by Eq. 1.17. One consequence of such coupling is apparent in Fig. 1.4. If s_1 is increased or decreased relative to its in-focus value, not only does the image become blurred (the rays do not converge properly), but the magnification increases or decreases as well. Such axially dependent magnification is often problematic in microscopy applications (more on this in Chapter 5). Imaging systems that do not exhibit this problem and where magnification remains a constant independent of defocus are called telecentric systems. Modern microscopes are almost always telecentric, and we will consider these in detail in subsequent chapters. In the

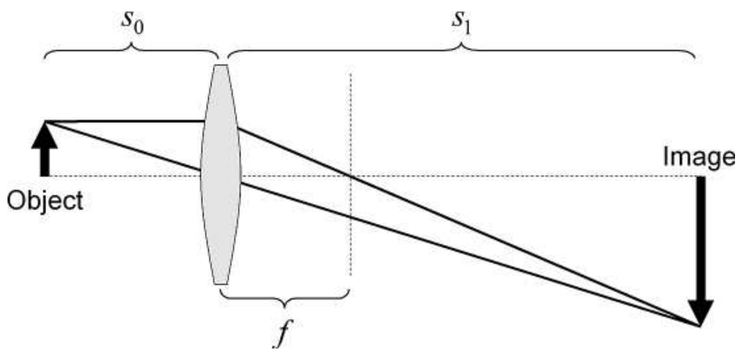

Figure 1.4. Imaging with a lens of focal length *f*.

meantime, we only note that the limit of telecentricity can be approached with a single lens when $|x_0| \ll |\theta_{x0} f/M|$, in which case Eq. 1.28 simplifies to

$$\begin{pmatrix} x_1 \\ \theta_{x1} \end{pmatrix} \rightarrow \begin{pmatrix} M x_0 \\ \approx \theta_{x0}/M \end{pmatrix} \tag{1.29}$$

Optical planes connected by an imaging operation are called conjugate planes, whereas those connected by a Fourier transform operation are called Fourier planes. We will make use of this terminology throughout this book.

Axial Magnification

Though we have considered only transverse imaging magnifications so far, we can readily derive an axial imaging magnification from the thin lens formula. For small axial displacements, this axial magnification is defined by

$$M_z = \frac{\mathrm{d}s_1}{\mathrm{d}s_0} \tag{1.30}$$

Taking the derivative of both sides of Eq. 1.25 with respect to s_0, we find

$$M_z = -\left(\frac{s_1}{s_0}\right)^2 = -M^2 \tag{1.31}$$

1.4.3 Thick Lens

The transfer matrix $\mathbf{M}_{\text{lens}}$ (Eq. 1.16) is applicable to thin lenses only, so thin that a ray traversing the lens incurs only a negligible lateral displacement. In practice, however, a lens always possesses a finite thickness. In many cases it can even comprise several optical elements and be quite thick indeed, an example being a microscope objective. Nevertheless, even a thick lens, if properly designed, possesses a well-defined focal length. This focal length is defined not from the center of the lens but rather from what is known as a principal plane, as illustrated in Fig. 1.5.

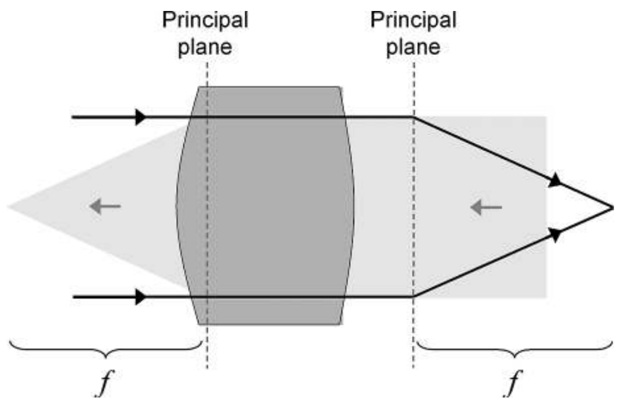

Figure 1.5. Principal planes of a thick lens of focal length *f*.

An important property of lenses in general is that they obey a principle of optical reciprocity wherein if a collimated bundle of rays parallel to the optical axis is focused to a point when incident from one side of the lens, then it is focused to another point when incident from the opposite side. In the case of a thin lens, these points are located a distance *f* from either side of the lens. In the case of a thick lens, however, they are located distances *f* from the associated principal planes of the lens, as illustrated in Fig. 1.5. The difference between thin and thick lenses is that for a thin lens the principal planes are merged into a single plane located exactly at the lens plane, whereas for a thick lens they are displaced from one another. This displacement can be quite large, often larger than the physical thickness of the lens itself. Moreover, the principal planes may be distributed asymmetrically relative to the physical lens center, and may even be on opposite sides of each other from what might be expected. In any case, regardless of the location of the principal planes the focal lengths of a thick lens in either direction are the same, provided only that the index of refraction of the media on either side of the lens is also the same.

It should be emphasized that this principle of reciprocity does not mean that a thick lens works just as well in either orientation. The angular acceptance of a thick lens can be quite different depending on its orientation, leading to very different fields of view in an imaging application. More will be said about this in Chapter 6.

Effect of Surrounding Media

So far, we have mostly considered systems surrounded by free space, and only made general allowances for the possibility of surrounding media other than free space. We close this chapter with a more detailed examination of the effects of different indices of refraction. For example, let us consider a thick lens, itself of refractive index n_f, surrounded on one side by refractive index n_0 and on the other by refractive index n_1, as shown in Fig. 1.6. The thin-lens formula in this case becomes more complicated (cf. [5]), and assumes the more general form given by

$$\frac{n_0}{s_0} + \frac{n_1}{s_1} = \frac{n_f - n_0}{R_0} - \frac{n_f - n_1}{R_1} \tag{1.32}$$