We square both sides of the equation and carry out the square of the remaining binomial term to obtain

$$16a^2x^2 - 4ab^2x + b^4 = 4b^2\left(x^2 - 2ax + a^2 + y^2\right),$$

and then cancel and regroup terms to arrive at

$$\left(16a^2 - 4b^2\right)x^2 - 4b^2y = 4a^2b^2 - b^4.$$

Finally, we divide both sides of the equation by the term on the right to obtain the (hopefully) familiar form of an ellipse

$$\frac{x^2}{\left(\frac{b^2}{4}\right)} + \frac{y^2}{\left(\frac{b^2}{4} - a^2\right)} = 1$$

9.4 Paraxial Rays and ABCD Matrices

In the remainder of this chapter we develop a formalism for describing the effects of mirrors and lenses on rays of light. Keep in mind that when describing light as a collection of rays rather than as waves, the results can only describe features that are macroscopic compared to a wavelength. The rays of light at each location in space describe approximately the direction of travel of the wave fronts at that location. Since the wavelength of visible light is extraordinarily small compared to the macroscopic features that we perceive in our day-to-day world, the ray approximation is often a very good one. This is the reason that ray optics was developed long before light was understood as a wave.

We consider ray theory within the *paraxial approximation*, meaning that we restrict our attention to rays that are near and almost parallel to an *optical axis* of a system, say the z-axis. It is within this approximation that the familiar imaging properties of lenses occur. An image occurs when all rays from a *point* on an *object* converge to a corresponding *point* on what is referred to as the *image*. To the extent that the paraxial approximation is violated, the clarity of an image can suffer, and we say that there are *aberrations* present. Very often in the field of optical engineering, one is primarily concerned with minimizing aberrations in cases where the paraxial approximation is not strictly followed. This is done so that, for example, a camera can take pictures of subjects that occupy a fairly wide angular field of view, where rays violate the paraxial approximation. Optical systems are typically engineered using the science of *ray tracing*, which is described briefly in section 9.9.

As we develop paraxial ray theory, we should remember that rays impinging on devices such as lenses or curved mirrors should strike the optical component at near normal incidence. To quantify this statement, the paraxial approximation is valid to the extent that we have

$$\sin\theta \cong \theta \tag{9.24}$$

and similarly

$$\tan\theta \cong \theta \tag{9.25}$$

Here, the angle θ (in radians) represents the angle that a particular ray makes with respect to the optical axis. There is an important mathematical reason for this approximation.

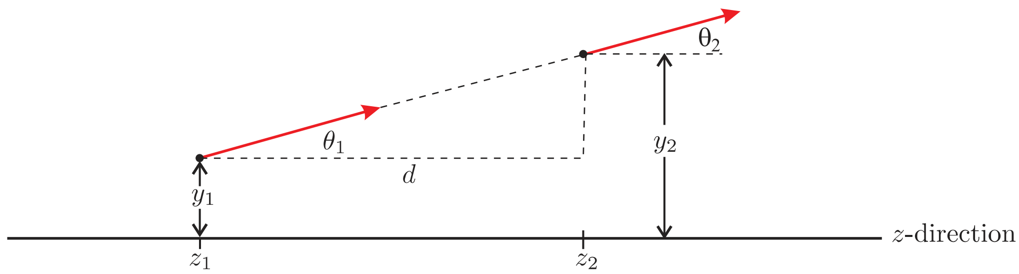

Figure 9.6 The behavior of a ray as light traverses a distance d.

The sine is a nonlinear function, but at small angles it is approximately linear and can be represented by its argument. It is this linearity that is crucial to the process of forming images. The linearity also greatly simplifies the formulation since it reduces the problem to linear algebra. Conveniently, we will be able to keep track of imaging effects with a 2×2 matrix formalism.

Consider a ray confined to the y–z plane where the optical axis is in the z-direction. Let us specify a ray at position z_1 by two coordinates: the displacement from the axis y_1 and the orientation angle θ_1 (see Fig. 9.6). The ray continues along a straight path as it travels through a uniform medium. This makes it possible to predict the coordinates of the same ray at other positions, say at z_2. The connection is straightforward. First, since the ray continues in the same direction, we have

$$\theta_2 = \theta_1 \tag{9.26}$$

By referring to Fig. 9.6 we can write y_2 in terms of y_1 and θ_1:

$$y_2 = y_1 + d \tan\theta_1 \tag{9.27}$$

where $d \equiv z_2 - z_1$. Equation (9.27) is nonlinear in θ_1. However, in the paraxial approximation (9.25) it becomes linear, which after all is the point of the approximation. In this approximation the expression for y_2 becomes

$$y_2 = y_1 + d\theta_1 \tag{9.28}$$

Equations (9.26) and (9.28) describe a linear transformation which in matrix notation can be consolidated into the form

$$\begin{bmatrix} y_2 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ \theta_1 \end{bmatrix} \quad \text{(propagation through a distance } d\text{)} \tag{9.29}$$

Here, the vectors in this equation specify the essential information about the ray before and after traversing the distance d, and the matrix describes the effect of traversing the distance. This type of matrix is called an ABCD matrix.

Suppose that the distance d is subdivided into two distances, a and b, such that $d = a+b$. If we consider individually the effects of propagation through a and through b, we have

$$\begin{aligned} \begin{bmatrix} y_{\text{mid}} \\ \theta_{\text{mid}} \end{bmatrix} &= \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ \theta_1 \end{bmatrix} \\ \begin{bmatrix} y_2 \\ \theta_2 \end{bmatrix} &= \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_{\text{mid}} \\ \theta_{\text{mid}} \end{bmatrix} \end{aligned} \tag{9.30}$$

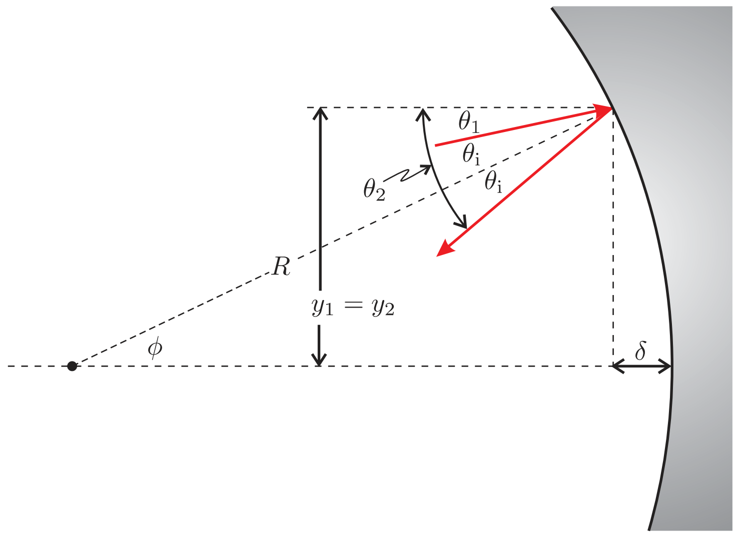

Figure 9.7 A ray depicted in the act of reflection from a curved surface.

where the subscript "mid" refers to the ray in the middle position after traversing the distance a. If we combine the equations, we get

$$\begin{bmatrix} y_2 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ \theta_1 \end{bmatrix} \tag{9.31}$$

which is in complete agreement with (9.29) since the ABCD matrix for the entire displacement is

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & a+b \\ 0 & 1 \end{bmatrix} \tag{9.32}$$

9.5 Reflection and Refraction at Curved Surfaces

We next consider the effect of reflection from a spherical surface as depicted in Fig. 9.7. We consider only the act of reflection without considering propagation before or after the reflection takes place. Thus, the incident and reflected rays in the figure are symbolic only of the direction of propagation before and after reflection; they do not indicate any amount of travel. Upon reflection we have

$$y_2 = y_1 \tag{9.33}$$

since the ray has no chance to go anywhere.

We adopt the widely used convention that, upon reflection, the positive z-direction is reoriented so that we consider the rays still to travel in the positive z sense. Notice that in Fig. 9.7, the reflected ray approaches the z-axis. In this case θ_2 is a negative angle (as opposed to θ_1 which is drawn as a positive angle) and is equal to

$$\theta_2 = -(\theta_1 + 2\theta_i) \tag{9.34}$$

where θ_i is the angle of incidence with respect to the normal to the spherical mirror surface. By the law of reflection, the reflected ray also occurs at an angle θ_i referenced to the surface normal. The surface normal points towards the center of curvature, which we assume is on the z-axis a distance R away. By convention, the radius of curvature R is a positive number if the mirror surface is *concave* and a negative number if the mirror surface is *convex.*

We must eliminate θ_i from (9.34) in favor of θ_1 and y_1. By inspection of Fig. 9.7 we can write

$$\frac{y_1}{R} = \sin\phi \cong \phi \tag{9.35}$$

where we have applied the paraxial approximation (9.24). (Note that the angles in the figure are exaggerated.) We also have

$$\phi = \theta_1 + \theta_i \tag{9.36}$$

and when this is combined with (9.35), we get

$$\theta_i = \frac{y_1}{R} - \theta_1 \tag{9.37}$$

With this we are able to put (9.34) into a useful linear form:

$$\theta_2 = -\frac{2}{R}y_1 + \theta_1 \tag{9.38}$$

Equations (9.33) and (9.38) describe a linear transformation that can be concisely formulated as

$$\begin{bmatrix} y_2 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -2/R & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ \theta_1 \end{bmatrix} \quad \text{(concave mirror)} \tag{9.39}$$

The ABCD matrix in this transformation describes the act of reflection from a concave mirror with radius of curvature R. The radius R is negative when the mirror is convex.

The final basic element that we shall consider is a spherical interface between two materials with indices n_i and n_t (see Fig. 9.8). This has an effect similar to that of the curved mirror, which changes the direction of a ray without altering its distance y_1 from the optical axis. Please note that here the radius of curvature is considered to be positive for a convex surface (opposite convention from that of the mirror). Again, we are interested only in the act of transmission without any travel before or after the interface. As before, (9.33) applies (i.e. $y_2 = y_1$).

To connect θ_1 and θ_2 we must use Snell's law which in the paraxial approximations is

$$n_i\theta_i = n_t\theta_t \tag{9.40}$$

As seen in the Fig. 9.8, we have

$$\theta_i = \theta_1 + \phi \tag{9.41}$$

and

$$\theta_t = \theta_2 + \phi \tag{9.42}$$

As before, (9.35) applies (i.e. $\phi \cong y_1/R$). When this is used in (9.41) and (9.42), Snell's law (9.40) becomes

$$\theta_2 = \left(\frac{n_i}{n_t} - 1\right)\frac{y_1}{R} + \frac{n_i}{n_t}\theta_1 \tag{9.43}$$

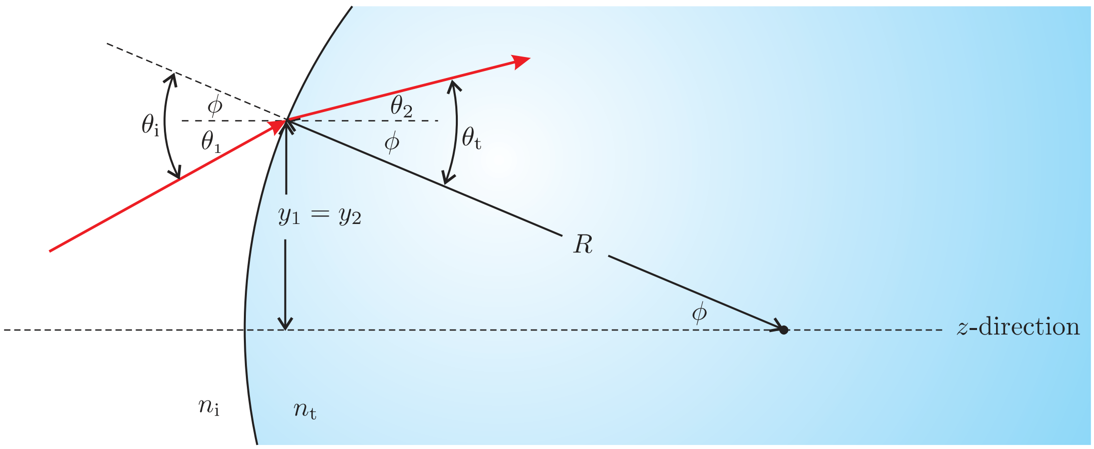

Figure 9.8 A ray depicted in the act of transmission at a curved material interface.

The compact matrix form of (9.33) and (9.43) turns out to be

$$\begin{bmatrix} y_2 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ (n_i/n_t - 1)/R & n_i/n_t \end{bmatrix} \begin{bmatrix} y_1 \\ \theta_1 \end{bmatrix} \text{ (from } n_i \text{ to } n_t\text{; interface radius } R) \tag{9.44}$$

In summary, we have developed three basic ABCD matrices seen in (9.29), (9.39), and (9.44). All other ABCD matrices that we will use are composites of these three. For example, one can construct the ABCD matrix for a lens by using two matrices like those in (9.44) to represent the entering and exiting surfaces of the lens. A distance matrix (9.29) can be inserted to account for the thickness of the lens. It is left as an exercise to derive the ABCD matrix for such a thick lens (see P 9.6).

The three ABCD matrices discussed can be used for many different composite systems. As another example, consider a ray that propagates through a distance a, followed by a reflection from a mirror of radius R, and then propagates through a distance b. This example is depicted in Fig. 9.9. The vector depicting the final ray in terms of the initial one is computed as follows:

$$\begin{aligned} \begin{bmatrix} y_2 \\ \theta_2 \end{bmatrix} &= \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -2/R & 1 \end{bmatrix} \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ \theta_1 \end{bmatrix} \\ &= \begin{bmatrix} 1 - 2b/R & a + b - 2ab/R \\ -2/R & 1 - 2a/R \end{bmatrix} \begin{bmatrix} y_1 \\ \theta_1 \end{bmatrix} \end{aligned} \tag{9.45}$$

The ordering of the matrices is important. The first effect that the light experiences is the matrix to the right, in the position that first operates on the vector representing the initial ray.

We have continually worked within the y–z plane as indicated in Figs. 9.6–9.9. This may have given the impression that it is necessary to work within that plane, or a plane containing the z-axis. However, within the paraxial approximation, our ABCD matrices are still valid for rays contained in planes that do not include the optical axis (as long as the rays are nearly parallel to the optical axis.

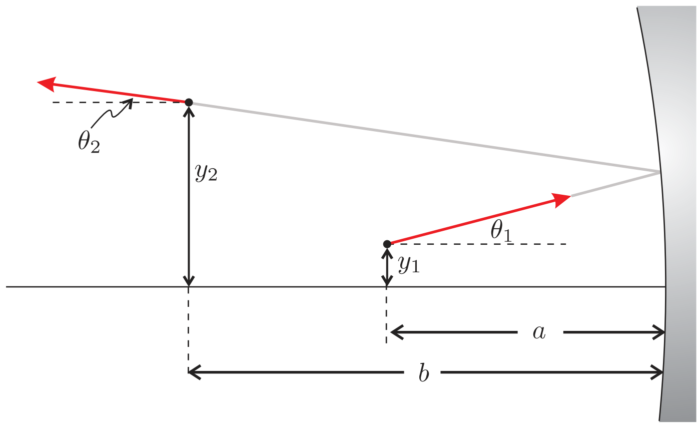

Figure 9.9 A ray that travels through a distance a, reflects from a mirror, and then travels through a distance b.

Imagine a ray contained within a plane that is parallel to the y–z plane but for which $x > 0$. One might be concerned that when the ray meets, for example, a spherically concave mirror, the radius of curvature *in the perspective of the y–z dimension* might be different for $x > 0$ than for $x = 0$ (at the center of the mirror). This concern is actually quite legitimate and is the source of what is known as spherical aberration. Nevertheless, in the paraxial approximation the intersection with the curved mirror of all planes that are parallel to the optical axis always give the same curvature.

To see why this is so, consider the curvature of the mirror in Fig. 9.7. As we move away from the mirror center (in either the x or y-dimension or some combination thereof), the mirror surface deviates to the left by the amount

$$\delta = R - R\cos\phi \tag{9.46}$$

In the paraxial approximation, we have $\cos\phi \cong 1 - \phi^2/2$. And since in this approximation we may also write $\phi \cong \sqrt{x^2 + y^2}\Big/ R$, (9.46) becomes

$$\delta \cong \frac{x^2 + y^2}{2R} \tag{9.47}$$

In the paraxial approximation, we see that the curve of the mirror is parabolic, and therefore separable between the x and y dimensions. That is, the curvature in the x-dimension (i.e. $\partial\delta/\partial x = x/R$) is independent of y, and the curvature in the y-dimension (i.e. $\partial\delta/\partial y = y/R$) is independent of x. A similar argument can be made for a spherical interface between two media within the paraxial approximation.

This allows us to deal conveniently with rays that have positioning and directional components in both the x and y dimensions. Each dimension can be treated separately without influencing the other. Most importantly, the identical matrices, (9.29), (9.39), and (9.44), are used for either dimension. Figs. 9.6–9.9 therefore represent projections of the actual rays onto the y–z plane. To complete the story, one would also need corresponding figures representing the projection of the rays onto the x–z plane.

9.6 Image Formation by Mirrors and Lenses

Consider the example shown in Fig. 9.9 where a ray travels through a distance a, reflects from a curved mirror, and then travels through a distance b. From (9.45) we know that the ABCD matrix for the overall process is

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1-2b/R & a+b-2ab/R \\ -2/R & 1-2a/R \end{bmatrix} \tag{9.48}$$

As is well known, it is possible to form an image with a concave mirror. Suppose that the initial ray is one of many which leaves a point on an object positioned at $d_o = a$ before the mirror. In order for an image to occur at $d_i = b$, it is essential that all rays leaving the original point on the object converge to a single point on the image. That is, we want rays leaving the point y_1 on the object (which may take on a range of angles θ_1) all to converge to a single point y_2 at the image. In the following equation we need y_2 to be independent of θ_1:

$$\begin{bmatrix} y_2 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} y_1 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} Ay_1 + B\theta_1 \\ Cy_1 + D\theta_1 \end{bmatrix} \tag{9.49}$$

The condition for image formation is therefore

$$B = 0 \quad \text{(condition for image formation)} \tag{9.50}$$

When this condition is applied to (9.48), we obtain

$$d_o + d_i - \frac{2d_o d_i}{R} = 0 \Rightarrow \frac{2}{R} = \frac{1}{d_o} + \frac{1}{d_i} \tag{9.51}$$

which is the familiar imaging formula for a mirror, in agreement with (9.1). When the object is infinitely far away (i.e. $d_o \to \infty$), the image appears at $d_i \to R/2$. This distance is called the *focal length* and is denoted by

$$f = \frac{R}{2} \quad \text{(focal length of a mirror)} \tag{9.52}$$

Please note that d_o and d_i can each be either positive (*real* as depicted in Fig. 9.9) or negative (*virtual* or behind the mirror).

The magnification of the image is found by comparing the size of y_2 to y_1. From (9.48)–(9.51), the magnification is found to be

$$M \equiv \frac{y_2}{y_1} = A = 1 - \frac{2d_i}{R} = -\frac{d_i}{d_o} \tag{9.53}$$

The negative sign indicates that for positive distances d_o and d_i the image is inverted.

Another common and very useful example is that of a thin lens, where we ignore the thickness between the two surfaces of the lens. Using the ABCD matrix in (9.44) twice, we find the overall matrix for the thin lens is

$$\begin{aligned} \begin{bmatrix} A & B \\ C & D \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ \frac{1}{R_2}(n-1) & n \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{1}{R_1}\left(\frac{1}{n}-1\right) & \frac{1}{n} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ -(n-1)\left(\frac{1}{R_1}-\frac{1}{R_2}\right) & 1 \end{bmatrix} \qquad \text{(Thin Lens)} \end{aligned} \tag{9.54}$$

Galileo Galilei

(1564–1642, Italian)

While Galileo did not invent the telescope, he was one of the few people of his time who knew how to build one. He also constructed a compound microscope. He attempted to measure the speed of light by having his assistant position himself on a distant hill and measuring the time it took for his assistant to uncover a lantern in response to a light signal. He was, of course, unable to determine the speed of light. His conclusion was that light is "really fast" if not instantaneous.

where we have taken the index outside of the lens to be unity while that of the lens material to be n. R_1 is the radius of curvature for the first surface which is positive if convex, and R_2 is the radius of curvature for the second surface which is also positive if convex *from the perspective of the rays which encounter it*.

Notice the close similarity between (9.54) and the matrix in (9.39). The ABCD matrix for either a thin lens or a mirror can be written as

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1/f & 1 \end{bmatrix} \tag{9.55}$$

where in the case of the thin lens the focal length is given by the lens maker's formula

$$\frac{1}{f} = (n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \quad \text{(focal length of thin lens)} \tag{9.56}$$

All of the arguments about image formation given above for the curved mirror work equally well for the thin lens. The only difference is that the focal length (9.56) is used in place of (9.52). That is, if we consider a ray traveling though a distance d_o impinging on a thin lens whose matrix is given by (9.55), and then afterwards traveling a distance d_i, the overall ABCD matrix is exactly like that in (9.48):

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 - d_\text{i}/f & d_\text{o} + d_\text{i} - d_\text{o}d_\text{i}/f \\ -1/f & 1 - d_\text{o}/f \end{bmatrix} \tag{9.57}$$

When we use the imaging condition (9.50), the imaging formula (9.1) emerges naturally.

9.7 Image Formation by Complex Optical Systems

A complicated series of optical elements (e.g. a sequence of lenses and spaces) can be combined to form a composite imaging system. The matrices for each of the elements are multiplied together (the first element that rays encounter appearing on the right) to form

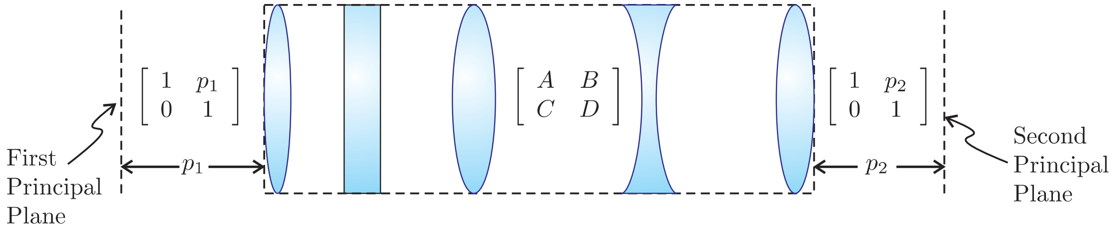

Figure 9.10 A multi-element system represented as an ABCD matrix for which principal planes always exist.

the overall composite ABCD matrix. We can study the imaging properties of a composite ABCD matrix by combining the matrix with the matrices for the distances from an object to the system and from the system to the image formed:

$$\begin{aligned} \begin{bmatrix} 1 & d_\mathrm{i} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} 1 & d_\mathrm{o} \\ 0 & 1 \end{bmatrix} &= \begin{bmatrix} A + d_\mathrm{i}C & d_\mathrm{o}A + B + d_\mathrm{o}d_\mathrm{i}C + d_\mathrm{i}D \\ C & d_\mathrm{o}C + D \end{bmatrix} \\ &= \begin{bmatrix} A' & B' \\ C' & D' \end{bmatrix} \end{aligned} \tag{9.58}$$

Imaging occurs according to (9.50) when $B' = 0$, or

$$d_\mathrm{o}A + B + d_\mathrm{o}d_\mathrm{i}C + d_\mathrm{i}D = 0, \quad \text{(general condition for image formation)} \tag{9.59}$$

with magnification

$$M = A + d_\mathrm{i}C \tag{9.60}$$

There is a convenient way to simplify this analysis.

For every ABCD matrix representing a (potentially) complicated optical system, there exist two principal planes located (in our convention) a distance p_1 before entering the system and a distance p_2 after exiting the system. When the matrices corresponding to the (appropriately chosen) distances to those planes are appended to the original ABCD matrix of the system, the overall matrix simplifies to one that looks like the matrix for a simple thin lens (9.55). With knowledge of the positions of the principal planes, one can treat the complicated imaging system in the same way that one treats a simple thin lens. The only difference is that d_o is the distance from the object to the first principal plane and d_i is the distance from the second principal plane to the image. (In the case of an actual thin lens, both principal planes are at $p_1 = p_2 = 0$. For a composite system, p_1 and p_2 can be either positive or negative.)

Next we demonstrate that p_1 and p_2 can always be selected such that we can write

$$\begin{aligned} \begin{bmatrix} 1 & p_2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} 1 & p_1 \\ 0 & 1 \end{bmatrix} &= \begin{bmatrix} A + p_2C & p_1A + B + p_1p_2C + p_2D \\ C & p_1C + D \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ -1/f_\mathrm{eff} & 1 \end{bmatrix} \end{aligned} \tag{9.61}$$

The final matrix is that of a simple thin lens, and it takes the place of the composite system including the distances to the principal planes. Our task is to find the values of p_1 and

p_2 that make this matrix replacement work. We must also prove that this replacement is always possible for physically realistic values for A, B, C, and D.

We can straightaway make the definition

$$f_{\text{eff}} \equiv -1/C \tag{9.62}$$

We can also solve for p_1 and p_2 by setting the diagonal elements of the matrix to 1. Explicitly, we get

$$p_1 C + D = 1 \Rightarrow p_1 = \frac{1-D}{C} \tag{9.63}$$

and

$$A + p_2 C = 1 \Rightarrow p_2 = \frac{1-A}{C} \tag{9.64}$$

It remains to be shown that the upper right element in (9.61) (i.e. $p_1 A + B + p_1 p_2 C + p_2 D$) automatically goes to zero for our choices of p_1 and p_2. This may seem unlikely at first, but we can invoke an important symmetry in the matrix to show that it does in fact vanish for our choices of p_1 and p_2.

When (9.63) and (9.64) are substituted into the upper right matrix element of (9.61) we get

$$\begin{aligned} p_1 A + B + p_1 p_2 C + p_2 D &= \frac{1-D}{C}A + B + \frac{1-D}{C}\frac{1-A}{C}C + \frac{1-A}{C}D \\ &= \frac{1}{C}\left[1 - AD + BC\right] \\ &= \frac{1}{C}\left(1 - \begin{vmatrix} A & B \\ C & D \end{vmatrix}\right) \end{aligned} \tag{9.65}$$

This equation shows that the upper right element of (9.61) vanishes when the determinant of the original ABCD matrix equals one. Fortunately, this is always the case *as long as we begin and end in the same index of refraction*. Therefore, we have

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = 1 \tag{9.66}$$

Notice that the determinants of the matrices in (9.29), (9.39), and (9.55) are all one, and so ABCD matrices constructed of these will also have determinants equal to one. The determinant of (9.44) is not one. This is because it begins and ends in different indices, but when this matrix is used in succession to form a lens or even a strange conglomerate of successive material interfaces, the resulting matrix will have a determinant equal to one as long as the beginning and ending indices are the same. Table 9.1 is a summary of ABCD matrices of common optical elements. All of the matrices obey (9.66).

9.8 Stability of Laser Cavities

As a final example of the usefulness of paraxial ray theory, we apply the ABCD matrix formulation to a laser cavity. The basic elements of a laser cavity include an amplifying medium and mirrors to provide feedback. Presumably, at least one of the end mirrors is

$\begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix}$	(Distance within any material, excluding interfaces)
$\begin{bmatrix} 1 & d/n \\ 0 & 1 \end{bmatrix}$	(Window, starting and stopping in air)
$\begin{bmatrix} 1 & 0 \\ -1/f & 1 \end{bmatrix}$	(Thin lens or a mirror with $f = R/2$)
$\begin{bmatrix} 1 + \frac{d}{R_1}\left(\frac{1}{n} - 1\right) & \frac{d}{n} \\ (1-n)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) + \frac{d}{R_1 R_2}\left(2 - \frac{1}{n} - n\right) & 1 - \frac{d}{R_2}\left(\frac{1}{n} - 1\right) \end{bmatrix}$	(Thick lens)

Table 9.1 Summary of ABCD matrices for common optical elements.

partially transmitting so that energy is continuously extracted from the cavity. Here, we dispense with the amplifying medium and concentrate our attention on the optics providing the feedback. As might be expected, the mirrors must be carefully aligned or successive reflections might cause rays to “walk” continuously away from the optical axis, so that they eventually leave the cavity out the side. If a simple cavity is formed with two flat mirrors that are perfectly aligned parallel to each other, one might suppose that the mirrors would provide ideal feedback. However, all rays except for those that are perfectly aligned to the mirror surface normals eventually wander out of the side of the cavity as illustrated in Fig. 9.11a. Such a cavity is said to be *unstable*. We would like to do a better job of trapping the light in the cavity.

To improve the situation, a cavity can be constructed with concave end mirrors to help confine the beams within the cavity. Even so, one must choose carefully the curvature of the mirrors and their separation L. If this is not done correctly, the curved mirrors can “overcompensate” for the tendency of the rays to wander out of the cavity and thus aggravate the problem. Such an unstable scenario is depicted in Fig. 9.11b.

Figure 9.11c depicts a cavity made with curved mirrors where the separation L is chosen appropriately to make the cavity stable. Although a ray, as it makes successive bounces, can strike the end mirrors at a variety of points, the curvature of the mirrors keeps the “trajectories” contained within a narrow region so that they cannot escape out the sides of the cavity.

There are many ways to make a stable laser cavity. For example, a stable cavity can be made using a lens between two flat end mirrors as shown in Fig. 9.11d. Any combination of lenses (perhaps more than one) and curved mirrors can be used to create stable cavity configurations. *Ring cavities* can also be made to be stable where in no place do the rays retro-reflect from a mirror but circulate through a series of elements like cars going around a racetrack.

We now find the conditions that have to be met in order for a cavity to be stable. The

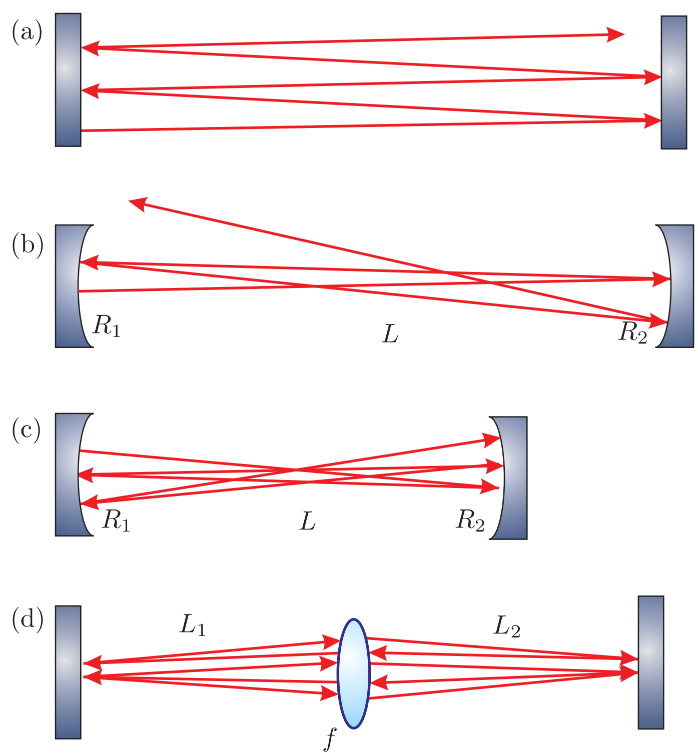

Figure 9.11 (a) A ray bouncing between two parallel flat mirrors. (b) A ray bouncing between two curved mirrors in an unstable configuration. (c) A ray bouncing between two curved mirrors in a stable configuration. (d) Stable cavity utilizing a lens and two flat end mirrors.

ABCD matrix for a round trip in the cavity is useful for this analysis. For example, the round-trip ABCD matrix for the cavity shown in Fig. 9.11c is

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & L \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -2/R_2 & 1 \end{bmatrix} \begin{bmatrix} 1 & L \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -2/R_1 & 1 \end{bmatrix} \tag{9.67}$$

where we have begun the round trip just after a reflection from the first mirror. The round-trip ABCD matrix for the cavity shown in Fig. 9.11d is

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & 2L_1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1/f & 1 \end{bmatrix} \begin{bmatrix} 1 & 2L_2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1/f & 1 \end{bmatrix} \tag{9.68}$$

where we have begun the round trip just after a transmission through the lens moving to the right. It is somewhat arbitrary where the round trip begins.

To determine whether a given configuration of a cavity will be stable, we need to know what a ray does after making many round trips in the cavity. To find the effect of propagation through many round trips, we multiply the round-trip ABCD matrix together N times, where N is the number of round trips that we wish to consider. We can then examine what happens to an arbitrary ray after making N round trips in the cavity as follows:

$$\begin{bmatrix} y_{N+1} \\ \theta_{N+1} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}^N \begin{bmatrix} y_1 \\ \theta_1 \end{bmatrix} \tag{9.69}$$

At this point students might be concerned that taking an ABCD matrix to the N^{th} power can be a lot of work. (It is already a significant amount of work just to compute the ABCD matrix for a single round trip.) In addition, we are interested in letting N be very large, perhaps even infinity. Students can relax because we have a neat trick to accomplish this daunting task.

We use Sylvester's theorem from appendix 0.4, which states that if

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = 1 \tag{9.70}$$

then

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^N = \frac{1}{\sin\theta} \begin{bmatrix} A\sin N\theta - \sin(N-1)\theta & B\sin N\theta \\ C\sin N\theta & D\sin N\theta - \sin(N-1)\theta \end{bmatrix} \tag{9.71}$$

where

$$\cos\theta = \frac{1}{2}(A+D). \tag{9.72}$$

As we have already discussed, (9.70) is satisfied if the refractive index is the same before and after, which is guaranteed for any round trip. We therefore can employ Sylvester's theorem for any N that we might choose, including very large integers.

We would like the elements of (9.71) to remain finite as N becomes very large. If this is the case, then we know that a ray remains trapped within the cavity and stays reasonably close to the optical axis. Since N only appears within the argument of a sine function, which is always bounded between -1 and 1 for real arguments, it might seem that the elements of (9.71) always remain finite as N approaches infinity. However, it turns out that θ can become imaginary depending on the outcome of (9.72), in which case the sine becomes a hyperbolic sine, which can "blow up" as N becomes large. In the end, the condition for cavity stability is that a real θ must exist for (9.72), or in other words we need

$$-1 < \frac{1}{2}(A + D) < 1 \qquad \text{(condition for a stable cavity)} \tag{9.73}$$

It is left as an exercise to apply this condition to (9.67) and (9.68) to find the necessary relationships between the various element curvatures and spacing in order to achieve cavity stability.

9.9 Aberrations and Ray Tracing

The paraxial approximation places serious limitations on the performance of optical systems (see (9.24) and (9.25)). To stay within the approximation, all rays traveling in the system should travel very close to the optic axis with very shallow angles with respect to the optical axis. To the extent that this is not the case, the collection of rays associated with a single point on an object may not converge to a single point on the associated image. The resulting distortion or "blurring" of the image is known as *aberration*.

Common experience with photographic and video equipment suggests that it is possible to image scenes that have a relatively wide angular extent (many tens of degrees), in apparent serious violation of the paraxial approximation. The paraxial approximation is indeed violated in these devices, so they must be designed using more complicated analysis techniques than those we have learned in this chapter. The most common approach is to use a computationally intensive procedure called *ray tracing* in which $\sin\theta$ and $\tan\theta$ are rendered exactly. The nonlinearity of these functions precludes the possibility of obtaining analytic solutions describing the imaging performance of such optical systems.

The typical procedure is to start with a collection of rays from a test point such as shown in Fig. 9.12. Each ray is individually traced through the system using the exact representation of geometric surfaces as well as the exact representation of Snell's law. On close analysis, the rays typically do not converge to a distinct imaging point. Rather, the rays can be "blurred" out over a range of points where the image is supposed to occur. Depending on the angular distribution of the rays as well as on the elements in the setup, the spread of rays around the image point can be large or small. The engineer who designs the

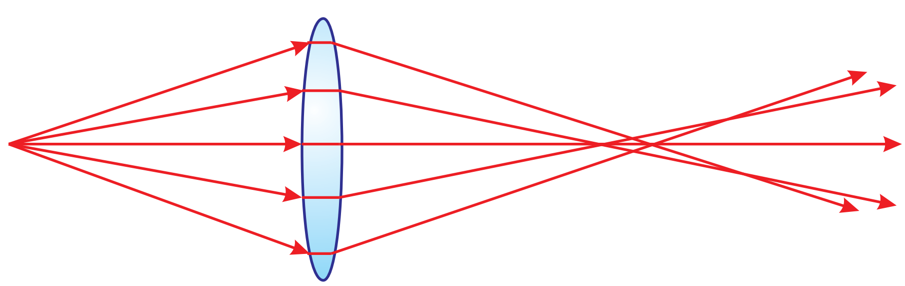

Figure 9.12 Ray tracing through a simple lens.