EEAP 349

GEOMETRIC & GAUSSIAN OPTICS NOTES

FRANK MERAT

Geometric Optics

I. Introduction

Geometric optics is that field of optics devoted to the analysis of the transformation of light rays by optical elements such as lenses or mirrors. This definition requires that a light ray be defined. The following is how an engineer would define a light ray. Let L represent a very small light source (i.e., a point source). Suppose this illuminates a screen O through a hole in an intermediate screen S. The result will be an image of L on O as shown below.

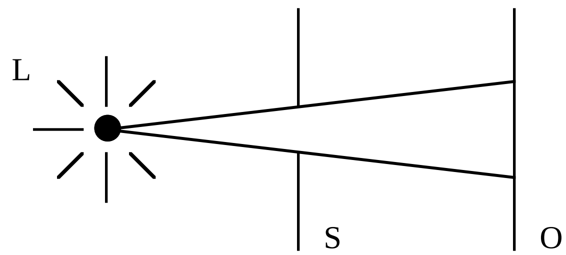


Figure 1.

Note that light is drawn as if it traveled along straight line paths (This will be proven in the geometric optics limit later in these notes). Only a portion of O is illuminated by L. As the hole in S shrinks in diameter the illuminated portion of O also shrinks. This process may be continued until the hole is on the order of 0,1 to 0,3 mm in diameter (for holes of smaller size diffraction effects come into play and the spot size on O begins to increase). Thus, we may image a very narrow ray of light (ray diameter on the order of 0.1-0.3 mm) traveling in a straight line from L to O. Because of this straight line path property the behavior of light can be analyzed geometrically, hence, geometric optics. In most optical instruments the beams of light are fairly wide which allows a geometric optics analysis and, consequently, makes geometric optics a very important area of optics.

Before proceeding to develop the principles of geometric optics it is interesting to show that light does travel in a straight line and that this is consistent with the wave equation and Maxwell's equations.

Starting with the wave equation

$$\nabla^2\phi = \frac{n^2}{c^2}\frac{\partial^2\phi}{\partial t^2} \tag{1}$$

We can now assume a solution of the form

$$\phi(\bar{r},t) = A(\bar{r})e^{ik\left(L(\bar{r})-ct\right)} \tag{2}$$

where $L(\bar{r})$ is the optical path traveled by the light from some reference source (usually the origin of the coordinate system) and may be rigorously defined as

$$L(\bar{r}) = \int_{P_1}^{P_2} n(s)ds$$

where P_1 is the starting point of the ray, P_2 is the final point of the ray, and n(s) is the index of refraction integrated along the ray path from P_1 to P_2, i.e., a line integral. In general a ray refers to a normal to a propagating wave front which is a surface. Using (2) we evaluate

$$\nabla^2\phi = \left\{\nabla\cdot\nabla\left(e^{a+ikL}\right)\right\}e^{-ikct}$$

where we have written $A(\bar{r}) = e^{a(\bar{r})}$ for notational convenience. Continuing,

$$\nabla^2\phi = \left\{\nabla\cdot\left[e^{a+ikL}\nabla(a+ikL)\right]\right\}e^{-ikct} = \left\{\left[\nabla(a+ikL)\right]^2 + \nabla^2(a+ikL)\right\}e^{a+ikL-ikct}$$

$$\nabla^2\phi = \left\{\left[\nabla(a+ikL)\right]^2 + \nabla^2(a+ikL)\right\}\phi \qquad (3)$$

Likewise, evaluating $\frac{n^2}{c^2}\frac{\partial^2\phi}{\partial t^2}$ using (2)

$$\frac{n^2}{c^2}\frac{\partial^2\phi}{\partial t^2} = \frac{n^2}{c^2}\left(-ikc^2\right)\phi \qquad (4)$$

Substituting (3) and (4) into (1) we obtain

$$\left((\nabla a)^2 + 2ik\nabla a\cdot\nabla L - k^2(\nabla L)^2 + \nabla^2 a + ik\nabla^2 L\right)\phi = -n^2k^2\phi$$

Equating the real parts

$$\left((\nabla a)^2 - k^2(\nabla L)^2 + \nabla^2 a\right)\phi = -n^2k^2\phi$$

$$\left(k^2(\nabla L)^2 - n^2k^2\right)\phi = \left[(\nabla a)^2 + \nabla^2 a\right]\phi$$

$$\left[(\nabla L)^2 - n^2\right]\phi = \frac{\lambda^2}{4\pi^2}\left[(\nabla a)^2 + \nabla^2 a\right]\phi$$

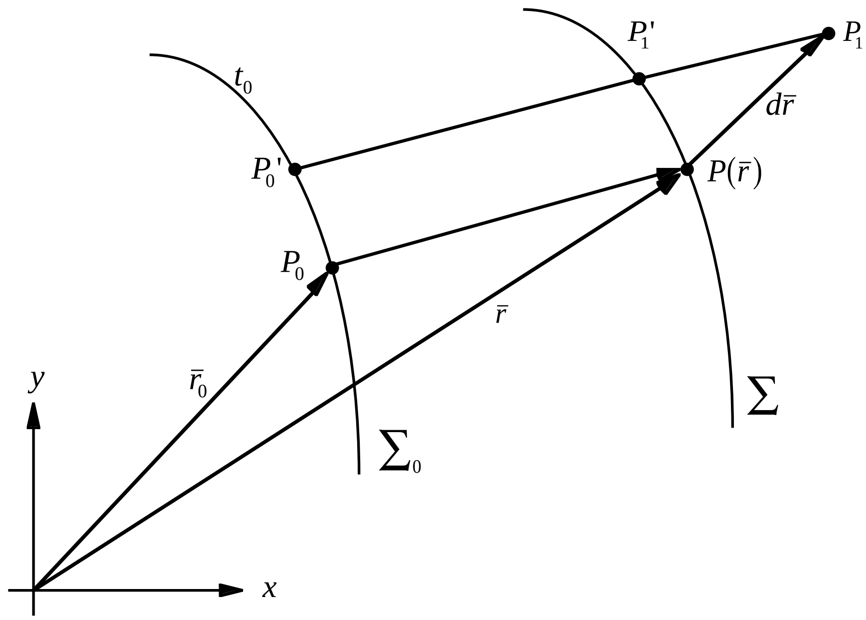
where we have used $k = \frac{2\pi}{\lambda}$. Taking the limit as $\lambda \to 0$ we get the result that $(\nabla L)^2 - n^2 = 0$ or

$$(\nabla L)^2 = n^2 \qquad (5)$$

This result is known as the eikonal equation and implies that light travels in straight lines, at least in the zero wavelength limit. Taking the square root of (5)

$$\nabla L = n\bar{\tau}$$

where $\bar{\tau}$ will be shown to be the unit tangent to $L(\bar{r})$.



$$L(\bar{r}) = \int_{\bar{r}_0}^{\bar{r}} n(s)ds$$

$$\phi(\bar{r},t) = A(\bar{r})e^{jk(L(\hat{r})-ct)}$$

Figure 2.

Suppose many rays leave surface $\sum_0$ at time t_0. The surface $\sum$ consists of all points reached by rays starting from $\sum_0$ in a time interval $(t-t_0)$. Thus, for any ray leaving $\sum$, $L(\bar{r}) = c(t-t_0)$ - by definition the optical distance between $\sum_0$ and $\sum$. Pick any point P_1 beyond the surface $\sum$. The optical path from P_0' to P_1' is $L(\bar{r})$ by definition of $\sum$. The optical distance from P_1' to P_1 is $dL = n|P_1'P_1| = nd\bar{r}\cdot\bar{\tau}$ where $\bar{\tau}$ is tangent to the ray P_0P at $\bar{r}$. (It is assumed that for $|d\bar{r}|$ small, $P_1'P$ will be parallel to P_0P at the surface $\sum$.)

$$dL = nd\bar{r}\cdot\bar{\tau} \tag{6}$$

Note now that , in general, $dL = \frac{\partial L}{\partial x}dx + \frac{\partial L}{\partial y}dy + \frac{\partial L}{\partial z}dz$ or

$$dL = \nabla L\cdot d\bar{r} \tag{7}$$

Equating (6) and (7) we again get the eikonal equation

$$\nabla L\cdot d\bar{r} = n\bar{\tau}\cdot d\bar{r}$$

$$\nabla L = n\bar{\tau} \tag{8}$$

The derivative of any function f along some curve is $\frac{df}{ds} = \nabla f\cdot\bar{a}$ where $\bar{a}$ is the unit vector tangent to the curve and s is the displacement along the curve. Let $\bar{\tau}$ be the unit tangent as defined by (8), then

$$\frac{df}{ds} = \nabla f\cdot\bar{\tau} = \nabla f\cdot\frac{\nabla L}{n} = \frac{1}{n}(\nabla L\cdot\nabla f) \tag{9}$$

Now pick $f = \frac{\partial L}{\partial x}$. The reason for this will become clear shortly.

$$\frac{d}{ds}\left(\frac{\partial L}{\partial x}\right) = \frac{1}{n}\nabla L \cdot \nabla\left(\frac{\partial L}{\partial x}\right) = \frac{1}{n}\left\{\frac{\partial L}{\partial x}\cdot\frac{\partial}{\partial x}\frac{\partial L}{\partial x} + \frac{\partial L}{\partial y}\cdot\frac{\partial}{\partial y}\frac{\partial L}{\partial x} + \frac{\partial L}{\partial z}\cdot\frac{\partial}{\partial z}\frac{\partial L}{\partial x}\right\}$$

$$\frac{d}{ds}\left(\frac{\partial L}{\partial x}\right) = \frac{1}{n}\left\{\frac{\partial L}{\partial x}\frac{\partial^2 L}{\partial x^2} + \frac{\partial L}{\partial y}\frac{\partial}{\partial x}\frac{\partial L}{\partial y} + \frac{\partial L}{\partial z}\frac{\partial}{\partial x}\frac{\partial L}{\partial z}\right\}$$

Rearranging,

$$\frac{d}{ds}\left(\frac{\partial L}{\partial x}\right) = \frac{1}{2n}\frac{\partial}{\partial x}\left\{\left(\frac{\partial L}{\partial x}\right)^2 + \left(\frac{\partial L}{\partial y}\right)^2 + \left(\frac{\partial L}{\partial z}\right)^2\right\}$$

$$\frac{d}{ds}\left(\frac{\partial L}{\partial x}\right) = \frac{1}{2n}\frac{\partial}{\partial x}\left|\nabla L\right|^2 = \frac{1}{2n}\frac{\partial}{\partial x}n^2 = \frac{\partial n}{\partial x}$$

since $\left|\nabla L\right|^2 = n^2$. This manipulation may be repeated for $f = \frac{\partial L}{\partial y}$ and $f = \frac{\partial L}{\partial z}$ and the results summed to give

$$\frac{d}{ds}(\nabla L) = \nabla n = \frac{d}{ds}(n\bar{\tau}) \tag{10}$$

If the medium is homogeneous this reduces to $\frac{d\bar{\tau}}{ds} = 0$, i.e., $\bar{\tau}$=constant (since homogeneity is equivalent to n=constant). If the tangent vector to a curve is constant that that curve must be a straight line; this, in the $\lambda = 0$ limit light rays travel in straight lines in homogeneous media.

This result together with Fermat's principle permits one to derive the fundamental properties of reflection and refraction. Fermat's principle is that a ray of light will traverse a medium in such a way that the total optical path assumes an extreme value. Stated mathematically,

$$\delta L = \delta\left\{\int_{P_1}^{P_2} n(s)ds\right\} = 0$$

where the δ operation indicates a variation in the following quantity and ds is the differential distance along the ray path connecting P_1 to P_2.

To illustrate how this principle may be used consider the case of a light ray from P_1 being reflected off a surface S to a second point P_2 as shown below.

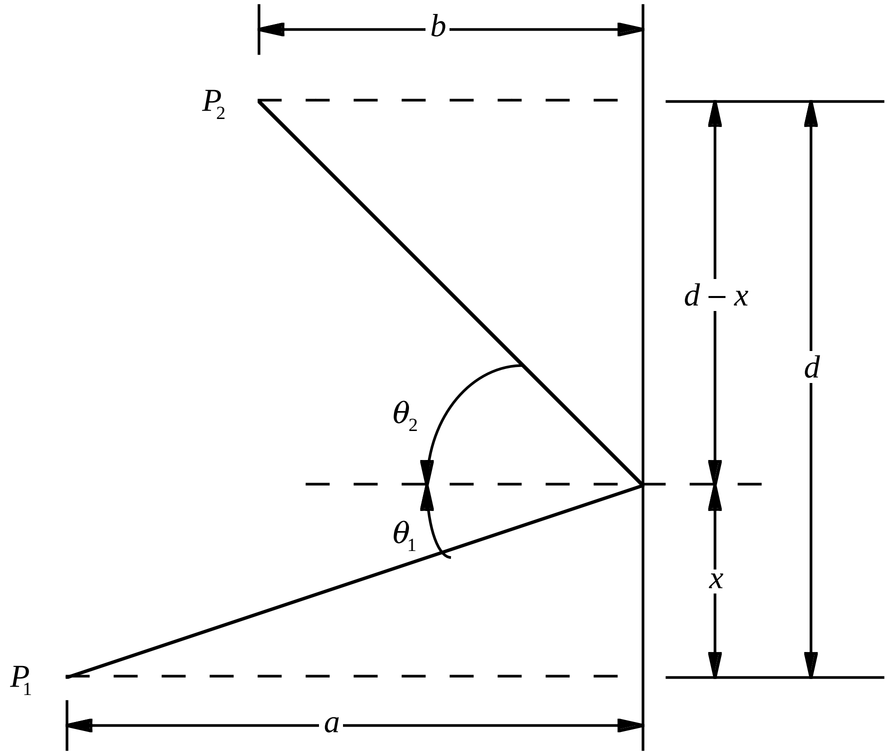


Figure 3.

The optical path L is given by

$$L = n\left(\sqrt{a^2+x^2} + \sqrt{b^2+(d-x)^2}\right)$$

Differentiating,

$$dL = n\left\{\frac{1}{2}\frac{1}{\sqrt{a^2+x^2}}2x + \frac{1}{2}\frac{1}{\sqrt{b^2+(d-x)^2}}2(d-x)(-1)\right\}dx = 0$$

Note that for our purposes the variational operator $\delta(\ )$ is equivalent to the differential operator $d(\ )$. Neglecting the trivial solution $dx = 0$ we must have

$$\frac{x}{\sqrt{a^2+x^2}} = \frac{d-x}{\sqrt{b^2+(d-x)^2}}$$

From the figure $\sin\theta_1 = \dfrac{x}{\sqrt{a^2+x^2}}$ and $\sin\theta_2 = \dfrac{d-x}{\sqrt{b^2+(d-x)^2}}$ where θ_1 and θ_2 are measured with respect to a normal to the reflecting surface. Thus, $\sin\theta_1 = \sin\theta_2$ or, the angle if incidence equals the angle of refraction.

It is a very similar problem to derive Snell's Law governing refraction. The geometry is as given below.

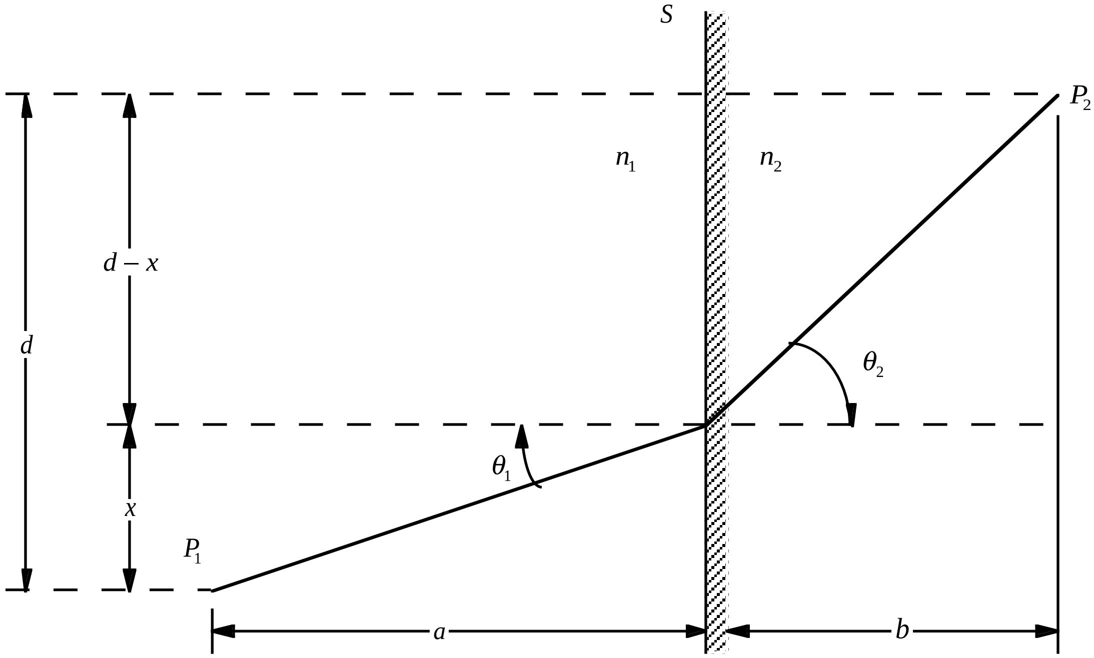


Figure 4.

As before construct the optical path function L.

$$L = n_1\sqrt{a^2 + x^2} + n_2\sqrt{b^2 + (d-x)^2}$$

Differentiating,

$$dL = n_1\frac{1}{\sqrt{a^2 + x^2}} + n_2\frac{(d-x)}{\sqrt{b^2 + (d-x)^2}} = 0$$

by Fermat's Principle. This requires that

$$n_1\frac{1}{\sqrt{a^2 + x^2}} = n_2\frac{(d-x)}{\sqrt{b^2 + (d-x)^2}}$$

or $n_1 \sin\theta_1 = n_2 \sin\theta_2$ (where θ_1 and θ_2 are measured with respect to the surface S) which is Snell's Law.

In general, Snell's Law and the law of refraction are valid for arbitrarily curved surfaces as long as the angles are measured with respect to a perpendicular to the tangent to the curved surface.

II. Optical Transformations and the Ray Matrix

In general the effect of an optical element such as a reflecting or refracting surface upon an incident ray can be modeled as a transformation of the incident ray's slope and displacement with respect to the optic axis. Specifically,

$$n_2 r_2' = A n_1 r_1' + B r_1$$
$$r_2 = C n_1 r_1' + D r_1$$

The unprimed letters represent the ray's displacement from the optic axis (usually the axis of symmetry for optical systems), primed letters the ray's slope with respect to the optic axis, the subscript “1” the incident ray, the subscript “2” the reflected or transmitted ray depending upon the particular optical element, n_1 the index of refraction of the medium the incident ray is traveling in, and n_2 the index of refraction of the medium the transmitted or reflected ray is traveling in.

The quantities A, B, C and D describe the effect of the optical element upon the incident ray. The transformation may be conveniently written in matrix form as

$$\begin{bmatrix} n_2 r_2' \\ r_2 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} n_1 r_1' \\ r_1 \end{bmatrix}$$

The square matrix containing A, B, C and D is commonly referred to as the ray matrix and will be shown to be capable of describing many optical elements.

To demonstrate the derivation of the elements of the ray matrix we will consider the propagation of a ray in the direction from the $z=z_1$ to the $z=z_2=z_1+d$ planes as shown below.

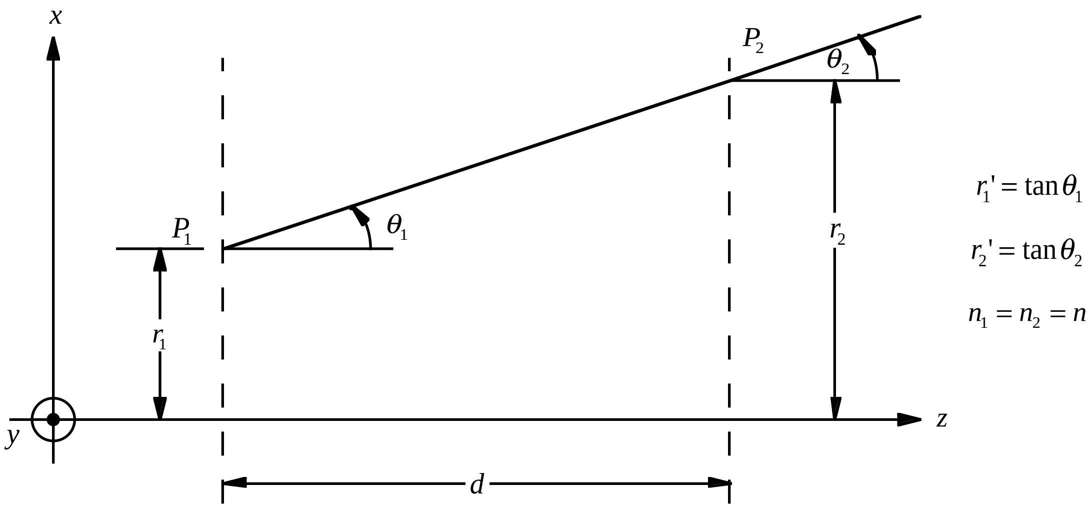


Figure 5.

At point P_1 the ray has displacement r_1 from the optic axis and slope r_1’ with respect to the optic axis. Because the ray propagates along a straight line the slope r_2’ will be the same as r_1’; however, the displacement r_2 at the z_2 plane will not be r_1 but, rather, r_1+r_1’d. Thus, a length d of open space may be characterized by the set of transformations

$$nr_2' = nr_1'$$

$$r_2 = \frac{d}{n}\left(nr_1'\right) + r_1$$

or in the equivalent matrix form

$$\begin{bmatrix} n_2 r_2' \\ r_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{d}{n} & 1 \end{bmatrix} \begin{bmatrix} n_1 r_1' \\ r_1 \end{bmatrix}$$

(The quantity d/n is often referred to as the reduced distance.) This result defines the basic ray matrix describing the translation of a ray through space; hence, we define the basic translation ray matrix T to be

$$T = \begin{bmatrix} 1 & 0 \\ \frac{d}{n} & 1 \end{bmatrix}$$

Before developing any more ray matrices it will be necessary to state some sign conventions commonly used with ray matrices:

1. light proceeds from left to right unless otherwise indicated;
2. distances measured in the direction light is traveling are positive;
3. a distance is always measured from a refracting surface or a principle plane (to be defined later);
4. a radius of curvature is positive if the direction from the vertex of a surface to the center of curvature is from left to right (Note: the vertex of a curved refracting surface is its intersection point with the optic axis.);
5. surfaces are numbered in the order in which light passes through them; and
6. a reflecting surface requires the use of a negative index of refraction for the medium following the surface to account for the change in the direction of the ray or the "folding" of the optical system as it is sometimes called.

We will now derive the ray matrix for a mirror using the paraxial ray approximation that the displacement of the ray from the optic axis and its slope with respect to the optic axis is small. This assumption is not as restrictive as it sounds and many interesting optical systems can be examined using the paraxial ray approximation. Returning to our derivation we will consider the mirror diagrammed below.

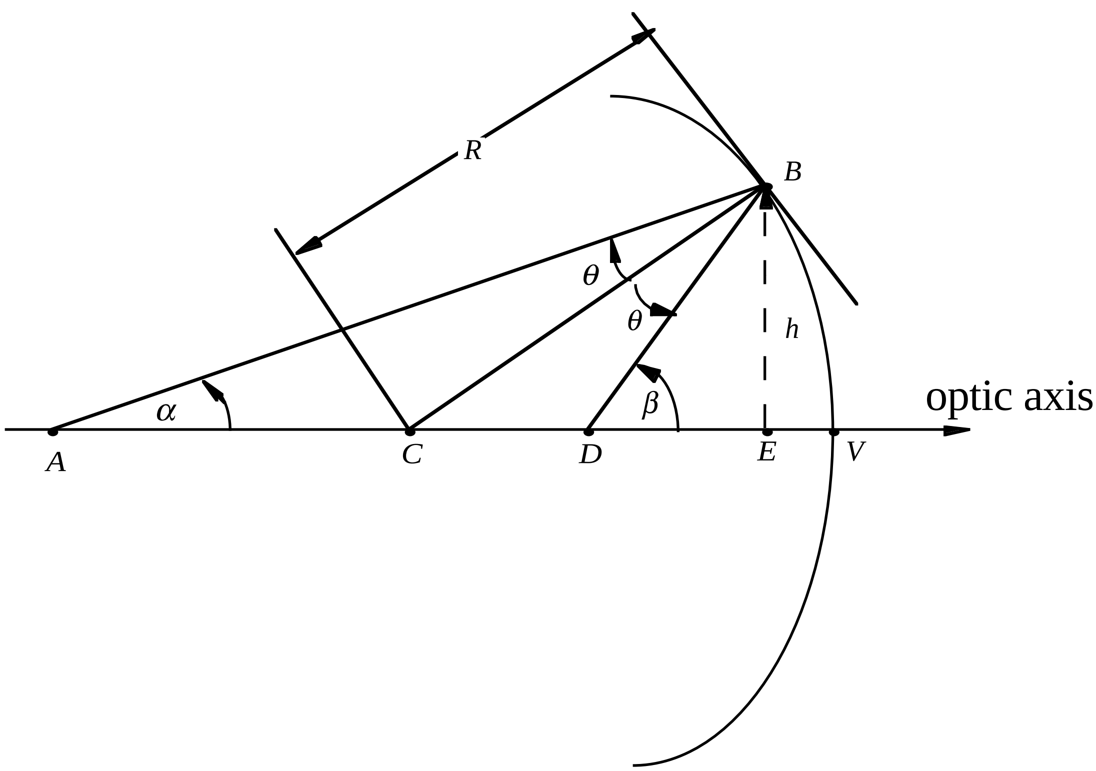


Figure 6.

By sign convention #4 the distance from V to C is from right to left indicating that the mirror curvature is negative. For triangle CBE we then have

$$\sin(\beta - \theta) = \frac{h}{R}$$

By the paraxial ray approximation this reduces to $\beta - \theta \approx \frac{h}{R}$ or

$$\theta \approx \beta - \frac{h}{R}$$

There are many ways to do the geometry. Here will will use the theorem from geometry that the exterior angle of a triangle is equal to the sum of the two interior angles to write $\beta = \alpha + 2\theta$. Using our previous expression for θ in this result and solving for β we get

$$\beta = \alpha + 2\left(\beta - \frac{h}{R}\right)$$

$$\beta = 2\frac{h}{R} - \alpha$$

By our adopted sign convention $r_1' = +\frac{h}{\overline{AE}}$ and $r_2' = -\frac{h}{\overline{DE}}$ so that

$$r_1' = \tan\alpha \approx \alpha$$
$$r_2' = -\tan\beta \approx -\beta$$

where we have agai n invoked the paraxial ray approximation. Substituting these results into the expression for β we get

$$-r_2' = -r_1' + 2\frac{h}{R}$$

Noting that $r_2 = r_1 = h$ we may multiply through by n (the index of refraction) to obtain

$$nr_2' = nr_1' - \frac{2n}{R}r$$

Thus, the ray matrix R^- is

$$R^- = \begin{bmatrix} 1 & -\frac{2n}{R} \\ 0 & 1 \end{bmatrix}$$

where R^- is the ray matrix for a mirror of curvature R<0. The more general form for R is

$$R = \begin{bmatrix} 1 & \frac{2n}{R} \\ 0 & 1 \end{bmatrix}$$

where R is a signed quantity.

Let us now develop the ray matrix for the curved dielectric interface illustrated below. Note that this system has a positive radius of curvature according to our adopted convention.

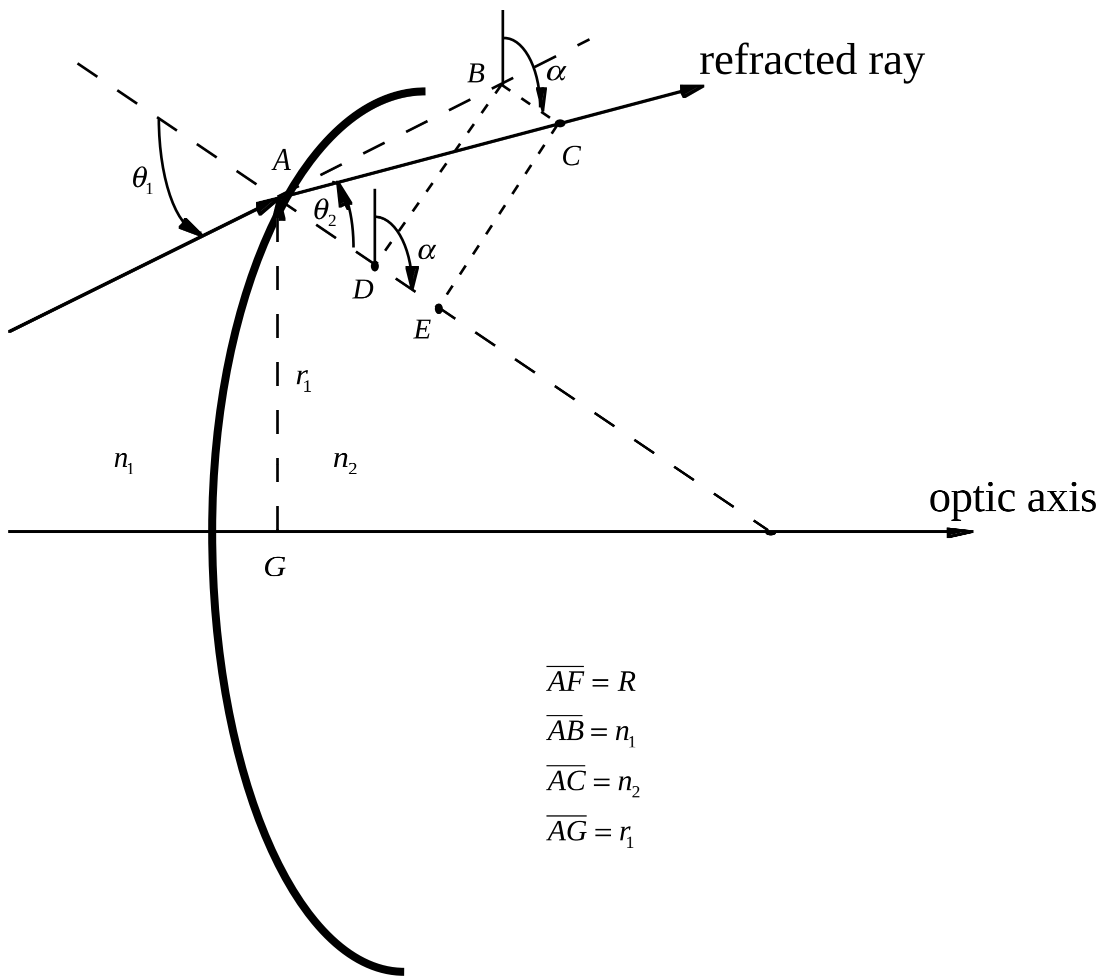


Figure 7

Construct AB along the path of OA. The distance $\overline{AB}$ may be drawn proportional to n_1 and, for our purposes, we may take $\overline{AB} = n_1$. In like fashion locate C on the refracted ray AC such that $\overline{AC} = n_2$. Note that $\overline{BD} = \overline{AB}\sin\theta_1 = n_1 \sin\theta_1$ and that $\overline{CE} = \overline{AC}\sin\theta_2 = n_2 \sin\theta_2$. But $n_1 \sin\theta_1 = n_2 \sin\theta_2$ by Snell's Law which means that $\overline{BD} = \overline{CE}$ and, therefore, that $BC \parallel DE$. Let α be as indicated in Figure 7 and consider triangle AFG. It follows that $\cos(\pi - \alpha) = \frac{r_1}{R} = -\cos(\alpha)$ or $\cos(\alpha) = -\frac{r_1}{R}$. Let us now examine triangle ABC in greater detail (See Figure 8).

By definition, $r_1' = \tan\beta_1$ and $r_2' = \tan\beta_2$. Re-writing using $\tan(x) = \frac{\sin(x)}{\cos(x)}$

$$r_1' \cos\beta_1 = \sin\beta_1 = \frac{\overline{A'B'}}{n_1}$$

$$r_2' \cos\beta_2 = \sin\beta_2 = \frac{\overline{A'C'}}{\overline{AC}} = \frac{\overline{A'C'}}{n_2}.$$

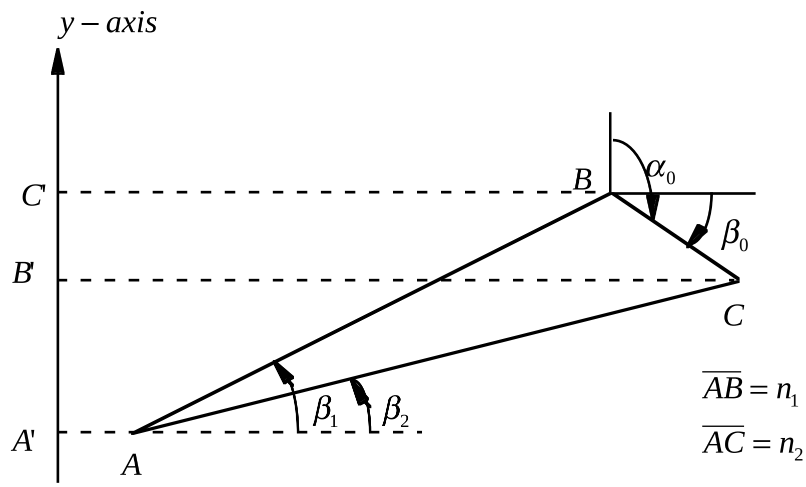


Figure 8.

Rewriting again

$$\overline{A'B'} = n_1 r_1' \cos\beta_1$$
$$\overline{A'C'} = n_2 r_2' \cos\beta_2$$

Note that $\sin\beta_0 = \dfrac{\overline{B'C'}}{\overline{BC}}$ or $\overline{B'C'} = \overline{BC}\sin\beta_0$. Along the x-axis

$$\overline{A'B'} = \overline{A'C'} + \overline{B'C'}$$

or, substituting our expressions for these quantities,

$$(n_1 r_1')\cos\beta_1 = (n_2 r_2')\cos\beta_2 + \overline{BC}\sin\beta_0$$

(NOTE: This is a complex way of deriving these matrices.)

But $\beta_0 = \alpha - \dfrac{\pi}{2}$ so $\sin(\beta_0) = \sin\left(\alpha - \dfrac{\pi}{2}\right) = -\cos(\alpha)$. From the previous page $\cos(\alpha) = -\dfrac{r_1}{R}$ so that

$$n_1 r_1' \cos\beta_1 = n_2 r_2' \cos\beta_2 + \frac{r_1}{R}\overline{BC}$$

Examining the diagram again we see that along AF we have

$$\overline{BC} = \overline{AE} - \overline{AD} = n_2\cos\theta_2 - n_1\cos\theta_1$$

$$n_1 r_1' \cos\beta_1 = n_2 r_2' \cos\beta_2 + \frac{r_1}{R}\left(n_2\cos\theta_2 - n_1\cos\theta_1\right) \qquad (11)$$

We may now linearize this using the paraxial ray approximations $\cos\beta_1 \cong \cos\beta_2 \cong 1$and $\cos\theta_1 \cong \cos\theta_2 \cong 1$. The result is

$$(n_2 r_2') = (n_1 r_1') - a r_1 \qquad (12)$$

where we have defined $a = \dfrac{n_2 - n_1}{R}$. This coefficent a is known as the optical power of the surface and has natural units of meter^{-1}. This unit is known in the optical industry as a diopter and is often indicated by writing $P = \dfrac{n_2 - n_1}{R}$ instead of a as above.

Deviations from (12) for a real lens are called abberations and may be predicted by using a Taylor series expansion of $\cos x$ in (12). Recall that $\cos x = 1 - \frac{x^3}{3!} + \frac{x^5}{5!} + \ldots$ where $|x| < 1$. We used only the first term $\cos x \cong 1$ to derive (12). Most optical aberrations that are significant in real optical systems can be described by retaining third and fifth order terms in the cosine expansion.

Equation (12) together with $r_2 = r_1$ allows us to write the refraction matrix $R_{refraction}$

$$R_{refraction} = \begin{bmatrix} 1 & -a \\ 0 & 1 \end{bmatrix}$$

It is useful at this point to compare this result with the ray matrix for reflection from a curved surface

$$R_{reflection} = \begin{bmatrix} 1 & +\frac{2n}{R} \\ 0 & 1 \end{bmatrix}$$

Note that if $n_2 = -n_1$ in $R_{refraction}$ then $R_{refraction} = R_{reflection}$. This is the basis for sign convention #6 and allows reflection to be represented as a special case of refraction.

The Lens

A lens is simply two curved dielectric interfaces separated by a small distance d as shown below.

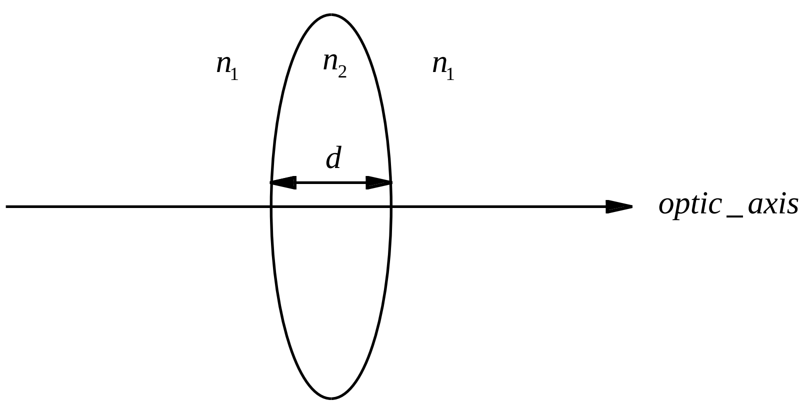


Figure 9

The ray matrix describing such a lens can be developed by following a light ray through the lens. At the front surface of the lens the ray is refracted by a surface of positive radius of curvature. If the input ray is described by r_1 and r_1' the resulting ray (r_2, r_2') can be computed using the refraction matrix

$$\begin{bmatrix} n_2 r_2' \\ r_2 \end{bmatrix} = \begin{bmatrix} 1 & -\frac{n_2 - n_1}{r_1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} n_1 r_1' \\ r_1 \end{bmatrix} = R_1 \begin{bmatrix} n_1 r_1' \\ r_1 \end{bmatrix}$$

This is the ray entering the lens. The light ray incident on the second lens surface is found by considering the translation of (r_2, r_2') as it passes through the lens, i.e.,

$$\begin{bmatrix} n_3 r_3' \\ r_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{d}{n_2} & 1 \end{bmatrix} \begin{bmatrix} n_2 r_2' \\ r_2 \end{bmatrix} = TR_1 \begin{bmatrix} n_1 r_1' \\ r_1 \end{bmatrix}$$

The second dielectric interface refracts this ray to yield

$$\begin{bmatrix} n_4 r_4' \\ r_4 \end{bmatrix} = \begin{bmatrix} 1 & -\dfrac{n_1 - n_2}{r_2} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} n_2 r_3' \\ r_3 \end{bmatrix} = R_2 T R_1 \begin{bmatrix} n_1 r_1' \\ r_1 \end{bmatrix}$$

The overall transformation of the incident ray is described by the system matrix $S \equiv R_2 T R_1$. Before examining this matrix in detail it is convenient to define the following powers of the lens refractive surfaces: $P_1 \equiv \dfrac{n_2 - n_1}{r_1}$ (the refractive power of the first surface) and $P_2 \equiv \dfrac{n_1 - n_2}{r_2}$ (the refractive power of the second surface). Using these definitions the system matrix S may be written as

$$S = \begin{bmatrix} 1 - \dfrac{d}{n_2} P_2 & -P_1 - P_2 + \dfrac{d}{n_2} P_1 P_2 \\ \dfrac{d}{n_2} & 1 - \dfrac{d}{n_2} P_1 \end{bmatrix}$$

In general, <u>any</u> optical system can be described by a system matrix S where

$$S = \begin{bmatrix} b & -a \\ -d & c \end{bmatrix}$$

The quantities a, b, c and d are called the Gaussian coefficients of the system. Their significance will become apparant when we describe image formation.

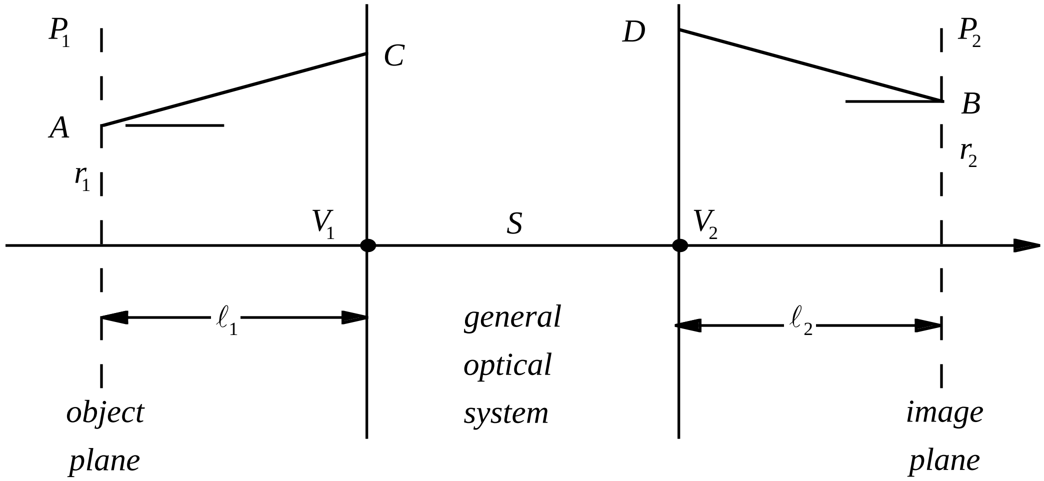
The matrix S has the interesting property that $\det S = ad - bc = 1$. To show that this is so recall that S is a product of translation and refraction matrices. The determinant of a refraction matrix is 1; likewise, the determinant of a translation matrix is 1. Since the determinant of a product matrix is the product of the determinants, i.e., $\det(AB) = \det(A)\det(B)$, where A and B are 2x2 matrices. The result is that any system matrix describing a paraxial ray system using combinations of refracting, reflecting and translating matrices will have a unity determinant.

<u>Image Formation</u>

Consider an optical system described by the system matrix S of Gaussian coefficents, i.e.,

$$S = \begin{bmatrix} b & -a \\ -d & c \end{bmatrix}$$

Let us now write the transform for this system for incident rays in a plane P_1 located ℓ_1 to the left of the optical system and P_2 located ℓ_2 to the right as indicated below.



The transformation of a ray between these planes can be given by

$$S_{P_1P_2} = \begin{bmatrix} 1 & 0 \\ \ell_2 & 1 \end{bmatrix} \begin{bmatrix} b & -a \\ -d & c \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\ell_1 & 1 \end{bmatrix}$$

$$S_{P_1P_2} = \begin{bmatrix} b + a\ell_1 & -a \\ b\ell_2 + a\ell_1\ell_2 - d - c\ell_1 & c - a\ell_2 \end{bmatrix}$$

Note that ℓ_1 is measured from V_1 and is negative; hence the minus sign.

Point B in P_2 (the image plane) is called the image of point A in P_1 (the object plane) when $r_2 = \beta r_1$ independent of the slope of the incoming ray AC. β is a constant and is the system magnification. Let the system input be $\begin{bmatrix} n_1 r_1' \\ r_1 \end{bmatrix}$ and the output $\begin{bmatrix} n_1 r_2' \\ r_2 \end{bmatrix}$. The resulting transformation leading to r_2 is

$$r_2 = \left(b\ell_2 + a\ell_1\ell_2 - d - c\ell_1\right) n_1 r_1' + \left(c - a\ell_2\right) r_1$$

For r_2 to be independent of r_1' it is necessary that

$$b\ell_2 + a\ell_1\ell_2 - d - c\ell_1 = 0$$

Solving for ℓ_2 in this expression

$$\ell_2 = \frac{d + c\ell_1}{b + a\ell_1}$$

The magnification $\beta \equiv \dfrac{r_2}{r_1}$ is then given by

$$\beta = \frac{r_2}{r_1} = c - a\ell_2 = c - a\left(\frac{d + c\ell_1}{b + a\ell_1}\right) = \frac{cb - ad}{b + a\ell_1}$$

To simplify this result recall that $\det\left(S_{P_1P_2}\right) = cb - ad = 1$. Thus,

$$\beta = \frac{1}{b + a\ell_1}$$

β may be either positive or negative. A negative β merely indicates an inverted image.

We can now rewrite $S_{P_1P_2}$ as $_{P_1P_2} = \begin{bmatrix} \dfrac{1}{\beta} & -a \\ 0 & \beta \end{bmatrix}$

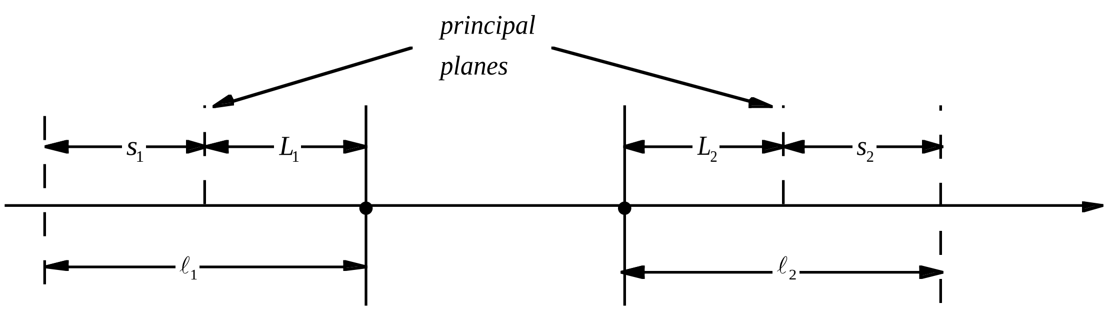
The planes P_1 and P_2 such that this matric describes the ray matrix transformation between the planes are called conjugate planes. The principal planes are conjugate planes for which $\beta = \pm 1$ (+1 in the following derivation), i.e., a 1:1 imaging relationship. Using $\beta = +1$ we solve for ℓ_1 and ℓ_2

$$\beta = \frac{1}{b + a\ell_1} = 1 \qquad \Rightarrow \qquad \ell_1 = \frac{1-b}{a}$$

$$\beta = c - a\ell_2 = 1 \qquad \Rightarrow \qquad \ell_2 = \frac{c-1}{a}$$

ℓ_1 and ℓ_2 locate the principal planes in terms of the elements a, b and c of the system matrix S. The principal points are the intersections of the principal planes with the optic axis. For later reference note that at the principal points we have $(n_1 r_2') = \frac{1}{\beta}(n_1 r_1') - a r_1$, or $r_2' = r_1'$. Any ray passing through the principal point in the object plane with slope r_1' will pass through the principal point in the image plane with the same slope. Points satisfying this 1:1 relationship in points are called nodal points.

It is common practice to locate the object and image planes relative to the principal planes rather than the vertices of the optical system. Let the distances from the vertices to the principal points be denoted by L_1 and L_2. Then $\ell_1 = s_1 + L_1$ and $\ell_2 = s_2 + L_2$.



As L_1 and L_2 locate the principal planes they satisfy

$$1 = \frac{1}{b + aL_1} = c - aL_2$$

Using this result

$$\beta = \frac{1}{b + a\ell_1} = \frac{1}{b + as_1 + aL_1} = \frac{1}{1 + as_1}$$

$$\beta = c - a\ell_2 = c - as_2 - aL_2 = 1 - as_2$$

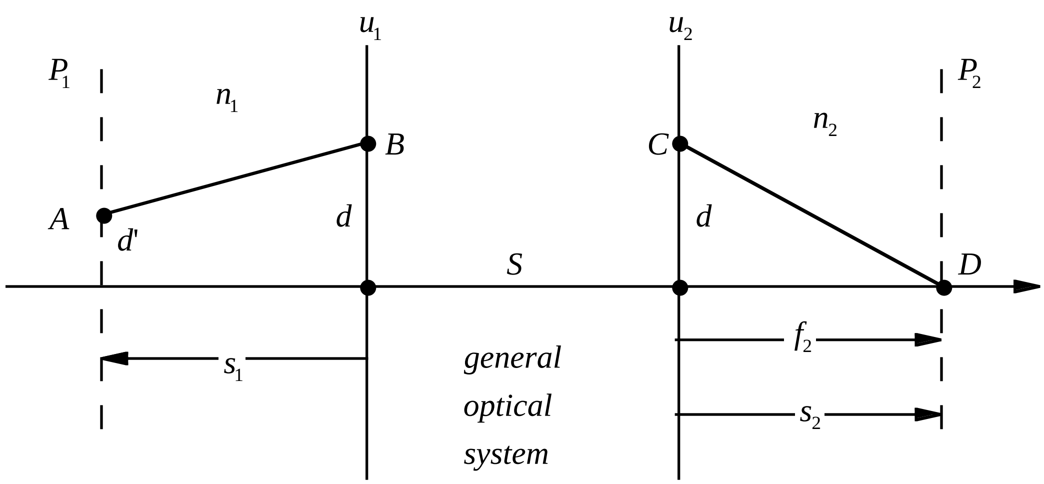
The system matrix may now be written as

$$S_{P_1P_2} = \begin{bmatrix} 1 + as_1 & -a \\ 0 & 1 - as_2 \end{bmatrix}$$

We will now derive the simple lens law

$$\frac{1}{s_2} = \frac{1}{s_1} + \frac{1}{f_2}$$

where f_2 is the focal length of the system for light incident from the left. The classical definition of the focal length of a system S is that any incident light rays parallel to the optic axis at P_1 will cross the optic axis at a point D a distance f_2 from the second principal plane.



Let P_1 and P_2 be conjugate planes such that the focal point of S is the conjugate point of P_2 (the intersection of a conjugate plane with the optic axis is the conjugate point). We now need to determine s_1. Suppose ray AB is not parallel to the optic axis. If AB is a distance d' away from the optic axis at P_1 and d at u_1. The slope of AB is then $r_1' = \dfrac{d' - d}{s_1}$. Since $r_1' = 0$ by the problem definition it follows that $s_1 \to -\infty$ (s_1 is negative from the drawing). To relate this to s_2 we recall that $\det\left(S_{P_1P_2}\right) = \dfrac{1 + as_1}{1 - as_2} = 1$ or, rewriting,

$$\frac{1}{s_2} = \frac{1}{s_1} + \frac{1}{a}$$

As $s_1 \to -\infty$ we see that $s_2 \to a$. All that remains is to relate f_2 and a to have the classical lens law. This may be done by noting that between the principal planes u_1 and u_2

$$n_2 r_2' = n_1 r_1' - a r_1$$

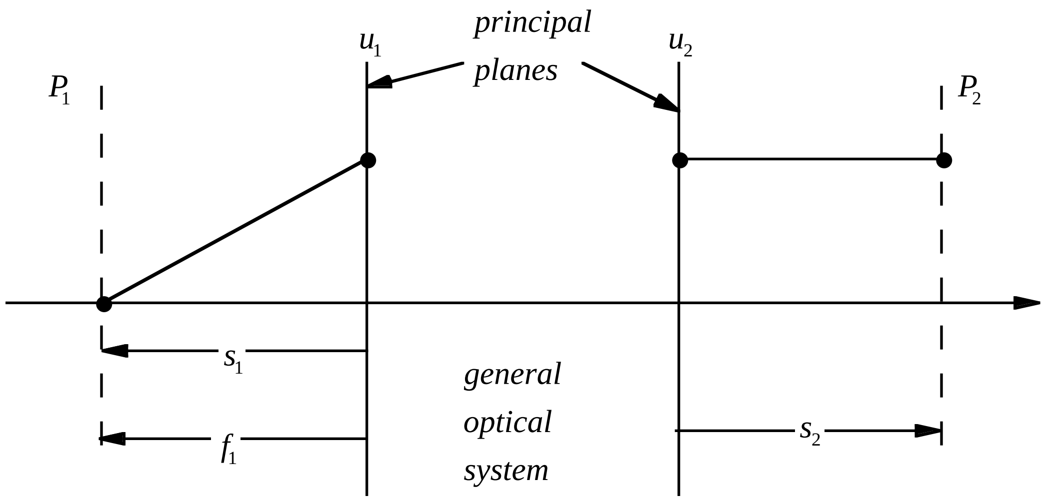
or using $r_1 = d$, $r_1' = 0$, $r_2' = \dfrac{\Delta Y}{\Delta X} = \dfrac{-d}{+f_2} = -\dfrac{d}{f_2}$ we get

$$a = +\frac{n_2}{f_2}$$

This is exactly the classical lens law whenever $n_2 = 1$

$$\frac{1}{s_2} = \frac{1}{s_1} + \frac{n_2}{f_2}$$

The system S has a second focal point such that light rays passing through this point with any slope will be transformed into parallel rays, i.e., $r_1' = anything$, $r_2' = 0$.



To solve this problem we first establish the image-object relationship between the conjugate planes P_1 and P_2. By analogy with the previous case $r_2' = 0$ implies $s_2 \to +\infty$. Using $\frac{1}{s_2} = \frac{1}{s_1} + \frac{1}{a}$ gives $s_1 = -a$. To relate a to f_1 we use the transformation

$$n_2 r_2' = n_1 r_1' - a r_1$$

where $r_1 = d$, $r_1' = \frac{+d}{-f_1}$, $r_2' = 0$ giving $n_1\left(\frac{+d}{-f_1}\right) = ad$, or

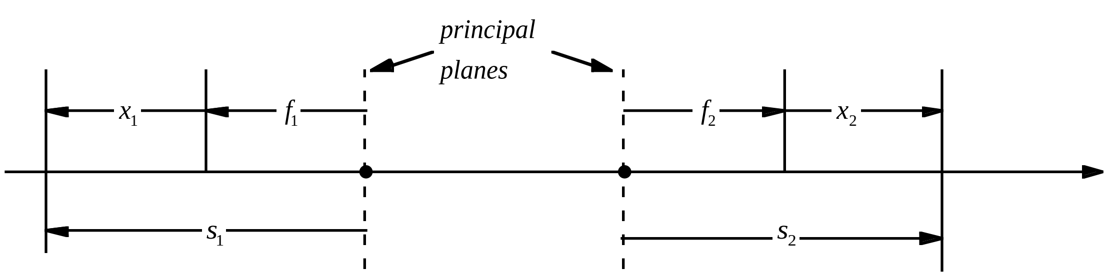
$$a = -\frac{n_1}{f_1}$$

These two expressions for a give the lens law

$$\frac{1}{s_2} - \frac{1}{s_1} = \frac{n_2}{f_2} = -\frac{n_1}{f_1}$$

For later reference we will derive the Newtonian form of the lens law:

$$x_1 x_2 = -f_2^{\,2}$$



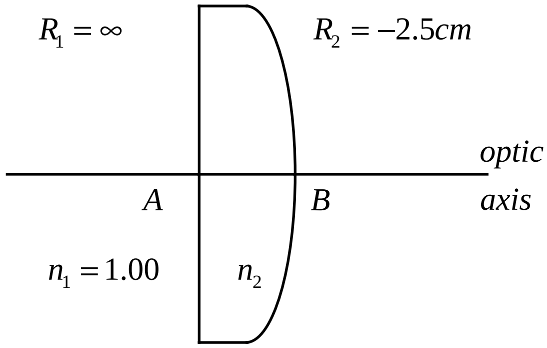
From the drawing we write:

$$s_1 = f_1 + x_1 = -f_2 + x_1$$
$$s_2 = f_2 + x_2$$

Substituting this into the lens law where $n_1 = n_2 = 1.000$ we get

$$x_1 x_2 = -f_2^{\,2}$$

Example: Plano-convex lens



$\overline{AB} = d = 0.6cm$

$n_2 = 1.5$

From page 13 the system matrix for planes passing through A and B perpendicular to the optical axis is:

$$S_{AB} = \begin{bmatrix} 1 - \frac{d}{n_2} P_2 & -P_1 - P_2 + \frac{d}{n_2} P_1 P_2 \\ \frac{d}{n_2} & 1 - \frac{d}{n_2} P_1 \end{bmatrix}$$

For the given lens

$$P_1 = \frac{n_2 - n_1}{r_1} = \frac{1.5 - 1}{\infty} = 0$$

$$P_2 = \frac{n_1 - n_2}{r_2} = \frac{1 - 1.5}{-2.5} = 0.2$$

$$\frac{d}{n_2} = \frac{0.6}{1.5} = 0.4$$

$$S_{AB} = \begin{bmatrix} 1 - (0.4)(0.2) & -0.2 \\ 0.4 & 1 \end{bmatrix} = \begin{bmatrix} 0.92 & -0.2 \\ 0.4 & 1 \end{bmatrix} = \begin{bmatrix} b & -a \\ -d & c \end{bmatrix}$$

As a check on our calculations

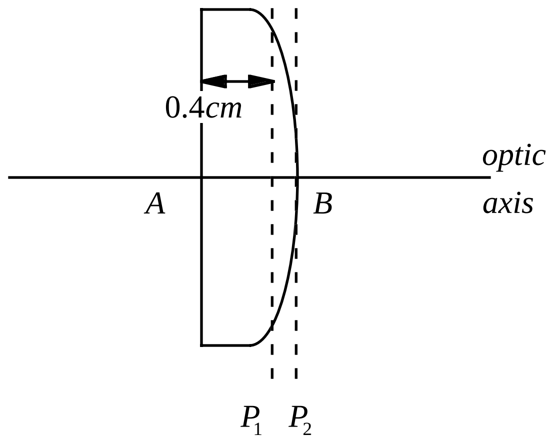
$$\det(S_{AB}) = (0.92)(1) - (0.2)(0.4) = 0.92 + 0.08 = 1.00$$

The location of the principal planes is given by

$$\ell_1 = \frac{1 - b}{a} = \frac{1 - 0.92}{0.2} = +0.4$$

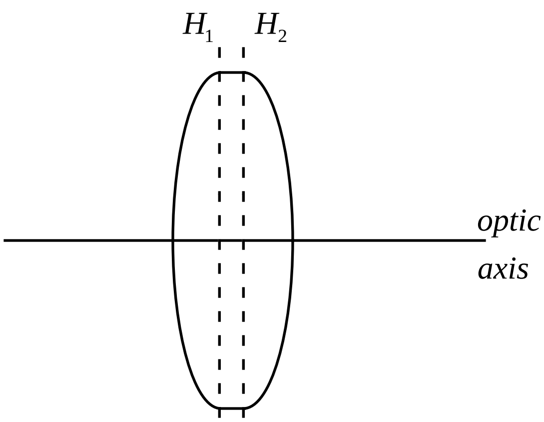
$$\ell_2 = \frac{c - 1}{a} = \frac{1 - 1}{0.2} = 0$$

The principal planes are then located as shown below

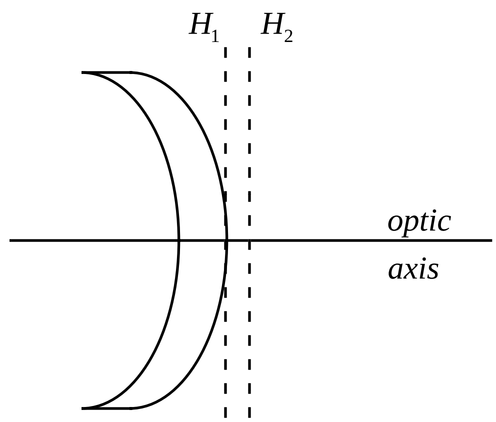


This type of lens is often found in optical instruments because a high quality flat surface is much easier to produce than a spherical surface; hence, a good plano-convex lens is cheaper than a comparable quality biconvex lens.

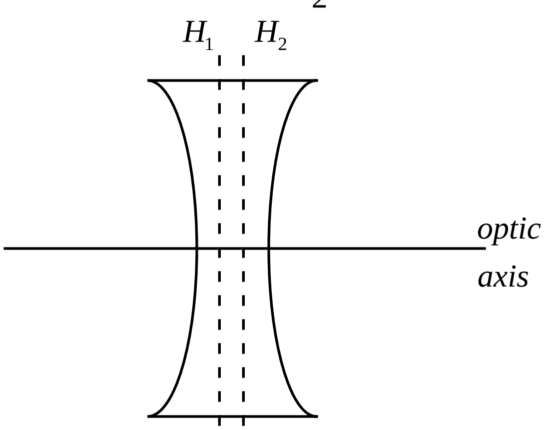
The location of the principal planes for some common lens shapes are shown below.



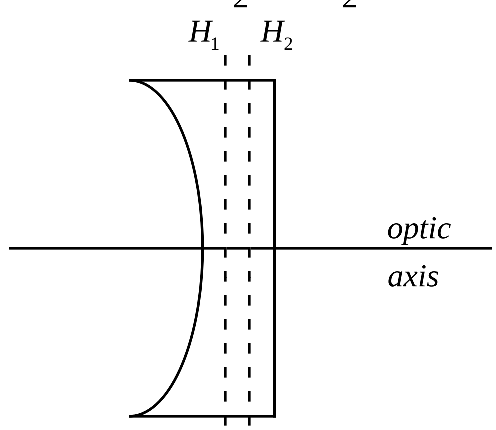
$$P_1 = P_2 = \frac{P}{2} > 0$$



$$P_1 = -\frac{P}{2};\ P_2 = \frac{3}{2}P > 0$$



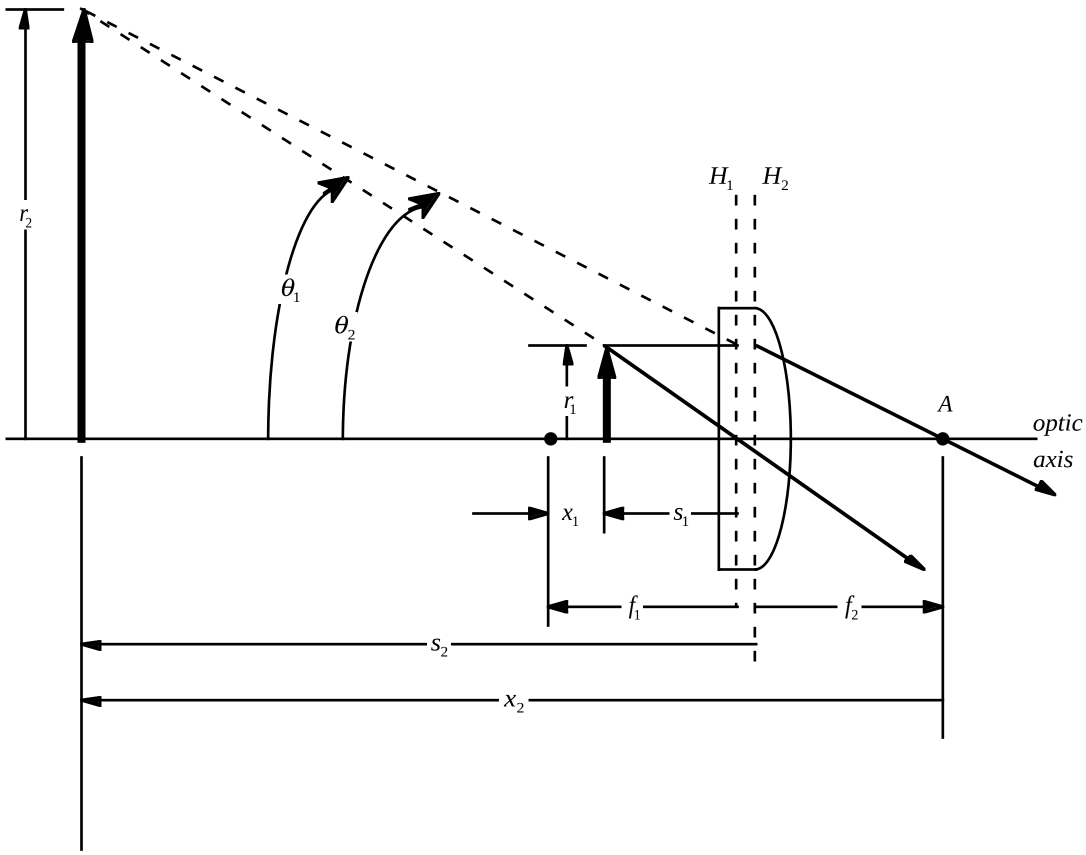
$$P_1 = P_2 = \frac{P}{2} < 0$$



$$P_1 = P < 0;\ P_2 = 0$$

In these sketches H_1 and H_2 are principal planes, P_1 and P_2 are the refractive powers of the first and second surfaces respectively, and $P = P_1 + P_2$, i.e., a thin lens where $d \approx 0$.

Simple magnifier:
We will now analyze a plano-convex lens as a simple magnifier as shown below.



Several features of this drawing are worth mentioning regarding graphical ray tracing. Note that the lens is assumed to be a thin lens, i.e., the distance between the principal planes H_1 and H_2 is small (≈ 0). In the drawing the object to be imaged is located s_1 in front of H_1. To locate the image we trace two rays from the object and graphically determine their intersection—this intersection locates the image. The first ray will be drawn parallel to the optic axis. By the definition of principal planes this ray must pass through the focal point A. The second ray is drawn from the object to the principal point of H_1. On page 17 we notes that principal points are nodal points; hence, the ray will leave the principal point of H_2 with the same slope as it had crossing H_1. These rays may be extended indefinitely until they intersect. A perpendicular to the optic axis from this point of intersection will locate the image.

As shown in the drawing the object is located near the first focal plane. The eye of the observer is located near the second focal plane. θ_2 is usually a good measure of the apparant size of the image. $Tan(\theta_2) = \dfrac{r_2}{-x_2}$ or, because of the paraxial ray approximation,

$$\theta_2 \approx \frac{r_2}{-x_2}$$

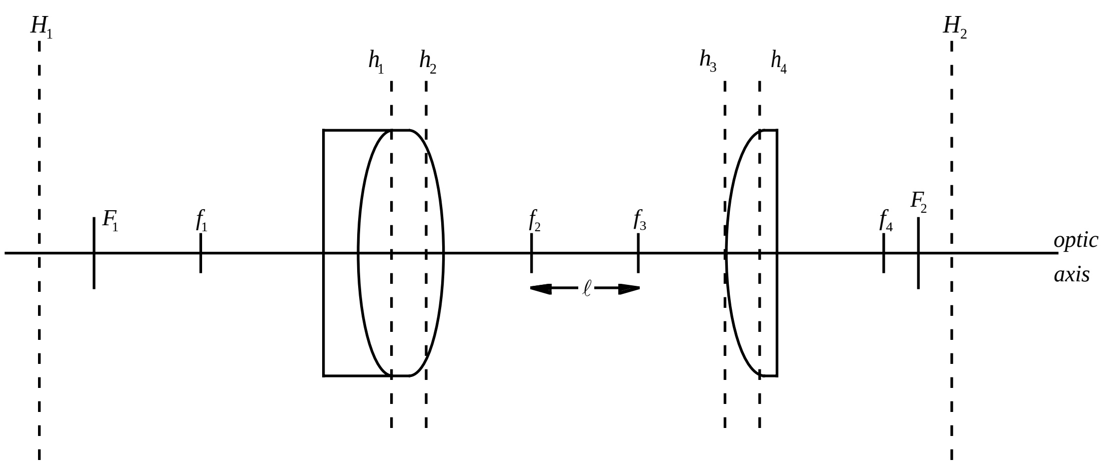
The magnification is given by $\beta = 1 - \frac{s_2}{f_2}$. If, as is usually the case, $\beta >> 1$ then $\beta \approx -\frac{s_2}{f_2}$. Substituting this result into the expression for θ_2 we get $\theta_2 \approx \frac{r_1}{-f_1}$ since $\theta_2 = \frac{-r_2}{x_2} = \frac{\left(\frac{s}{f_2}\right) r_1}{s_2 - f_2} \cong \frac{r_1}{f_2} = \frac{r_1}{-f_1}$. A comfortable viewing distance for the eye is approximately 10 inches (about 250mm). The angle θ' that the object would subtend if we viewed it from a distance of 250 mm unaided is $\theta' \approx \frac{r_1}{250mm}$. Common usage defines the magnification of the lens as

$$M = \frac{\theta_2}{\theta'} = \frac{-\frac{r_1}{f_1}}{\frac{r_1}{250mm}} = -\frac{250mm}{f_1}$$

Note that since f_1 is negative (see drawing) $M > 0$ and the image is upright.

<u>The Compound Microscope</u>

We will now consider a more complicated instrument, a compound microscope consisting of two lens separated by a distance d as shown by the following figure.



Items indicated by capitals are referring to the overall optical system; small letters refer to items characterizing the individual optical elements. Between planes h_1 and h_4

$$S_{h_1 h_4} = \begin{bmatrix} 1 & -\frac{1}{f_4} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ d & 1 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{f_2} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 - \frac{d}{f_4} & -\frac{1}{f_2} - \frac{1}{f_4} + \frac{d}{f_2 f_4} \\ d & 1 - \frac{d}{f_2} \end{bmatrix} = \begin{bmatrix} b & -a \\ -d & c \end{bmatrix}$$

$$-\frac{1}{F_2}=-\frac{1}{f_2}-\frac{1}{f_4}+\frac{d}{f_2 f_4}=\frac{\ell}{f_2 f_4} \text{ where } d=f_2+f_4+\ell$$

$$F_2=-\frac{f_2 f_4}{\ell}$$

$$L_1=\frac{1-b}{a}=\frac{1-\left(1-\frac{d}{f_4}\right)}{-\frac{\ell}{f_2 f_4}}=-\frac{df_4}{\ell}$$

$$L_2=\frac{c-1}{a}=\frac{1-\left(\frac{d}{f_2}-1\right)}{-\frac{\ell}{f_2 f_4}}=+\frac{df_4}{\ell}$$

For a typical microscope $f_2=f_4=16mm$ and $\ell=160mm$

$$L_1=-\frac{(16+16+160mm)(16mm)}{160mm}=-19.2mm$$

$$L_2=+\frac{(16+16+160mm)(16mm)}{160mm}=+19.2mm$$

Let us now locate the object relative to the system focal points. As before for good viewing the virtual image will be at $x_2\approx-250mm$. Then, using the Newtonian form of the lens law $x_1x_2=-F_2^2$.

$$x_1=-\frac{F_2^2}{x_2}=-\frac{(1.6)^2}{-250}=+0.01024mm$$

which is almost at the first focal point F_1. The system magnification is $M=\frac{250}{F_2}=-\frac{250}{f_2}\frac{\ell}{f_4}$ indicating an inverted image. For the numbers given $M\approx-156$.

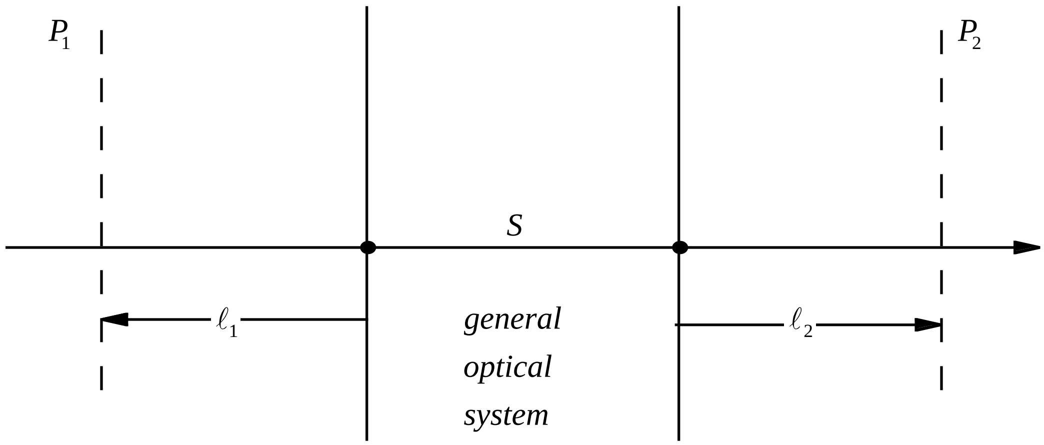
Let us now examine the intermediate image formed by the first lens. The object is very near the system focal point F_1 so that, relative to f_1, $x_1=-1.6mm$. Using the Newtonian lens law

$$x_2=-\frac{f_2^2}{x_1}=\frac{(16)^2}{-1.6}\cong 160mm$$

This indicates that the image is formed approximately at the focal point of the eyepiece. This intermediate image is real and inverted. The eyepiece may now be treated as a simple magnifier with magnification $M_e=-\frac{250}{f_1}=+\frac{250}{f_2}$. For the objective lens the magnification $M_o=1-\frac{s_2}{f_2}\approx\frac{-\ell}{f_2}$. The system magnification M is then seen to be approximately equal to the product of the eyepiece and objective magnifications.

The Telescope

A telescopic system is defined to be an optical system having a slope transformation that is independent of r_1, i.e., of the form $n_2r_2'=kn_1r_1'$ where k is a constant. Let us consider the optical system of the diagram below and examine the conditions under which it is telescopic.



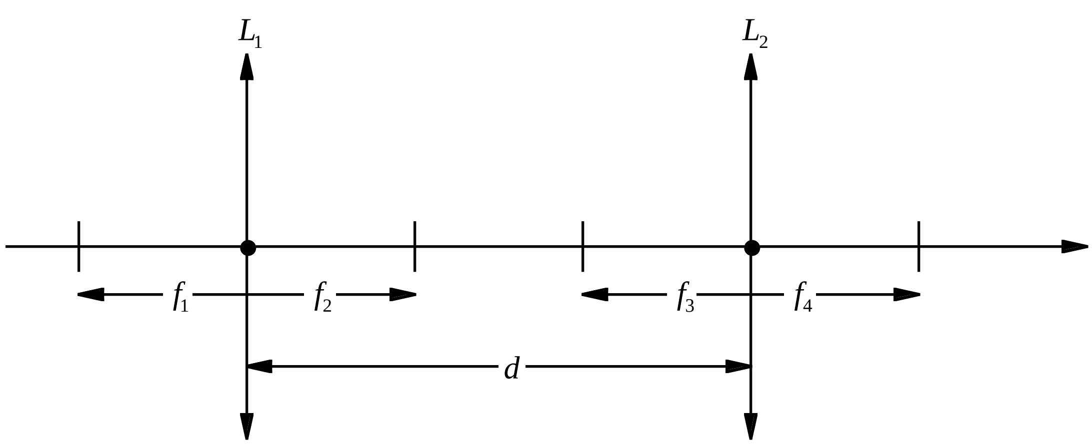
If S is a general matrix of Gaussian coefficients, i.e., $\begin{bmatrix} b & -a \\ -d & c \end{bmatrix}$, then

$$S_{P_1P_2} = \begin{bmatrix} 1 & 0 \\ \frac{\ell_2}{n_2} & 1 \end{bmatrix}\begin{bmatrix} b & -a \\ -d & c \end{bmatrix}\begin{bmatrix} 1 & 0 \\ -\frac{\ell_1}{n_1} & 1 \end{bmatrix} = \begin{bmatrix} b + a\frac{\ell_1}{n_1} & -a \\ b\frac{\ell_2}{n_2} + a\frac{\ell_1}{n_1}\frac{\ell_2}{n_2} - d - c\frac{\ell_1}{n_1} & c - a\frac{\ell_2}{n_2} \end{bmatrix}$$

The slope transformation between P_1 and P_2 is then

$$(n_2 r_2') = \left(b + a\frac{\ell_1}{n_1}\right)(n_1 r_1') - a r_1$$

For this transformation to be independent of r_1 it is necessary that $a = 0$; hence, $(n_2 r_2') = b(n_1 r_1')$. Note that once a is set equal to zero it remains zero for any choice of ℓ_1 and ℓ_2, i.e., it is invariant under translation. Consider the system shown below composed of two thin lenses separated by a distance d.



Between lenses L_1 and L_2

$$S_{L_1L_2} = \begin{bmatrix} 1 & -\frac{1}{f_4} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ d & 1 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{f_2} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 - \frac{d}{f_4} & -\frac{1}{f_2} - \frac{1}{f_4} + \frac{d}{f_2 f_4} \\ d & 1 - \frac{d}{f_2} \end{bmatrix}$$

For this system to qualify as telescopic

$$\frac{1}{f_2} + \frac{1}{f_4} - \frac{d}{f_2 f_4} = 0$$

Using this equality we can re-write $S_{L_1L_2}$ as

$$S_{L_1L_2} = \begin{bmatrix} -\frac{f_2}{f_4} & 0 \\ f_2 + f_4 & -\frac{f_4}{f_2} \end{bmatrix} = \begin{bmatrix} p_\alpha & 0 \\ f_2 + f_4 & \frac{1}{p_\alpha} \end{bmatrix}$$

where we have defined $p_\alpha = -\frac{f_2}{f_4}$. Note that the telescopic system requirement resulted in $d = f_2 + f_4$, i.e., the focal points of the two lenses must coincide. Writing out the transformations

$$r_2' = p_\alpha r_1'$$

$$r_2 = (f_2 + f_4) r_1' + \frac{r_1}{p_\alpha}$$

where we assumed that $n_2 = n_1 = 1$.

The quantity p_α is known as the angular magnification. In general, telescopes are capable of resolving objects at great distances because of their ability to magnify the small angular separation between such objects. With $S_{L_1L_2}$ being telescopic consider the transformation between a plane H_1 located ℓ_1 to the left of L_1 and H_2 located ℓ_2 to the right of L_2

$$S_{H_1H_2} = \begin{bmatrix} 1 & 0 \\ \ell_2 & 1 \end{bmatrix} \begin{bmatrix} p_\alpha & 0 \\ f_2 + f_4 & \frac{1}{p_\alpha} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\ell_1 & 1 \end{bmatrix} = \begin{bmatrix} p_\alpha & 0 \\ \ell_2 p_\alpha + f_2 + f_4 - \frac{\ell_1}{p_\alpha} & \frac{1}{p_\alpha} \end{bmatrix}$$

For image formation we let $\ell_2 p_\alpha + f_2 + f_4 - \frac{\ell_1}{p_\alpha} = 0$. Then, the transformation between H_1 and H_2 is

$$r_2' = p_\alpha r_1'$$

$$r_2 = \frac{1}{p_\alpha} r_1$$

Notice that high magnification and good angular separation are competing processes. The larger p_α is (better angular resolution) the smaller the system magnification ($\frac{1}{p_\alpha}$) is. The proper design of a telescopic system involves a trade-off between angular resolution and magnification.

Let us examine the longitudinal magnification $\frac{\Delta \ell_2}{\Delta \ell_1}$ as opposed to the transverse magnification $\frac{r_2}{r_1}$. Differentiating the expression $\ell_2 p_\alpha + f_2 + f_4 - \frac{\ell_1}{p_\alpha} = 0$ we get

$$\Delta\ell_2 p_\alpha - \frac{\Delta\ell_1}{p_\alpha} = 0$$

$$\frac{\Delta\ell_2}{\Delta\ell_1} = \frac{1}{p_\alpha^2}$$

For the system examined all magnifications are independent of image and/or object distances and the image is real and inverted. Any such telescopic system having the same signed magnifications as derived here is called Gailean.

Stops and Apetures

Up to now we have only been concerned with the image location and size. Two other important considerations are the system field of view and the brightness of the image. Stops are related to the determination of each of these factors and, in general, stops are defined to be those elements in the optical system that determine what fraction of the light from an object point will actually reach the corresponding image point.

Let us first examine an on-axis point as shown below.

For points P_1 and P_2 $x_1 x_2 = (-4)(+1) = -4cm^2 = -f_2^2 = -(2cm)^2 = -4cm^2$. To the observer at P_2 it appears that A_2 and B_2 limit the rays coming from P_1. We shall now show that the apeture A_2B_2 is merely the image of A_1B_1. To relate the apetures first note that they satisfy the Newtonian lens law $x_1x_2 = -f_2^2$ since $x_1 = +1$, $x_2 = -4$, and $f_2 = +2$. If point A_2 is the image of A_1 then their distances from the optic axis are related by $\beta = 1 - as_2 = \dfrac{1}{1 + as_1}$ where $a = \dfrac{1}{f_2}$ and s_1 and s_2 are measured from the principal planes of the lens. For a thin lens recall that the principal planes coincide. To put β in a more tractable form write $s_2 = f_2 + x_2$ and $s_1 = x_1 + f_1$. Substituting these expressions into those for β we get $\beta = -\dfrac{x_2}{f_2} = \dfrac{f_2}{x_1}$ where we have used the fact that $f_1 = -f_2$. Note that this equality is equivalent to the lens law as $-\dfrac{x_2}{f_2} = \dfrac{f_2}{x_1} \Rightarrow x_1x_2 = -f_2^2$. With the numbers given in the drawing $\beta = +2$. To show that their heights do obey this relation and have the ratio of 2:1 we note that the slope of P_1A_1 is ${}^{1cm}\!/_{5cm} = \dfrac{1}{5}$. The distance of B_1 from the optic axis is then $6 \times \dfrac{1}{5} = 1.2cm$. The slope of P_2B_1 is ${}^{\frac{6}{5}cm}\!\Big/_{3cm} = \dfrac{2}{5}$. The distance of A_2 from the optic axis is then $5cm \times \dfrac{2}{5} = 2cm$ which is twice that of A_1 confirming the magnification of 2. Actually this derivation could have been proved for arbitrary locations of the stop.

Returning to the cone of light rays coming from P_1 it is seen that A_1B_1 constitutes the apeture stop since it limits the angular spread (θ) of the rays coming from P_1 that will be imaged to P_2.