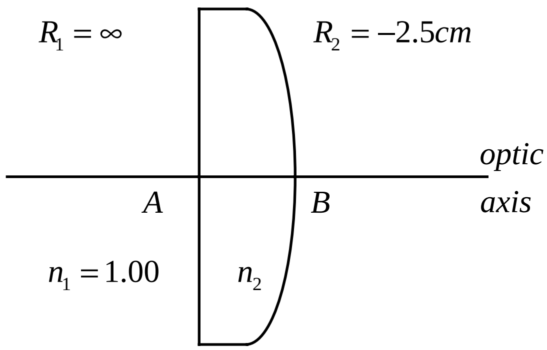
Example: Plano-convex lens



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

$n_2 = 1.5$

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

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

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

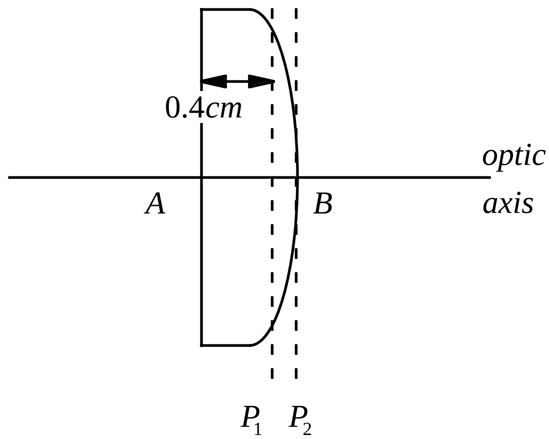
$$\det(S_{AB}) = (0.92)(1) - (0.2)(0.4) = 0.92 + 0.08 = 1.00$$

The location of the principal planes is given by

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

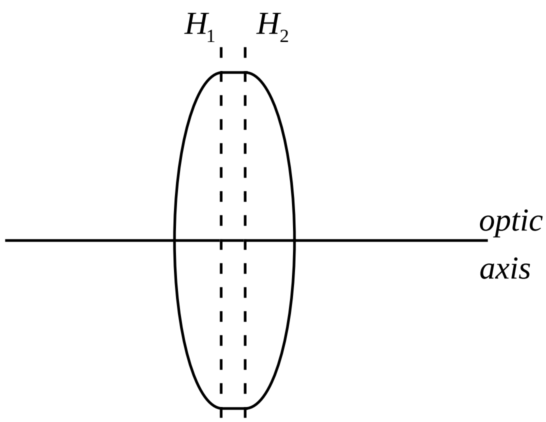
$$\ell_2 = \frac{c - 1}{a} = \frac{1 - 1}{0.2} = 0$$

The principal planes are then located as shown below

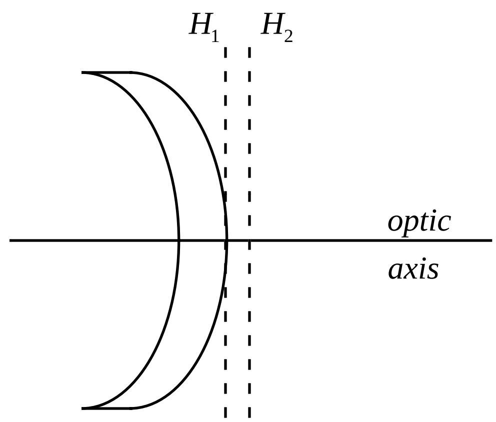


This type of lens is often found in optical instruments because a high quality flat surface is much easier to produce than a spherical surface; hence, a good plano-convex lens is cheaper than a comparable quality biconvex lens.

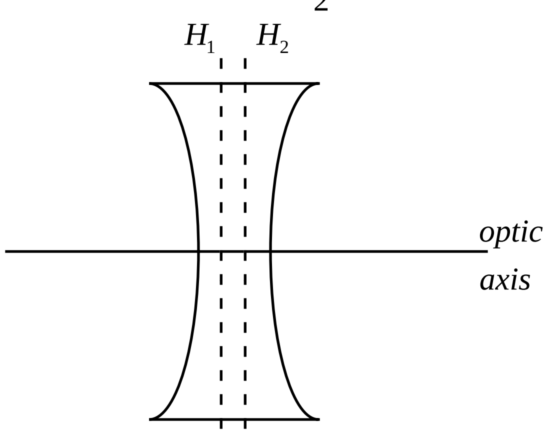
The location of the principal planes for some common lens shapes are shown below.



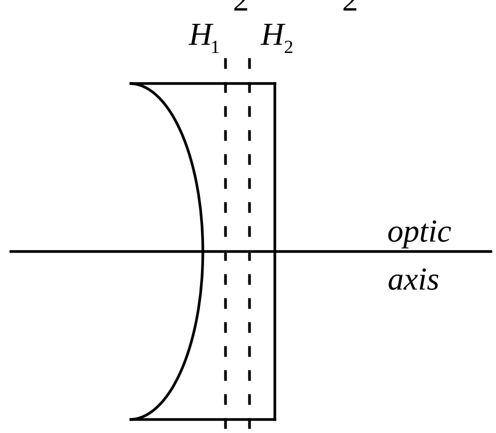
$$P_1 = P_2 = \frac{P}{2} > 0$$



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



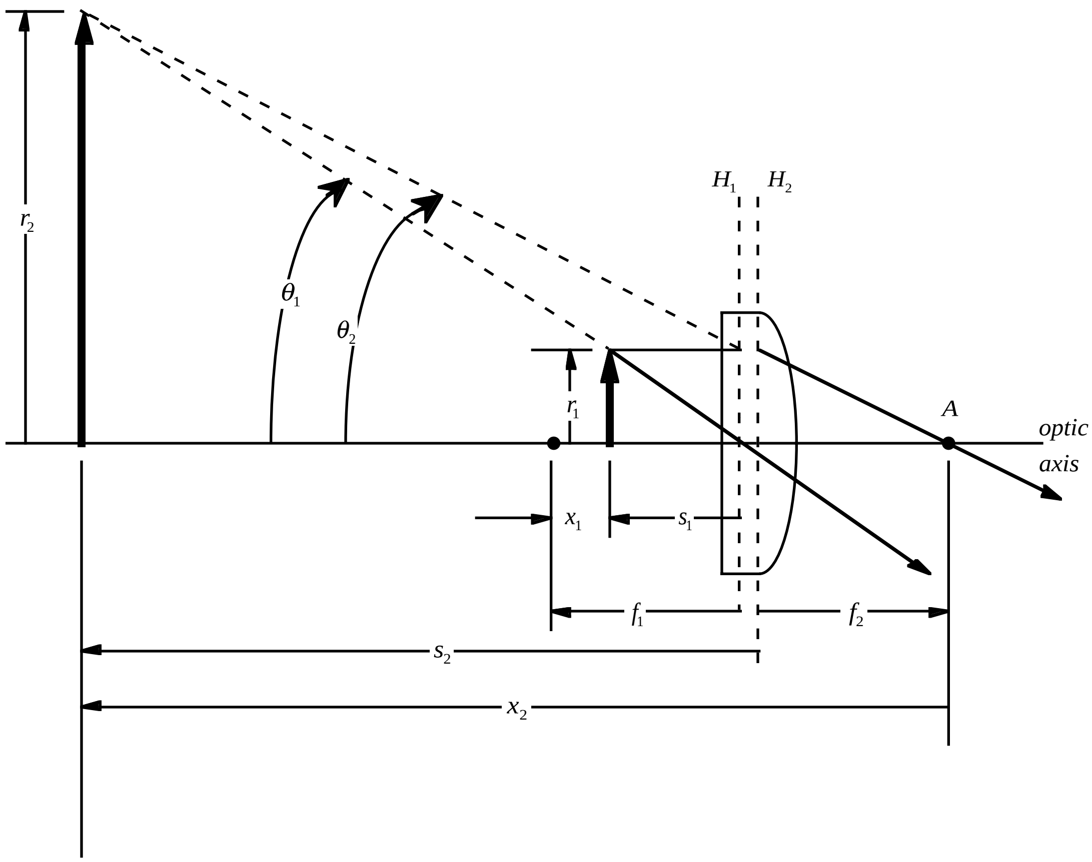
$$P_1 = P_2 = \frac{P}{2} < 0$$



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

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. $Tan(\theta_2) = \frac{r_2}{-x_2}$ or, because of the paraxial ray approximation,

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

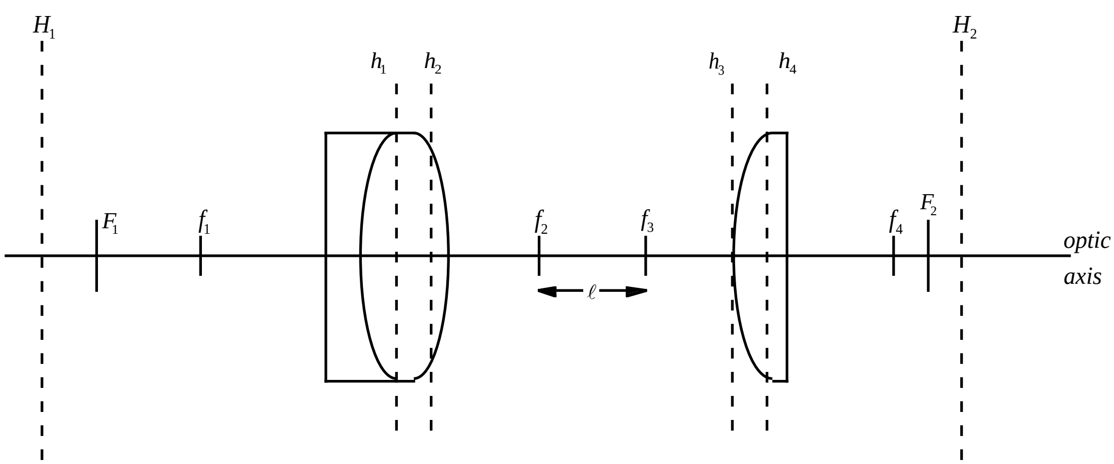
The magnification is given by $\beta = 1 - \frac{s_2}{f_2}$. If, as is usually the case, $\beta >> 1$ then $\beta \approx -\frac{s_2}{f_2}$. Substituting this result into the expression for θ_2 we get $\theta_2 \approx \frac{r_1}{-f_1}$ since $\theta_2 = \frac{-r_2}{x_2} = \frac{\left(\frac{s}{f_2}\right) r_1}{s_2 - f_2} \cong \frac{r_1}{f_2} = \frac{r_1}{-f_1}$. 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 $x_1 x_2 = -F_2^2$.

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

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 $M \approx -156$.

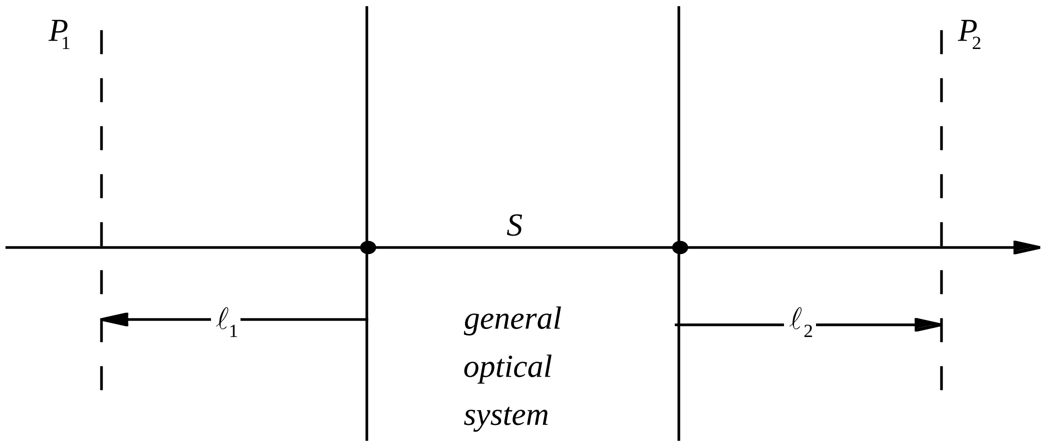
Let us now examine the intermediate image formed by the first lens. The object is very near the system focal point F_1 so that, relative to f_1, $x_1 = -1.6mm$. Using the Newtonian lens law

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

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 $n_2 r_2' = k n_1 r_1'$ where k is a constant. Let us consider the optical system of the diagram below and examine the conditions under which it is telescopic.



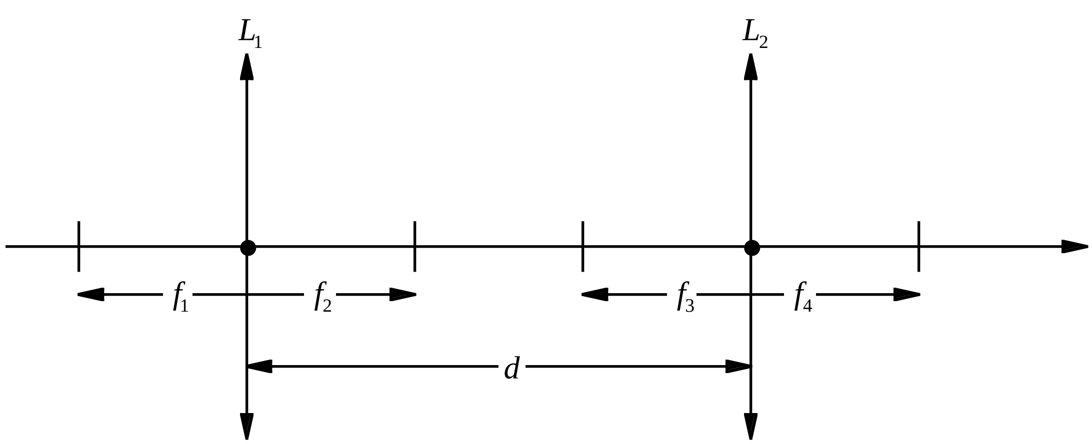
If S is a general matrix of Gaussian coefficients, i.e., $\begin{bmatrix} b & -a \\ -d & c \end{bmatrix}$, then

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

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

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

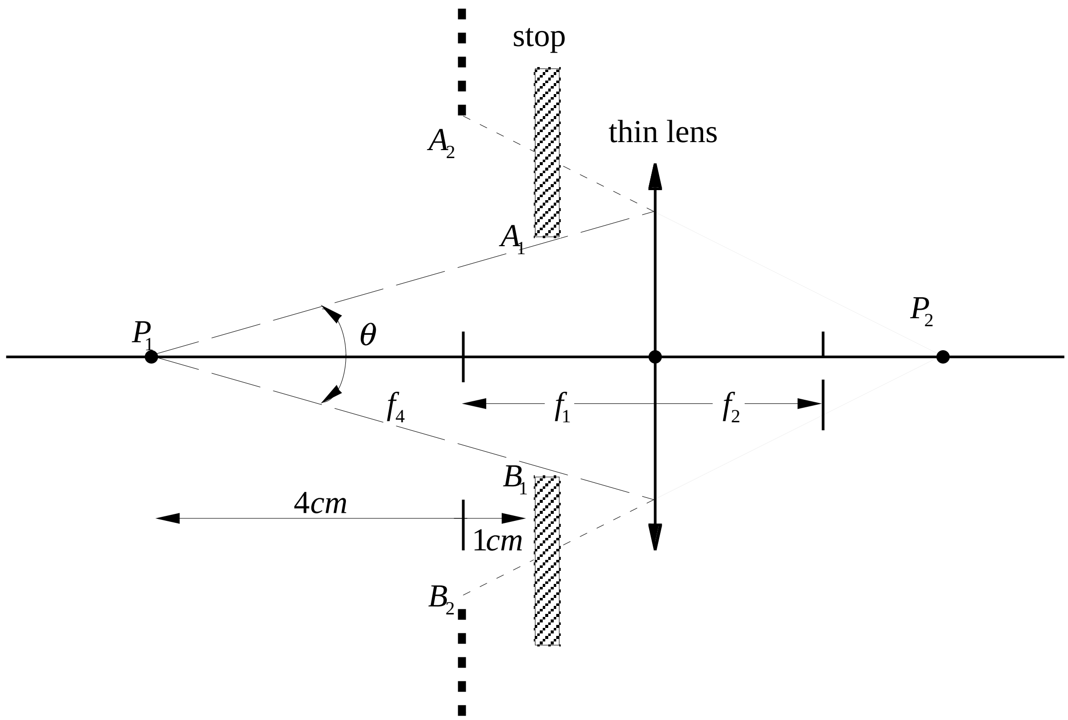
$$\frac{\Delta\ell_2}{\Delta\ell_1} = \frac{1}{p_\alpha^2}$$

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.

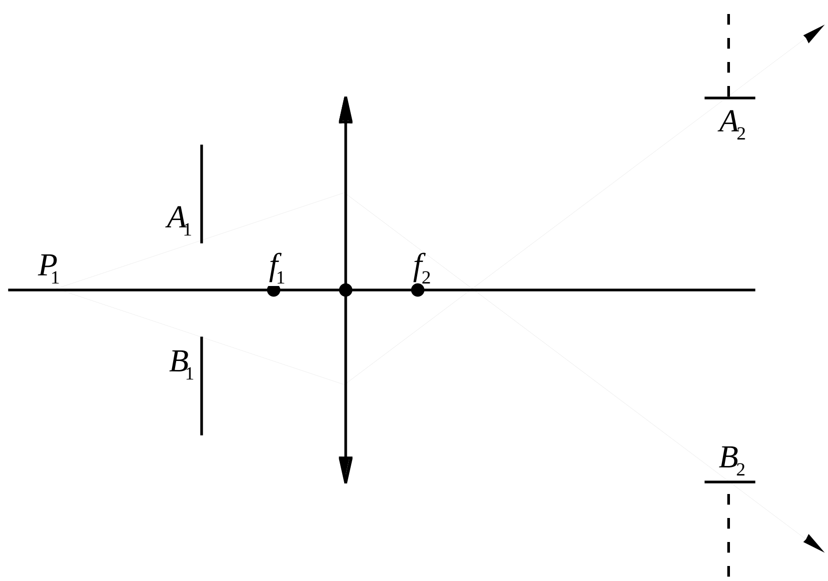
Let us first examine an on-axis point as shown below.



For points P_1 and P_2 $x_1x_2 = (-4)(+1) = -4cm^2 = -f_2^2 = -(2cm)^2 = -4cm^2$. 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 $\beta = 1 - as_2 = \frac{1}{1+as_1}$ where $a = \frac{1}{f_2}$ 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 $\beta = -\frac{x_2}{f_2} = \frac{f_2}{x_1}$ where we have used the fact that $f_1 = -f_2$. Note that this equality is equivalent to the lens law as $-\frac{x_2}{f_2} = \frac{f_2}{x_1} \Rightarrow x_1x_2 = -f_2^2$. 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 $1cm/5cm = \frac{1}{5}$. The distance of B_1 from the optic axis is then $6\times\frac{1}{5} = 1.2cm$. The slope of P_2B_1 is $\frac{6}{5}cm/3cm = \frac{2}{5}$. The distance of A_2 from the optic axis is then $5cm\times\frac{2}{5} = 2cm$ 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.

Suppose we move A_1B_1 to the left of the focal point as shown below.



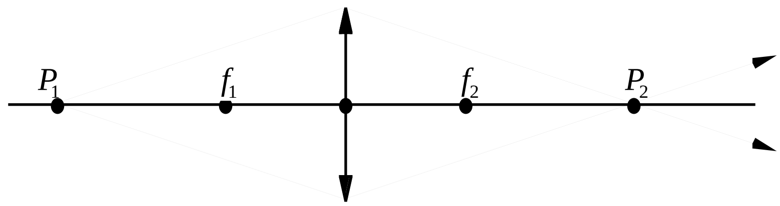
The apeture stop is still determined by A_1B_1 but to an observer to the right of the lens it is again A_2B_2 (the image of A_1B_1) that limits the cone of rays from P_1. Let us define a new concept —image space—as all physical objects to the right of the lens plus the images of all points to the left of the lens. With this definition we may define the apeture stop in image space as the exit pupil. In the two examples the cone of rays converging to or diverging from the image point is limited by the exit pupil A_2B_2.

Let us define object space as that space consisting of all physical objects to the left of the lens plus all objects located to the right of the lens. The image of the apeture stop in object space is defined to be the entrance pupil. In both examples the apeture stop is in object space; hence, it itself is the entrance pupil.

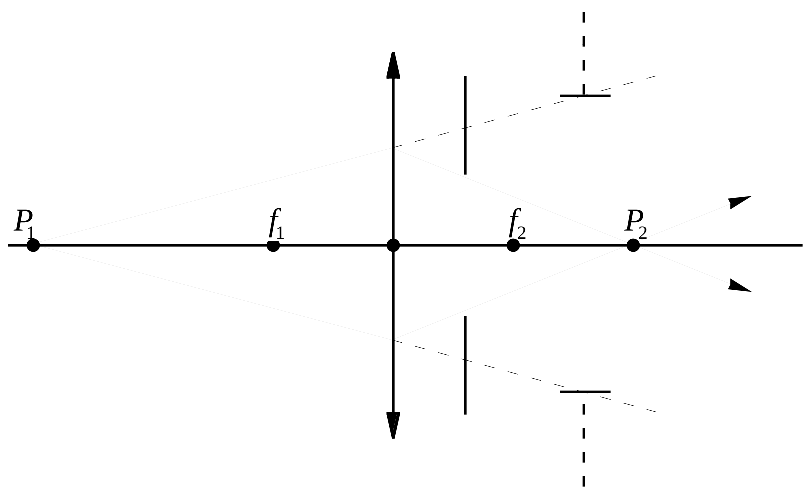
The term pupil is used to define these apetures because in optical systems to be used with the human eye the exit pupil corresponds to the pupil of the eye. Of the two examples only in the second example would such a correspondence be possible.

Consider now the imaging of off-axis points. A ray from an off-axis point (in the image plane) that passes through the center of the apeture stop is called a chief ray. Because the exit pupil and entrance pupil are images of the apeture stop the chief ray will also pass through the center of both pupils. The marginal ray is a ray from an off-axis point in the image plane which passes through the edge of the apeture stop, the entrance pupil and the exit pupil.

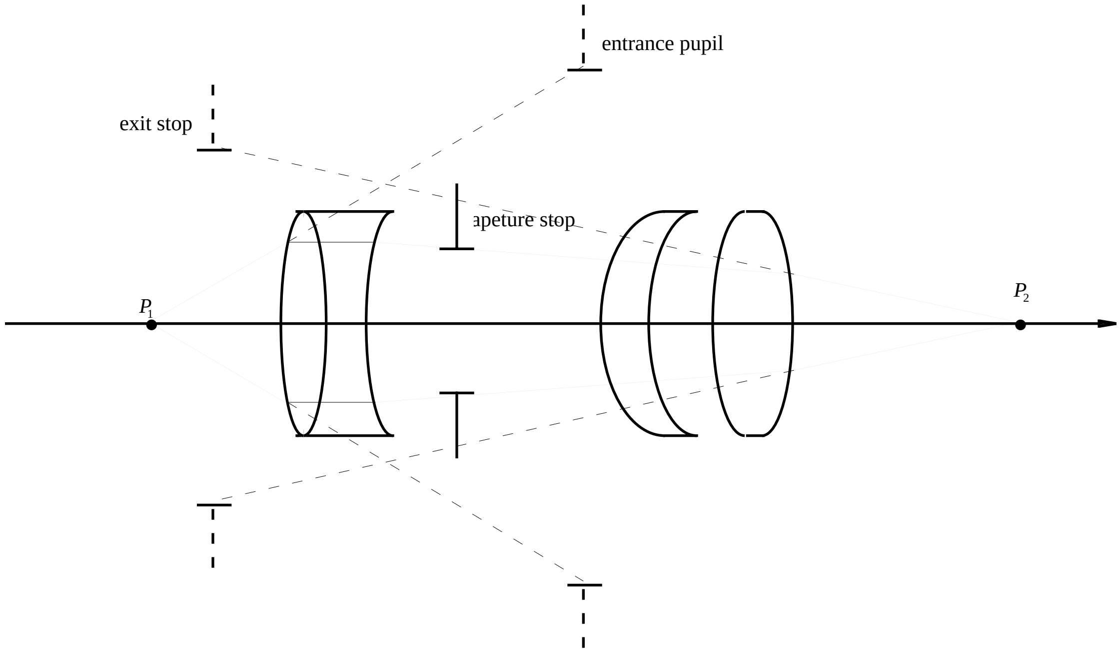
Note that in many optical systems it may not be a specific stop which limits the size of the light cone but the diameter of the lens itself as shown below.



Example of the apeture stop in object space.



In a compound lens system it is possible that the physical apeture stop will not be in image or object space as shown below.



(Note how large the entrance and exit pupils are for such a short system.) In multi-lens systems it is often difficult to determine the entrance or exit pupil at first glance; however, a systematic method of determining the limiting apeture is to image all stops and lens rims through the system: to the left to locate the entrance pupil, to the right to locate the exit pupil. The stop or lens rim allowing the smallest cone of rays through the system from the image point will be the appropriate entrance or exit pupil and the corresponding optical element will be the apeture stop.

The graphical techniques we have been using for the past several pages become rather complex in multi-lens systems such as the one on the previous page. A ray matrix analysis of the optical system may greatly simplify the process of determining the apeture stop and the corresponding pupils. Consider the general ray matrix transformation

$$\begin{bmatrix} r_2' \\ r_2 \end{bmatrix} = \begin{bmatrix} b & -a \\ -d & c \end{bmatrix} \begin{bmatrix} r_1' \\ r_1 \end{bmatrix}$$

where $n_2' = n_1' = 1$. Equivalently this may be written as

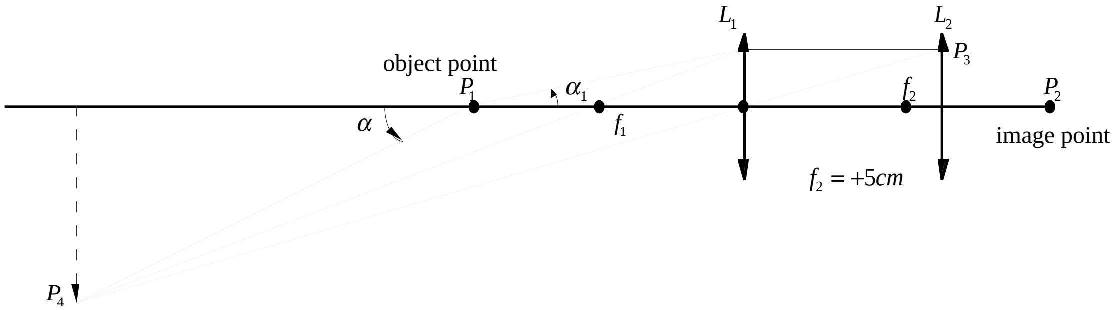
$$r_2' = br_1' - ar_1$$
$$r_2 = -dr_1' + er_1$$

To determine the apeture stop we take an object point on the optical axis so that $r_1 = 0$. Let the cone half-angle at P_1 be α_0 and let the radius of the apeture stop be ρ. Then the second transformation yields

$$\rho = -d\alpha_0 + c \times 0 = -d\alpha_0$$

or $\alpha_0 = -\dfrac{\rho}{d}$. The apeture stop for the system will have the smallest $\dfrac{\rho}{d}$ ratio of all stops and lens rims in the system. Note that d will be a component of the ray matrix relating P_1 to the lens or stop; not the matrix relating P_1 to P_2 (the object point).

Example:



The angle subtended by L_1 is $\alpha_1 = \frac{1cm}{10cm} = 0.1rad$. The image of L_2 through L_1 is located $x_1 = -\frac{f_2^2}{x_2} = -\frac{(5)^2}{+1} = -25cm$ locating the image of L_2 20cm to the left of P_1 as drawn above. To compute its size recall that between conjugate planes H_1 and H_2

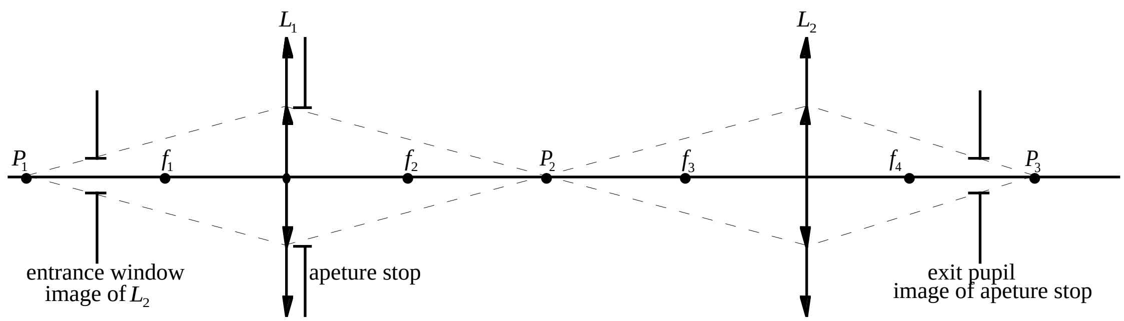
$$S_{H_1H_2} = \begin{bmatrix} \frac{1}{\beta} & -a \\ 0 & \beta \end{bmatrix}$$

where $a = \frac{1}{f_2}$ and $\beta = 1 - as_2$. Identifying s_2 as +30cm (we are imaging from right to left — point P_3 to P_4) we can write $\beta = 1 - \left(\frac{1}{5}\right)(+30) = -5$. The image then occludes the angle $\alpha_2 = \frac{5cm}{20cm} = \frac{1}{4}rad$. We could also have solved this by writing

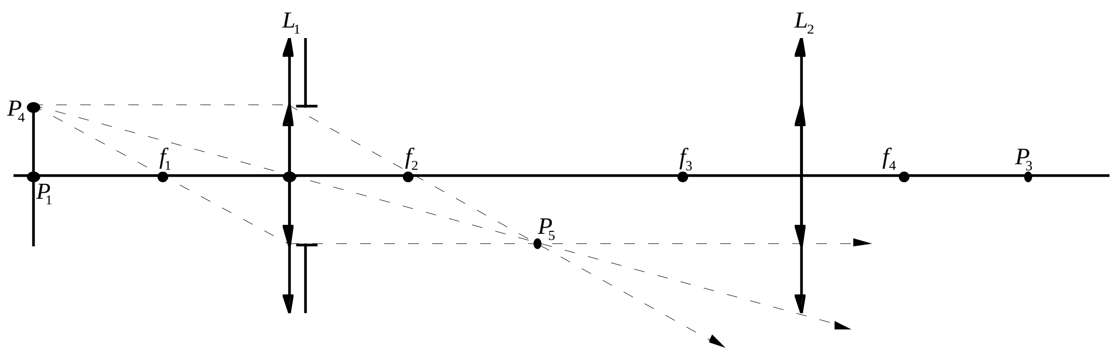
$$S_{P_3P_4} = \begin{bmatrix} 1 & 0 \\ 6 & 1 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{5} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 10 & 1 \end{bmatrix} = \begin{bmatrix} -1 & -\frac{1}{5} \\ +4 & -\frac{1}{5} \end{bmatrix}$$

Thus, $\alpha_2 = -\frac{\rho}{d} = -\frac{1cm}{4cm} = -\frac{1}{4}rad$. The minus sign here simply indicates that the image of P_3 through L_1 is inverted; however, we are only concerned with the magnitude of α_2 and not its sign. Irregardless of the method we reach the result that L_1 is the apeture stop.

The apeture stop determines the transmitted light cone for an on-axis point object, the field stop determines the transmitted light cone for off-axis points. A more formal definition is that the field stop limits the cone formed by the chief rays. The image of the field stop in object space is the entrance window); the image in image space is the exit window,



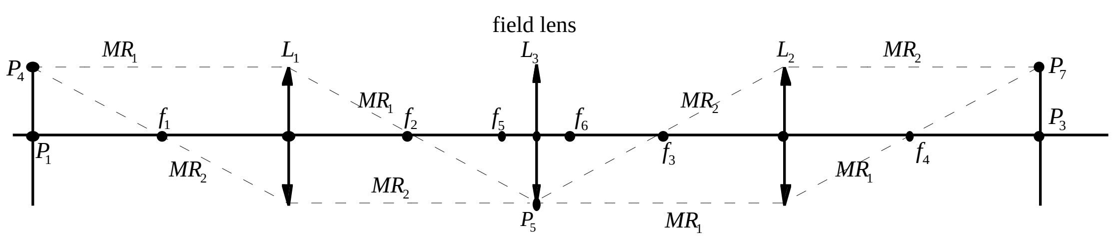
(a) on-axis imaging



(b) off-axis imaging

Lens L_2 will be the field stop since the light cone from P_4 is not intercepted by L_2. In fact, for P_4 as drawn, the light cone completely missed L_2. Since L_2 is the field stop its image in object space will be the entrance window and its image in image space the exit window. From the diagram L_2 is the exit window and the entrance window is as indicated. Note that at P_1 the entire light cone allowed by the apeture stop behind L_1 is transmitted through L_2. At P_4 the light cone transmitted by L_1 is totally blocked by the field stop — the rim of L_2; however, for P_6 located between P_1 and P_4 only part of the light cone passed by the apeture stop will be transmitted by L_2. This phenemona is known as vignetting.

Consider the effect of a stop located at P_2. If the image of this new stop subtends a smaller angle (as seen from the center of L_1) than L_2 does it will be the new field stop. It will reduce the field of view in the object plane, but does not remove the vignetting. To correct the vignetting we use a lens called a field lens at P_2. A lens located there will not effect the imaging of the on-axis point P_1 since because of the symmetry of the system L_3 simple images P_2 onto itself.



As a further argument that this is so consider the transformation of a thin lens

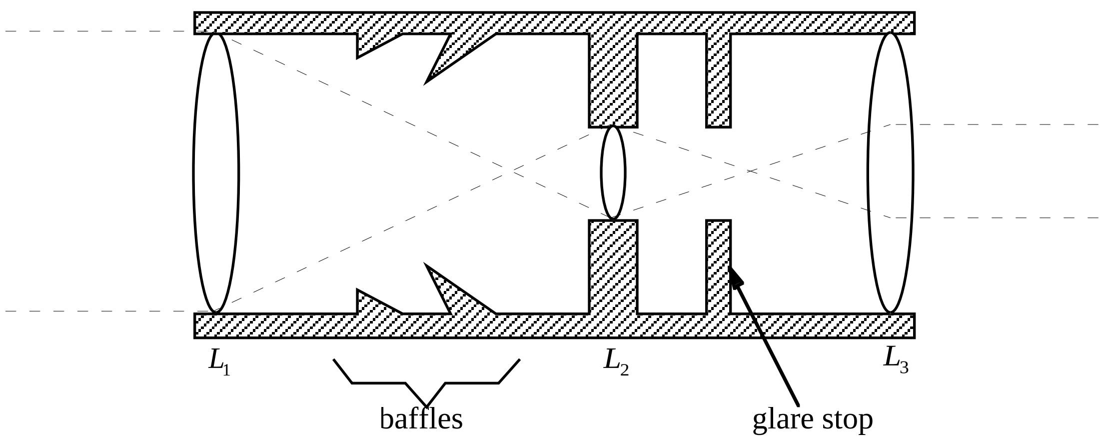
$$r_2' = r_1' - \frac{1}{f_6} r_1$$

$$r_2 = r_1$$

Since $r_1 = 0$ we have $r_2 = r_1 = 0$ and $r_2' = r_1'$ indicating that the light cone from P_1 is not altered by L_3. At this point we ask ourselves what L_3 does in the the optical system. To answer this L_3 deviates the entire light cone from P_4 onto L_2. Lens L_3 images L_1 onto L_2 if f_5 corresponds to f_2 and f_6 to f_3 thus forming a $\beta = -1$ imaging system between L_1 and L_2. Because of the total transmission of the light cone from P_4 to P_7 lens L_3 has completely eliminated vignetting. If P_4 were a slight distance further radially from P_1 no light from P_4 would reach L_2. This indicates that L_3 is the field stop for the system.

In the above example, the field lens was easily added; in more complex systems containing compound lenses it may be impossible to add a field lens. In fact, the concept of a field stop is really not too relevant to systems containing compound lenses.

Baffles may often be used in optical systems to prevent stray light from reaching the image. As an illustration consider the telescopic system shown below.



Lens L_2 is a field lens. L_1 functions as the apeture stop. Light may enter L_1 at an angle, reflect off the walls of the lens housing and be reflected or scattered into the image. One method of preventing this is to insert a glare stop which images through L_2 matches the apeture of L_1. In addition, baffles may be inserted outside the optical patch to "catch" stray light. As a further precaution all stops, baffles, etc. should be highly absorbent to suppress reflections, q.v., the use of flat black paint inside telescopes.

From page 1 we know that geometric optics breaks down when we attempt to image objects less than a few tenths of a millimeter in size. For objects smaller than this the image tends to blue and become fuzzy because of diffraction effects. If the only limit upon the image quality is diffraction then we say that the lens or optical system is diffraction limited. Such optics are very good optics. In most systems other limitations upon image quality are given the general name aberrations. Some aberrations may be due to inhomogenities in the glass, surface defects, or improper grinding. They are called irregular aberrations ince they are not subject to rogorous analysis except statistically. Certain other aberrations called regular aberrations are due to the breaking down of the paraxial ray approximation. An excellent example of how this affects our geometric optics results to this point may be fund on page 12. The refraction of a ray by a curved dielectric surface is given by

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

with the corresponding paraxial ray expression being

$$n_2 r_2' = n_1 r_1' - \left(\frac{n_2 - n_1}{R}\right) r_1$$

Note that it has been assumed that $\cos\beta_1 \approx \cos\beta_2 \approx 1$ and $\cos\theta_1 \approx \cos\theta_2 \approx 1$.

As soon as we begin working with large aperture lens this approximation beguns to break down as we deal with larger angles. In general

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \ldots$$

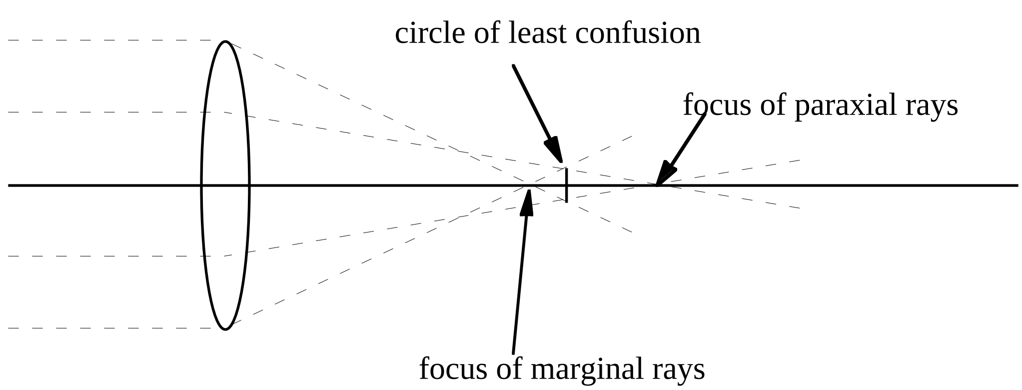
Regular aberrations are those deviations from paraxial ray results that may be predicted using higher order terms of the trigonometric expansions.

The results that we have obtained using the paraxial ray approximation are said to follow first-order theory, i.e., they include only the first terms in the trigonometric expansions. In almost every problem no more than the first five terms will ever be included and usually it is only the first two or three.

We shall not attempt to develop a rigorous theory of aberrations but will, instead, simply define the aberration, how it effects the imaging of the lens (or optical system, although we shall stick to describing the aberrations of a single lens) and how it may be corrected for.

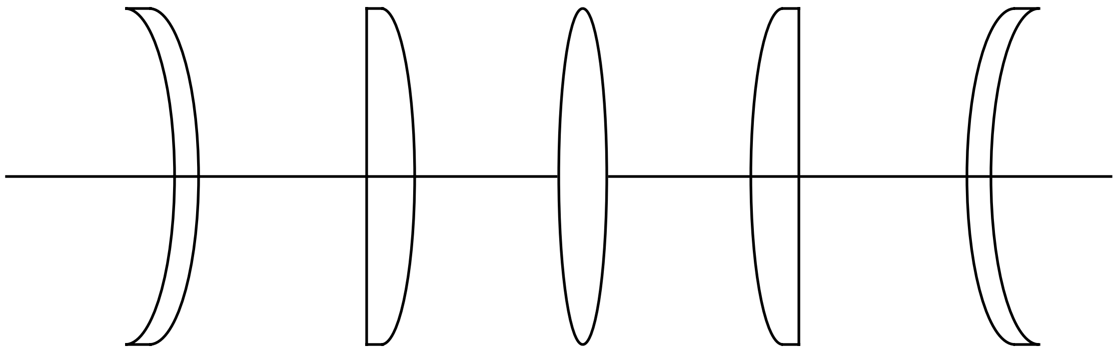
<u>Spherical Aberration</u>

This aberration is based directly upon the breakdown of the paraxial ray approximation for on-axis objects. A zone of a circularly symmetric lens is simply a region of the lens surface bounded by r and $r + dr$ where r is less than the lens radius. We may now define spherical aberration as the result of rays passing through different zones of a lens being focused to different points. Recall that marginal rays are those rays passing through the boundary of the lens. If the rays close to the optical axis are focused further away from the lens than marginal rays the lens is said to show positive longitudinal spherical aberration and the lens is undercorrected for spherical aberration.



Longitudinal spherical aberration is the distance between the focal points for paraxial and for marginal rays. Lateral spherical aberration is the spot size corresponding to the marginal rays in a plane perpendicular to the optic axis and containing the paraxial ray focal point.

Although we shall not show it spherical aberration varies according to how much the lenses are bent. One formal measure of this bending is the Coddington shape factor $\sigma = \dfrac{R_1 + R_2}{R_1 - R_2}$ where R_1 and R_2 are the signed radii of curvature of the left and right surfaces of the lens respectively (light traveling from left to right). Some typical lens shapes and their Coddington shape factors are given below.



$\sigma = -3$ $\sigma = -1$ $\sigma = 0$ $\sigma = +1$ $\sigma = +3$

Note: all of the above lenses are of equal diameter and equal paraxial ray focal length but have different spherical aberration. Spherical aberration is also a function of the image and object distances so we define the Cossington position factor $\pi = \dfrac{S_1 + S_2}{S_1 - S_2}$ where S_1 and S_2 are the signed object and image distances respectively.

Spherical aberration cannot be eliminated from a single lens but it can be minimized if the Coddington shape and position factors satisify

$$\sigma = -2\left(\frac{n^2 - 1}{n + 2}\right)\pi$$

where n is the index of refraction of the lens material.

To design a lens that has minimal spherical aberration we first determine the desired image-object ration which then fixes f and π. With these results we may then determine the desired shape of the lens. A useful set of equations for doing this may be obtained by substituting the expressions for π and σ into the lens law:

$$\frac{1}{S_2} - \frac{1}{S_1} = \frac{1}{f_2}$$

$$\pi = \frac{S_1 + S_2}{S_1 - S_2} = \frac{\dfrac{1}{S_1} + \dfrac{1}{S_2}}{\dfrac{1}{S_1} - \dfrac{1}{S_2}} = \frac{\dfrac{1}{S_2} - \dfrac{1}{f_2} + \dfrac{1}{S_2}}{\dfrac{1}{S_2} - \dfrac{1}{f_2} - \dfrac{1}{S_2}} = 1 - \frac{2 f_2}{S_2}$$

$$\pi = 1 - 2 f_2\left(\frac{1}{S_2}\right) = 1 - 2 f_2\left(\frac{1}{S_1} + \frac{1}{f_2}\right) = -1 - \frac{2 f_2}{S_1}$$

$$\therefore \pi = 1 - \frac{2 f_2}{S_2} = -1 - \frac{2 f_2}{S_1}$$

It is more complicated to relate σ to the lens parameters:

$$\frac{1}{S_2}-\frac{1}{S_1}=\frac{1}{f_2}=(n_2-n_1)\left(\frac{1}{r_1}-\frac{1}{r_2}\right)=\Delta n\left(\frac{r_2-r_1}{r_1r_2}\right)$$

where $\Delta n = n_2 - n_1$. Since $\sigma = \frac{r_2+r_1}{r_2-r_1}$, or $r_2 - r_1 = \frac{r_2+r_1}{\sigma}$ we can write

$$\frac{1}{f_2}=\Delta n\left(\frac{r_2-r_1}{r_1r_2}\right)=\frac{\Delta n}{\sigma}\left(\frac{r_2+r_1}{r_1r_2}\right)=\frac{\Delta n}{\sigma}\left(\frac{r_2-r_1}{r_1r_2}+\frac{2r_1}{r_1r_2}\right)$$

$$\frac{1}{f_2}=\frac{\Delta n}{\sigma}\frac{r_2-r_1}{r_1r_2}+\frac{2\Delta n}{\sigma r_2}=\frac{1_1}{\sigma f_2}+\frac{2\Delta n}{\sigma r_2}$$

Solving for r_2 gives $r_2 = \frac{2\Delta n f_2}{\sigma - 1}$

In like fashion

$$\frac{1}{f_2}=\Delta n\left(\frac{r_2+r_1}{r_1r_2}\right)=\frac{\Delta n}{\sigma}\left(\frac{2r_2-r_2+r_1}{r_1r_2}\right)=\frac{2\Delta n}{\sigma r_1}-\frac{\Delta n(r_2-r_1)}{\sigma r_1r_2}=\frac{2\Delta n}{\sigma r_1}-\frac{1}{f_2\sigma}$$

gives

$$\sigma=\frac{2\Delta n f_2}{r_1}-1 \text{ or } r_1=\frac{2\Delta n f_2}{\sigma+1}$$

Combining these results gives

$$\frac{r_1}{r_2}=\frac{\sigma-1}{\sigma+1}$$

To illustrate how to use these results consider the problem of determining the radii of curvature for a lens of $f_2 = +10cm$, $n = 1.5$ which has minimum spherical aberration for parallel incident light. Since the incident rays are parallel to the optic axis they are focused to the focal point. In this situation as we argued that $S_1 \to -\infty$ and $S_2 \to +10cm$ as per page 19 of these notes. The Coddington position factor is then

$$\pi=\frac{S_2+S_1}{S_2-S_1}=\frac{10-\infty}{10+\infty}\to -1$$

The shape factor which minimizes the spherical aberration is given by

$$\sigma=-2\left(\frac{n^2-1}{n+2}\right)\pi=2\left(\frac{2.25-1}{1.5+2}\right)(-1)=0.714$$

The radii of the lens surfaces will then be

$$r_1=\frac{2\Delta n f_2}{\sigma+1}=\frac{2(1.5-1)(10)}{0.714+1}=+5.83cm$$

and

$$r_2=\frac{2\Delta n f_2}{\sigma-1}=\frac{2(1.5-1)(10)}{0.714-1}=-35cm$$

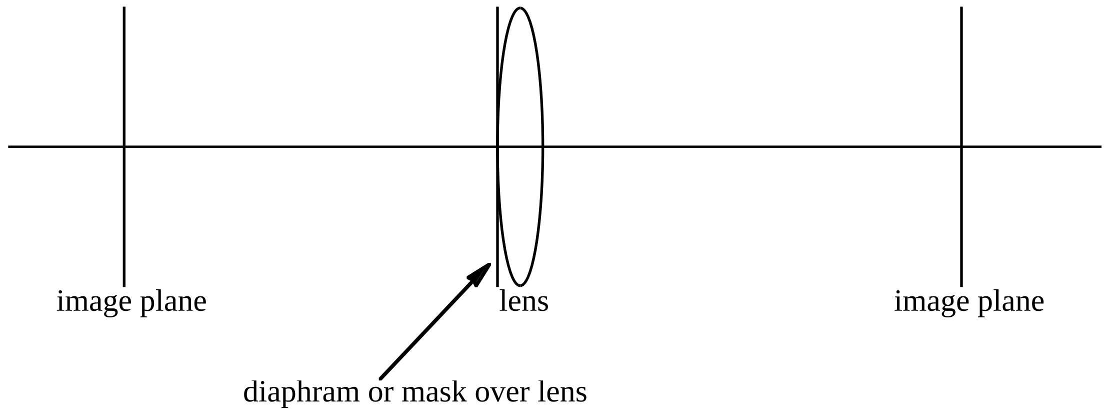
Thus, the final lens is a menicus.

Rigorous aberration theory shows how spherical aberration can be eliminated entirely for multiple-lens systems although we shall not pursue the subject in any greater depth.

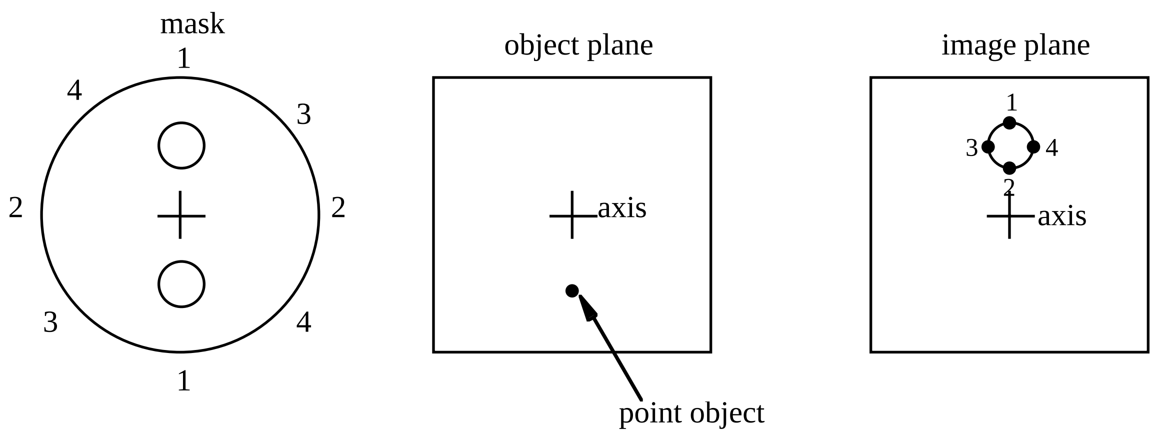
Coma

Coma refers to the comet-like appearance of the image of a point object that is located off the axis. Coma occurs when the incident rays make an angle with the optic axis; spherical abberation occured for an on-axis object and for rays parallel to the optic axis.

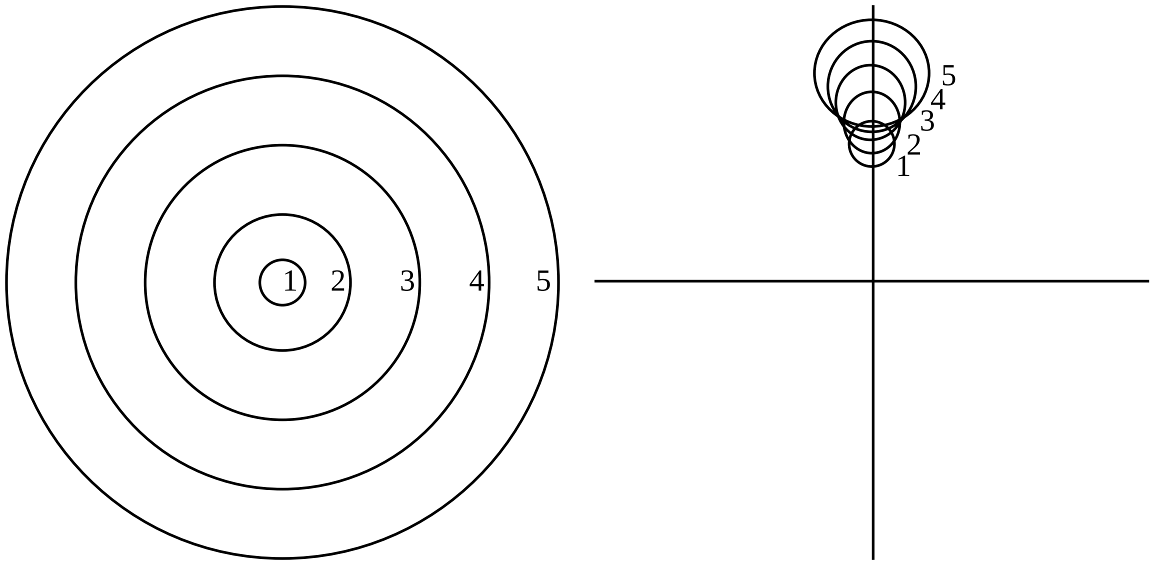
No expressions for coma will be developed; instead we will examine the results of an experim,ent. Consider the optical system shown below



where the image plane contains an off-axis point object. Let the mask be as shown below.



The mask is free to rotate about the optic axis. If the diaphram is oriented so that the holes correspond to the 1's the imahe point at 1 in the image place will result. If the mask is rotated so the holes correspond to the 2's the resultant point image will be at point 21 in the image place. Image points 3 and 4 are formed by orienting the mask holes to correspond with the mask positions 3 and 4. Note that the image points form a circular ring. This is called the comatic circle. The circle described by the maks holes is a zone of the lens. In general, zones of different radii will produce concentric circles of differing radius and location in the image place. The radius of the comatic circle is proportional to the square of the radius of the corresponding lens zone and is the distance between the center of the comatic circle and the optic axis. The sum total of all comatic circles, i.e., the total l,ight passing through the lens produces the characteristic comatic flare as shown below.

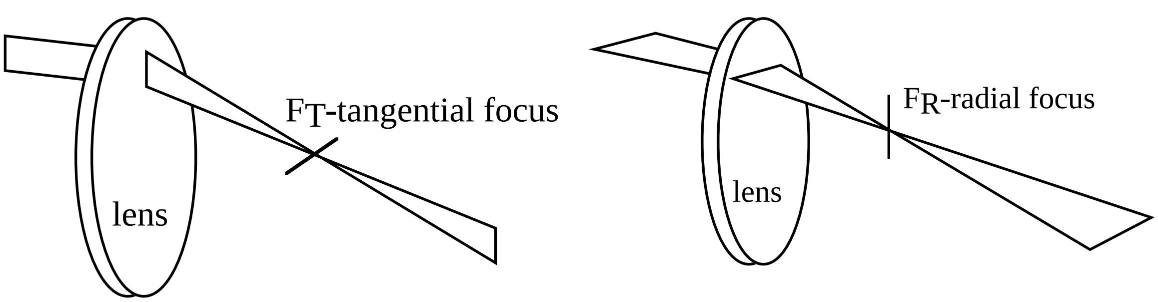


lens zones | comatic circles

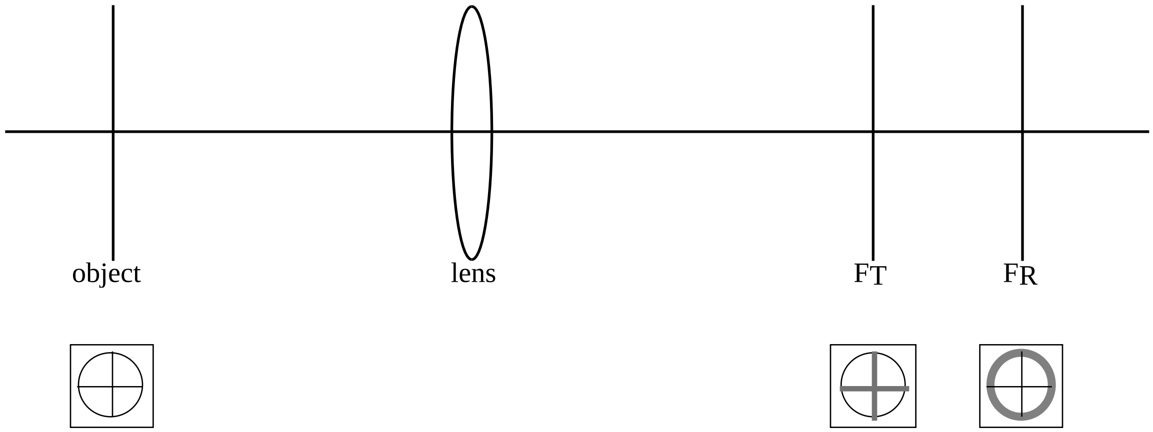
Note that coma is a function of the angle of obliquity — it is not present for point objects <u>on</u> the optic axis. A general condition for eliminating coma is to have $\frac{\sin \gamma_1}{\sin \gamma_2}$, where γ_1 is the slope angle at the object and γ_2 is the slope angle at the image, equal a constant. This relation must be true for all values of γ_1 and over the entire system aperture to completely eliminate coma. A system free of both coma and spherical aberration is called aplanatic.

<u>Astigmatism</u>

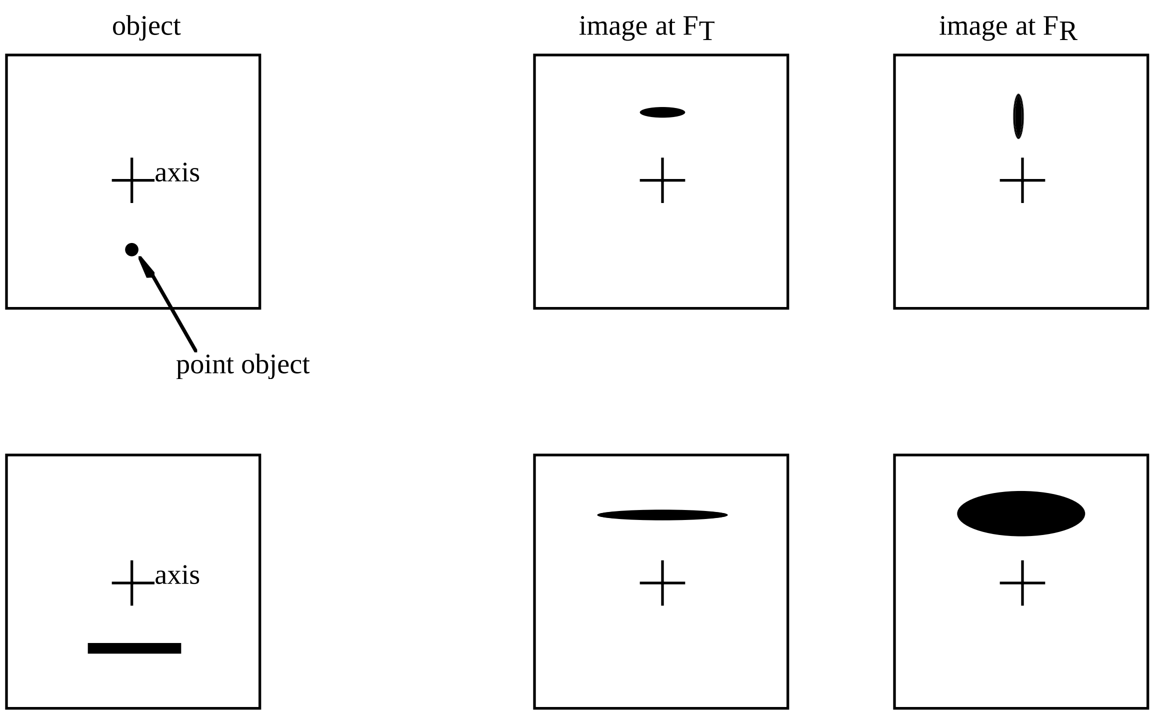
Consider a bunde of light rays of circular cross-section incident on a spherical lens surface some distance away from the optic axis. The projection of the circular bundle on the spherical surface will be an ellipse with its major axis along a lens radius and the minor axis perpendicular to this radius. Astignatism, then, is that property of a lens to focus rays along the major axis and the minor axis to different points. These points are called tangential and radial focal points.



The classical object used to illustrate astignatism is a spoked wheel as the object to be imaged. The spokes being oriented radially will focus to the place F_T, the rim being oriented prependicular to the radius will focus to a plane F_R.

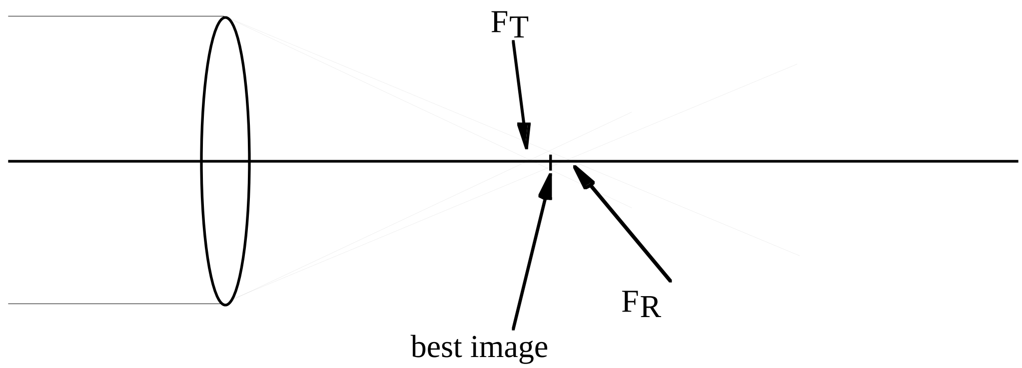


Some other examples are



The distance between F_T and F_R is the astigmatic interval or the Interval of Sturm. The "best" image is somewhere between F_T and F_R as we will illustrate with an example.

<u>Example:</u>
A 70mm diameter lens has a tangential focal length of 16.7cm and a radial focal length of 18.5cm.



The refractive powers corresponding to these focal lengths are $P_T = \frac{1}{F_T} = 6m^{-1}$ and

$P_R = \frac{1}{F_R} = 5.4m^{-1}$. The circle of least confusion is located at the "dioptic midpoint," i.e.,

$$P_{AVG} = \frac{6+5.4}{2} = 5.7m^{-1}.$$

That is, the circle of least confusion is located $\frac{1}{5.7m^{-1}} = 17.5cm$ from the lens. Notice that this is not simple the average of F_T and F_R which would be $17.6cm$. The diameter of the circle of least confusion is found by using similar triangles

$$\frac{70}{18.5} = \frac{d}{18.5-17.5}$$

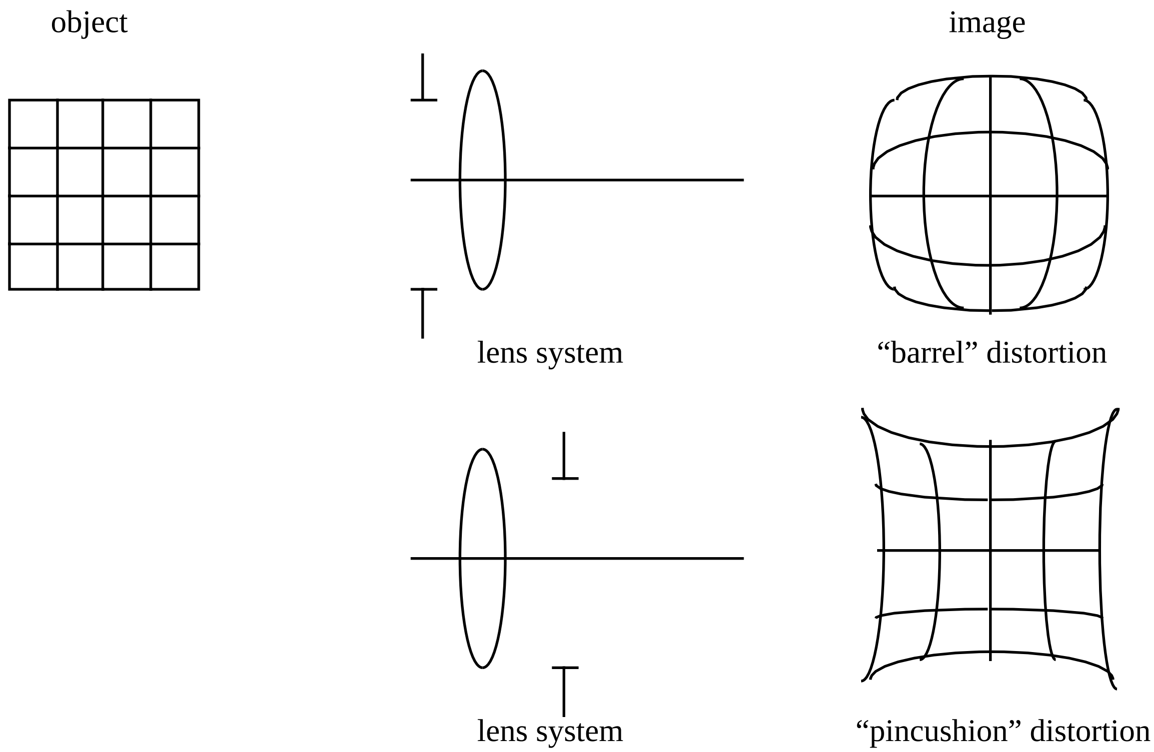
to get $d = 3.78mm$

What shall not deal with eliminating astigmatism in any detail. We note, however, that we have image formation in two surfaces — one corresponding to tangential and one to radial object details. The correction of astigmatism then lies in somehow bringing these two surfaces together. The details of doing this are beyond the scope of this course but the result is that two lenses slightly separated with an aperture stop placed between them can minimize astigmatism. It is not possible to correct for astigmatism with a single lens.

Curvature of Field & Distortion

These are two other aberrations related to astigmatism. Curvature of field occurs when a place object is not imaged into a place surface. The difference between astigmatism and curvature of field is that astigmatism varies with object distance while curvature of field does not.

Distortion has a very specific meaning in optics — the aberration due to the dependence of the transverse linear magnification upon the distance of the object point from the optic axis. This effect causes square objects to resemble barrels or pincushions and may be minimized by using a symmetrical doublet compound lens with a central stop.



Chromatic aberration

For a simple lens light of different wavelengths will have different focal points. The reason for this is a phenemona called dispersion — the property of a material that its index of refraction is a function of wavelength. Air has very little dispersion for visible light; vacuum, none at all. In visible optics it is customary to specify a material's index of refraction at three wavelengths rather than in graphic or functional terms. The three wavelengths most often used are the Fraunhofer lines of the solar spectrum: the blue F line at $\lambda \sim 486nm$, the yellow sodium D line at $\lambda \sim 589nm$, and the red C line at $\lambda \sim 656nm$. Let the corresponding indices of refraction be n_F, n_D, and n_C. The dispersion across the visible spectrum is $n_F - n_C$ and the dispersive power is defined as $\Delta = \frac{n_F - n_C}{n_D - 1}$. A low dispersion glass such as crown glass has a $\Delta \sim 0.020$ and a highly dispersive glass such as flint glass has a $\Delta \sim 0.033$. In terms of a simple lens recall that the focal point is given by

$$\frac{1}{f_2} = \left(n_2 - n_1\right)\left(\frac{1}{r_1} - \frac{1}{r_2}\right)$$

where r_1 and r_2 are the signed radii of curvature of the first and second surfaces respectively, n_1 is the index of refraction of the medium surrounding the lens, and n_2 is the refractive index of the lens material. This indicates that f_2 does indeed depend upon n_2. For a simple positive lens the focal length is shorter for blue light than for red light.

Chromatic aberrationm ay be further catagorized as longitudinal and lateral chromatic aberration. Longitudinal chromatic aberration is the distance between the image planes corresponding to the different wavelengths of light; lateral chromatic aberration is the difference in the location of the images from the optic axis. Chromatic aberration is called positive if the blue focus lies clsoer to the lens than the red.

Lenses corrected for chromatic aberration at two different wavelengths are called achromats; lens which correct at three wavelengths are called apochromats. The specific wavelengths for which the lens is corrected will depend upon the application. For example, film is more blue sensitive than the human eye so camera lenses are corrected for blues and blue-greens. The human eye's peak response is in the yellow-green so visual instruments are corrected for yellow-greens.

There are two methods of correcting for chromatic aberration. The first is to use a compound lens — a doublet — made of a flint glass lens and a crown glass lens placed in contact. The principle of operation is that the dispersion produced in one lens is canceled by the opposite dispersion of the other. Without derivation, if the power of the desired lens combination is P_A (A for achromatic) and Δ_1 and Δ_2 are the dispersions of the first and second lens respectively and P_1 and P_2 their respective power we have

$$P_1 = P_A \frac{\Delta_2}{\Delta_2 - \Delta_1}$$

and

$$P_2 = -P_A \frac{\Delta_1}{\Delta_2 - \Delta_1}$$

Example:
with the following data on crown and flint glass design an achromatic doublet with a +10cm focal length.

	n_F	n_D	n_C
crown	1.53	1.523	1.52
flint	1.63	1.62	1.61

For the crown glass,

$$\Delta_1 = \frac{n_F - n_C}{n_D - 1} = \frac{1.53 - 1.52}{1.523 - 1} = 0.019$$

and for the flint

$$\Delta_2 = \frac{n_F - n_C}{n_D - 1} = \frac{1.63 - 1.61}{1.62 - 1} = 0.032$$

The powers are then

$$P_1 = (10)\frac{0.032}{0.032 - 0.019} = 24.6m^{-1}$$

and

$$P_2 = -(10)\frac{0.019}{0.032 - 0.019} = -14.6m^{-1}$$

For a thin lens $P_{lens} = P_1 + P_2 = 24.6 - 14.6 = 10.0m^{-1}$

and since $P_{lens} = \frac{1}{f} = \frac{1}{0.1m} = 10m^{-1}$ we have our desired lens.

Knowing the focal lengths we may now determine the lens curvatures from the lensmaker's formula

$$P = \Delta n\left(\frac{1}{r_1} - \frac{1}{r_2}\right)$$

Suppose we let the first lens be a symmetric doublet, i.e., $r_1 = +R$ and $r_2 = -R$, so that

$$24.6m^{-1} = (1.523-1)\left(\frac{1}{R}+\frac{1}{R}\right)$$

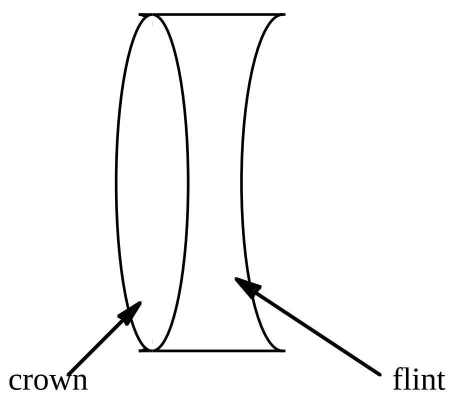
which gives

$$R = +4.25cm$$

The third surface will have r_3 given by

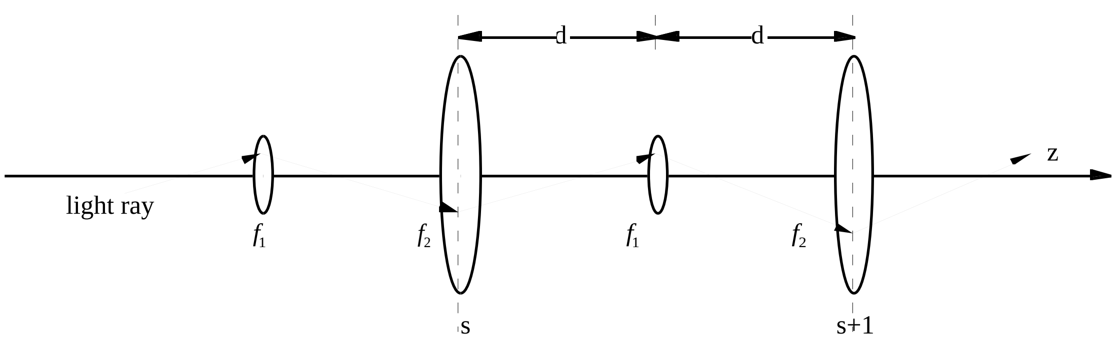
$$-14.6m^{-1} = (1.62-1)\left(-\frac{1}{R}+\frac{1}{r_3}\right)$$

or $r_3 = +20cm$. The resulting lens is of the form

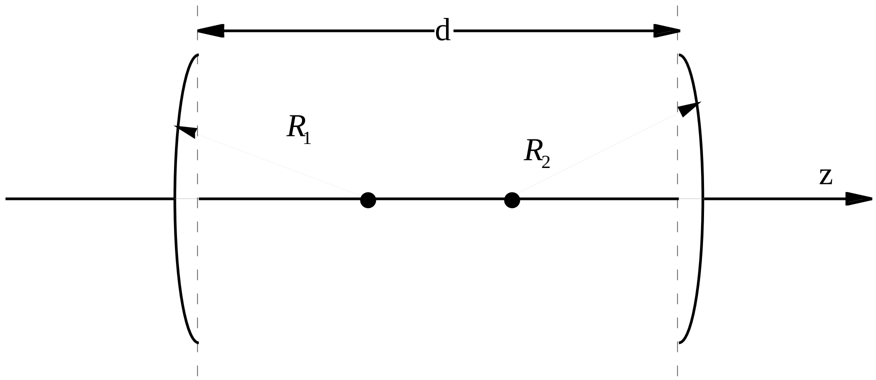


A second method of correcting for chromatic aberration is to use two positive lenses made of the same material and separated by a distacne equal to one-half the sum of their individual focal lengths, i.e., $d = \frac{1}{2}(f_1 + f_2)$. This is called a spaced doublet and its principle of operation will not be discussed here.

Propagation of a ray through a biperiodic lens sequence, i.e., a series of lenses of alternating focal lengths f_1 and f_2 separated by a distance d. Such a lens structure is called a lens waveguide and has been used as a guide structure for light. This lens system is also formally equivalent to an optical resonator formed by mirrors of radii $R_1 = 2f_1$ and $R_2 = 2f_2$, i.e., a laser cavity.



(a) light rays in a biperiodic lens sequence



(b) light rays in an optical resonator

For the lens sequence we will write the system matrix between planes s and $s+1$

$$\begin{bmatrix} r_{s+1}' \\ r_{s+1} \end{bmatrix} = \begin{bmatrix} 1 & -\frac{1}{f_2} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ d & 1 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{f_1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ d & 1 \end{bmatrix} \begin{bmatrix} r_s' \\ r_s \end{bmatrix}$$

$$\begin{bmatrix} r_{s+1}' \\ r_{s+1} \end{bmatrix} = \begin{bmatrix} 1-\frac{d}{f_2} & -\frac{1}{f_2} \\ d & 1 \end{bmatrix} \begin{bmatrix} 1-\frac{d}{f_1} & -\frac{1}{f_1} \\ d & 1 \end{bmatrix} \begin{bmatrix} r_s' \\ r_s \end{bmatrix}$$

$$\begin{bmatrix} r_{s+1}' \\ r_{s+1} \end{bmatrix} = \begin{bmatrix} \left(1-\frac{d}{f_2}\right)\left(1-\frac{d}{f_1}\right)-\frac{d}{f_2} & -\frac{1}{f_1}-\frac{1}{f_2}+\frac{d}{f_1 f_2} \\ d\left(2-\frac{d}{f_1}\right) & 1-\frac{d}{f_1} \end{bmatrix} \begin{bmatrix} r_s' \\ r_s \end{bmatrix} \qquad (1)$$

$$\begin{bmatrix} r_{s+1}' \\ r_{s+1} \end{bmatrix} = \begin{bmatrix} D & C \\ B & A \end{bmatrix} \begin{bmatrix} r_s' \\ r_s \end{bmatrix} \qquad (2)$$

In (2), A, B, C and D are NOT the Gaussian coefficents we have been working with. This particular notation has been chosen to be consistent with Yariv, Introduction to Optical Electronics. In the manner of Yariv, we may write (2) out for clarity as

$$r_{s+1}' = Dr_s' + Cr_s \qquad (3)$$

$$r_{s+1} = Br_s' + Ar_s \qquad (4)$$

From (4), $r_s' = \frac{1}{B}\left(r_{s+1} - Ar_s\right)$

Since this result must be true for all unit cells (basic lens units) of our system this result must also hold true for the transformation between planes $s+1$ and $s+2$, i.e.,

$$r_{s+1}' = \frac{1}{B}\left(r_{s+2} - Ar_{s+1}\right)$$

Substituting from (3),

$$Cr_s + Dr_s' = \frac{1}{B}\left(r_{s+2} - Ar_{s+1}\right)$$

and again from (4)

$$Cr_s + \frac{D}{B}(r_{s+1} - Ar_s) = \frac{1}{B}(r_{s+2} - Ar_{s+1})$$
$$BCr_s + Dr_{s+1} - ADr_s = r_{s+2} - Ar_{s+1}$$
and after rearranging
$$r_{s+2} - (A+D)r_{s+1} + (AD - BC)r_s = 0$$

Recalling that the determinant of a ray matrix is always 1 we have from (2) that $AD - BC = 1$. We also define $b = \frac{A+D}{2}$ where b is NOT a Gaussian coefficent but only a variable in the problem. Using these definitions and results we get
$$r_{s+2} - 2br_{s+1} + r_s = 0 \qquad (5)$$
To determine the solution of (5) we must first develop some ideas from what is called finite difference calculus. Define
$$\Delta r_s = r_{s+1} - r_s \qquad (6)$$
This is the first forward difference and may be likened to a derivative since here $\Delta s = (s+1) - s = 1$ and $\frac{\Delta r_s}{\Delta s} = \Delta r_s$. This is called finite difference calculus since $\Delta s = 1$ whereas in ordinary differential calculus we would take the limit as $s \to \infty$, i.e., $\lim_{s\to\infty} \frac{\Delta r_s}{\Delta s} = r_s'$. In a similar fashion to (6) we may define the second forward difference
$$\Delta^2 r_s = \Delta(\Delta r_s) = \Delta r_{s+1} - \Delta r_s = r_{s+2} - r_{s+1} - r_{s+1} + r_s$$
$$\Delta^2 r_s = r_{s+2} - 2r_{s+1} + r_s \qquad (7)$$
Rewriting (5) using (7) we get
$$\Delta^2 r_s + (2 - 2b)r_{s+1} = 0 \qquad (8)$$
This is analagous to the ordinary differential equation $\frac{d^2 f}{dx^2} + af = 0$ where a is a constant. which has solutions
$$f = c_1 e^{i\sqrt{a}x} + c_2 e^{-i\sqrt{a}x}$$
By analogy we may try a solution to (8) of the form $r_s = r_o e^{isq}$ where s is the independent variable and r_o and q are constants to be determined. Substituting this solution into (7) we get $r_o e^{i(s+2)q} - 2br_o e^{i(s+1)q} + r_o e^{isq} = 0$

This is a quadratic in e^{iq} and has solutions
$$e^{iq} = b \pm \sqrt{b^2 - 1} \qquad (9)$$
At this point let us define $b = \cos\theta$ so that $\sqrt{b^2 - 1} = i\sin\theta$ and (9) becomes
$e^{iq} = \cos\theta \pm i\sin\theta = e^{\pm i\theta}$

The general solution to (5) is then
$$r_s = c_1 e^{is\theta} + c_2 e^{-is\theta} \qquad (10)$$
where c_1 and c_2 are constants to be determined by the initial conditions. Equation (10) may also be written in the form
$$r_s = r_{\max} \sin(s\theta + \alpha) \qquad (11)$$
where $r_{\max}$ and α are found from the initial ray displacement and slope.

For this to be a real lens waveguide it is necessary that the light rays in the system always be incident upon the finite sized lenses. This requires that θ in (10) be real. For if θ were complex or imaginary we would have unbounded hyperbolic solutions. [Example: in (11) if $\alpha = 0$ and θ is complex, i.e., $\sin(si\theta') = \sinh(s\theta')$] For θ to be real it is necessary that $1 - b^2 \geq 0$ [from $b^2 - 1 = i\sin\theta$] or

$$b^2 \leq 1 \qquad (12)$$

From this and with the definition of b we get, after some algebra,

$$1 \geq 1 - \frac{d}{2f_1} - \frac{d}{2f_2} + \frac{d^2}{4f_1f_2} \geq 0$$

$$1 \geq \left(1 - \frac{d}{2f_1}\right)\left(1 - \frac{d}{2f_2}\right) \geq 0$$

The factors $1 - \frac{d}{2f_1}$ and $1 - \frac{d}{2f_2}$ are often called stability factors and are usually defined as

$$\left.\begin{aligned} g_1 &= 1 - \frac{d}{2f_1} \quad (R_1 = 2f_1) \\ g_2 &= 1 - \frac{d}{2f_2} \quad (R_2 = 2f_2) \end{aligned}\right\} \qquad (14)$$

This result is more general than it seems and we shall return to it when we discuss optical resonators. A plot of the condition (13) [$0 \leq g_1g_2 \leq 1$] is given below.