EUGENE HECHT/ALFRED ZAJAC Adelphi University

OPTICS

ADDISON-WESLEY PUBLISHING COMPANY

READING, MASSACHUSETTS · MENLO PARK, CALIFORNIA · LONDON · DON MILLS, ONTARIO

This book is in the
ADDISON-WESLEY SERIES IN PHYSICS

 Library of Congress Catalog Card No. 79-184159.

ISBN 0-201-02835-2
ABCDEFGHIJ-HA-79876543

Contents

Geometrical Optics— Paraxial Theory 5

5.1 INTRODUCTORY REMARKS

Suppose that we have an object which is either self-luminous or externally illuminated and imagine its surface as consisting of a large number of point sources. Each of these emits spherical waves, i.e. rays emanate radially in the direction of energy flow, or if you like, in the direction of the Poynting vector (Fig. 4.1). In this case, the rays *diverge* from a given point source S, whereas if the spherical wave were collapsing to a point, the rays would of course be *converging*. Generally one deals only with a small portion of a wavefront. *A point from which a portion of a spherical wave diverges, or one towards which the wave segment converges, is known as a focal point of the bundle of rays.*

Now envision the situation where we have a point source in the vicinity of some arrangement of reflecting and refracting surfaces representing an *optical system*. Of the infinity of rays emanating from S, generally speaking, only one will pass through an arbitrary point in space. Even so, it is possible to arrange for an infinite number of rays to arrive at a certain point P, as in Fig. 5.1. Thus, if for a cone of rays coming from S there is a corresponding cone of rays passing through P, the system is said to be *stigmatic* for these two points. The energy in the cone (apart from some inadvertent losses due to reflection, scattering and absorption) reaches P which is then referred to as a *perfect image* of S. The wave could conceivably arrive to form a finite patch of light or *blur spot* about P; it would still be an image of S but no longer a perfect one.

It follows from the principle of reversibility (Section 4.2.4) that a point source placed at P would equally well be imaged at S and accordingly these two are spoken of as *conjugate points*. In an *ideal optical system* every point of a three-dimensional region will be perfectly (or stigmatically) imaged in another region; the former being the *object space*, the latter the *image space*.

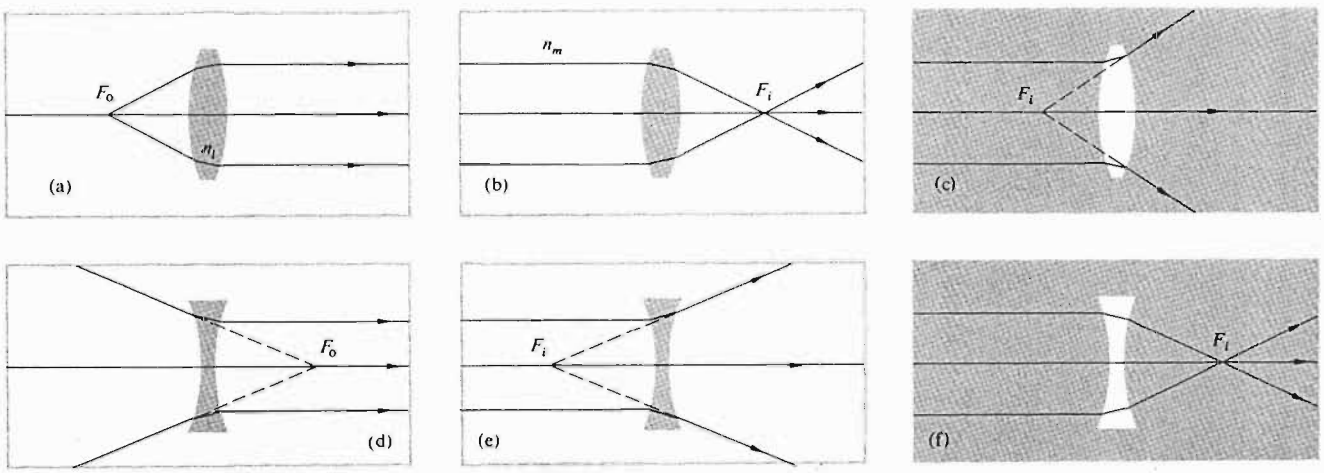


Fig. 5.18 Focal lengths for converging and diverging lenses.

Notice that in each instance it is particularly convenient to draw a ray through the center of the lens which, because it is perpendicular to both surfaces, is undeviated. Suppose however that an off-axis paraxial ray emerges from the lens parallel to its incident direction as in Fig. 5.19. We maintain that all such rays will pass through the point defined as the *optical center* of the lens O. To see this, draw two parallel planes, one on each side tangent to the lens at any pair of points A and B. This can easily be done by selecting A and B such that the radii $\overline{AC}_1$ and $\overline{BC}_2$ are themselves parallel. It is to be shown that the paraxial ray traversing $\overline{AB}$ enters and leaves the lens in the same direction. It is evident from the diagram that triangles AOC_1 and BOC_2 are similar, in the geometrical sense, and therefore their sides are proportional. Hence, $|R_1|(\overline{OC}_2) = |R_2|(\overline{OC}_1)$ and since the radii are constant, the location of O is constant, independent of A and B. As we saw earlier (Problem 4.15 and Fig. 4.44), a ray traversing a medium bounded by parallel planes will be displaced laterally, but will suffer no angular deviation. This displacement is proportional to the thickness, which for a thin lens is negligible. *Rays passing through O may, accordingly, be drawn as straight lines.* It is customary when dealing with thin lenses simply to place O midway between the vertices.

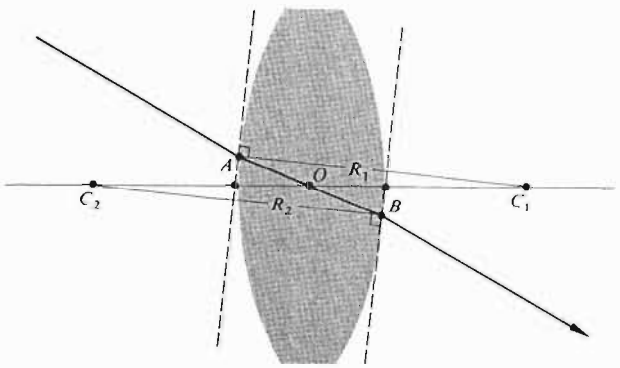
Recall that a bundle of parallel paraxial rays incident on a spherical refracting surface comes to a focus at a point on the optical axis (Fig. 5.11). As shown in Fig. 5.20, this implies that several such bundles entering in a narrow cone will be focused on a spherical segment σ, also centered on C. The undeviated rays normal to the surface, and therefore passing through C, locate the foci on σ. Since the ray cone must indeed be narrow, σ can satisfactorily be represented as a plane normal to the symmetry axis and passing through the image focus. It is known as a *focal plane*. In the same way, limiting ourselves to paraxial theory, a lens will focus all incident parallel bundles of rays* onto a surface called the *second or back focal plane* as in Fig. 5.21. Here each point on σ is located by the undeviated ray through O. Similarly, the *first or front focal plane* contains the object focus F_o.

iii) Finite Imagery

Thus far we've dealt with the mathematical abstraction of a single-point source, but now let's suppose there to be a great many such points combining to form a continuous finite object. For the moment, imagine the object to be a segment of a sphere, σ_o, centered on C as in Fig. 5.22. If σ_o is close to the spherical interface, point S will have a virtual image $P(s_i < 0$ and therefore on the left of $V)$. With S farther away, its image will be real ($s_i > 0$ and therefore on the right-hand side). In either case, each point on σ_o has a conjugate point on σ_i lying on a straight line through C. Within the restrictions

* Perhaps the earliest literary reference to the focal properties of a lens appears in Aristophanes' play, *The Clouds*, which dates back to 423 B.C. In it Strepsiades plots to use a burning glass to focus the sun's rays onto a wax tablet and thereby melt out the record of a gambling debt.

of paraxial theory, these surfaces can be considered as planar. Thus a small planar object normal to the optical axis will be imaged into a small planar region also normal to that axis. It should be noted that if σ_o is moved out to infinity, the cone of rays from each source point will become *collimated*, i.e. parallel, and the image points will lie on the focal plane (Fig. 5.21).



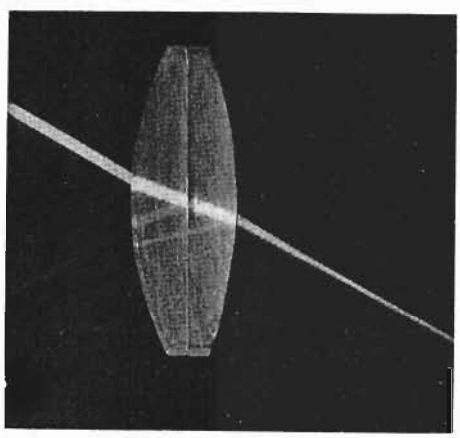


Fig. 5.19 The optical center of a lens. [Photo by E.H.].

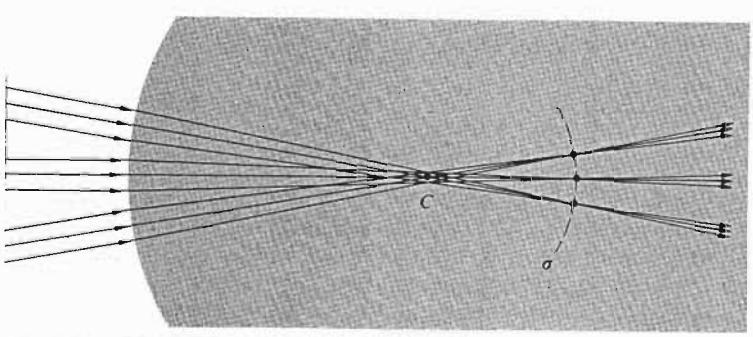


Fig. 5.20 Focusing of several ray bundles.

By cutting and polishing the right side of the piece depicted in Fig. 5.22, we can construct a thin lens just as was done in Section (i). Once again, the image (σ_i in Fig. 5.22) formed by the first surface of the lens will serve as the object for the second surface, which in turn will generate a final image. Suppose then that σ_i in Fig. 5.22(a) is the object for the second surface which is assumed to have a negative radius. We already know what will happen next—the situation is identical to Fig. 5.22(b) with the ray directions reversed. The *final image formed by a lens of a small planar object normal to the optical axis will itself be a small plane normal to that axis.*

The location, size and orientation of an image produced by a lens can be determined, particularly simply, using ray diagrams. To find the image of the object in Fig. 5.23, we must locate the image point corresponding to each object point. Since all rays issuing from a source point in a paraxial cone will arrive at the image point, any two such rays will suffice to fix that point. Knowing the positions of the focal points, there are three rays which are especially easy to apply. Two of these make use of the fact that a ray passing through the focal point will emerge from the lens parallel to the optical axis and vice versa; the other is the undeviated ray through *O*. Incidentally, this technique dates back to the work of Robert Smith as long ago as 1738.

This graphical procedure can be made even simpler by replacing the thin lens with a plane passing through its

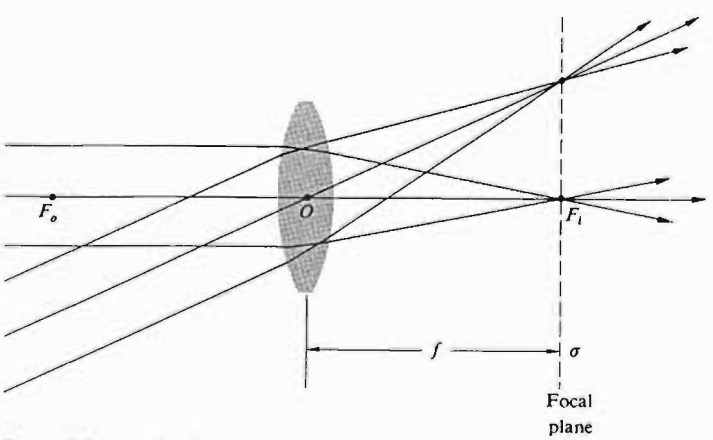


Fig. 5.21 The focal plane of a lens.

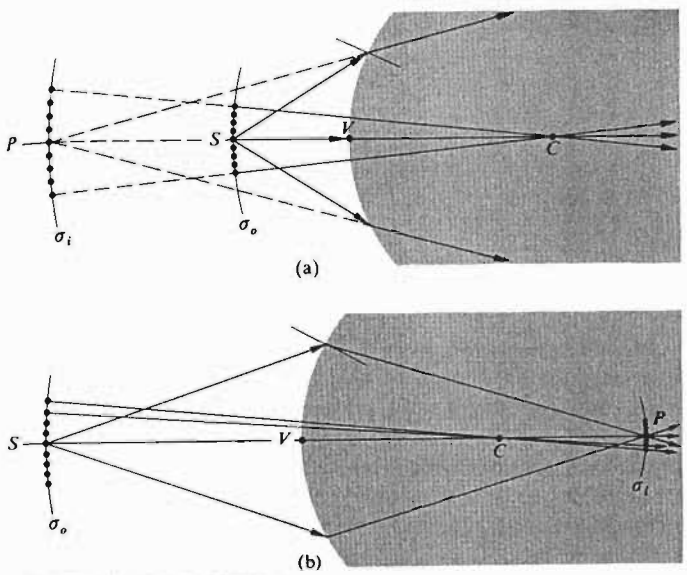


Fig. 5.22 Finite imagery.

center (Fig. 5.24). Presumably if we were to extend every incoming ray forward a little and every outgoing ray backward a bit, each pair would meet on this plane. Thus the total deviation of any ray can be envisaged as occurring all at once on that plane. This is equivalent to the actual process consisting of two separate angular shifts, one at each interface. (As we will see later, this is tantamount to saying that the two principal planes of a thin lens coincide.)

In accord with convention, transverse distances above the optical axis are taken as positive quantities while those below the axis are given negative numerical values. Therefore in Fig. 5.24 $y_o > 0$ and $y_i < 0$. Here the image is said to be *inverted* whereas if $y_i > 0$ when $y_o > 0$, it is *erect*. Observe that triangles AOF_i and $P_2P_1F_i$ are similar. Ergo

$$\frac{y_o}{|y_i|} = \frac{f}{(s_i - f)}, \tag{5.19}$$

Likewise, triangles S_2S_1O and P_2P_1O are similar and

$$\frac{y_o}{|y_i|} = \frac{s_o}{s_i}, \tag{5.20}$$

where all quantities other than y_i are positive. Hence

$$\frac{s_o}{s_i} = \frac{f}{(s_i - f)} \tag{5.21}$$

and

$$\frac{1}{f} = \frac{1}{s_o} + \frac{1}{s_i},$$

which is, of course, the Gaussian lens equation (5.17). Furthermore, triangles $S_2S_1F_o$ and BOF_o are similar and accordingly

$$\frac{f}{(s_o - f)} = \frac{|y_i|}{y_o}. \tag{5.22}$$

Using the distances measured from the focal points and combining this with Eq. (5.19) we have

$$x_o x_i = f^2. \tag{5.23}$$

This is the *Newtonian form* of the lens equation, the first statement of which appeared in Newton's *Opticks* in 1704. The signs of x_o and x_i are reckoned with respect to their concomitant foci. By convention x_o is taken to be positive

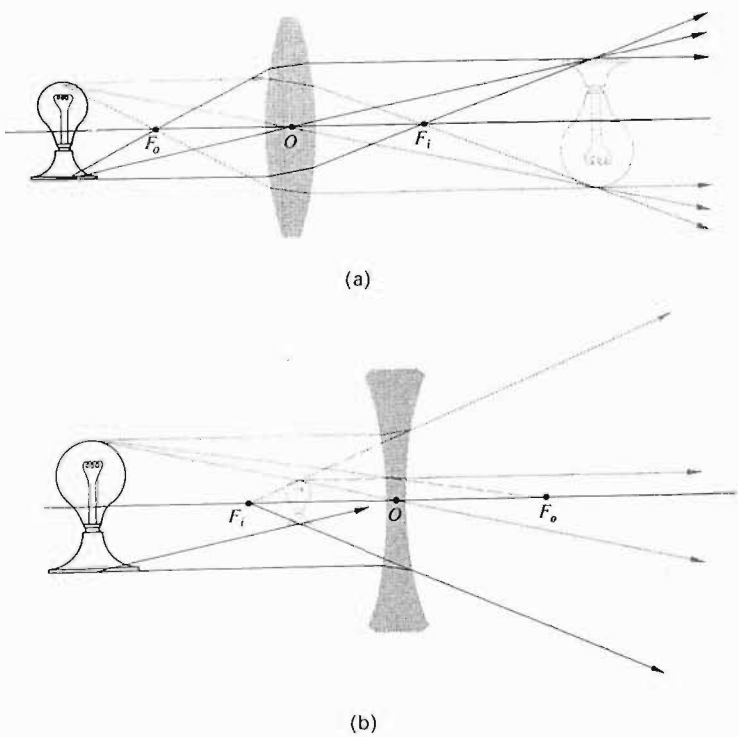


Fig. 5.23 (a) A real object and a positive lens. (b) A real object and a negative lens.

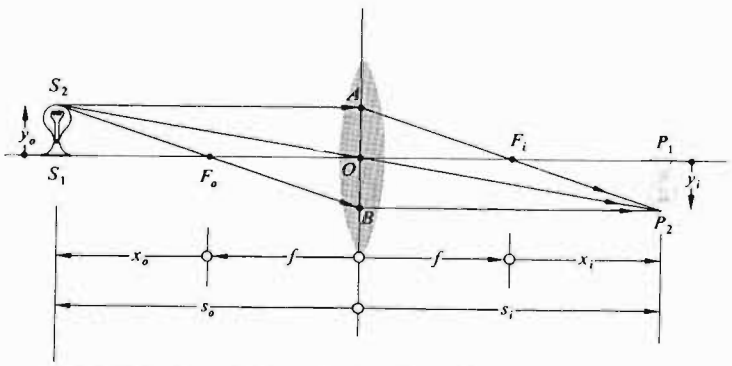


Fig. 5.24 Object and image location for a thin lens.

left of F_o whereas x_i is positive on the right of F_i. To be sure, it is evident from Eq. (5.23) that x_o and x_i have like signs and that means that *the object and image must be on opposite sides of their respective focal points.* This is a good thing for the neophyte to remember when making those hasty freehand ray diagrams for which he is already infamous.

The ratio of the transverse dimensions of the final image formed by any optical system to the corresponding dimension of the object is defined as the *lateral* or *transverse magnification* M_T, that is

$$M_T \equiv \frac{y_i}{y_o}. \tag{5.24}$$

Or from Eq. (5.20)

$$M_T = -\frac{s_i}{s_o}. \tag{5.25}$$

Table 5.2 Meanings associated with the signs of various thin lens and spherical interface parameters.

Quantity	Sign	
	+	−
s_o	Real object	Virtual object
s_i	Real image	Virtual image
f	Converging lens	Diverging lens
y_o	Erect object	Inverted object
y_i	Erect image	Inverted image
M_T	Erect image	Inverted image

Thus *a positive M_T connotes an erect image, while a negative value means the image is inverted* (see Table 5.2). Bear in mind that s_i and s_o are both positive for real objects and images. *Clearly then all such images formed by a single thin lens will be inverted.* The Newtonian expression for the magnification follows from Eqs. (5.19) and (5.22) and Fig. 5.24, whence

$$M_T = -\frac{x_i}{f} = -\frac{f}{x_o}. \tag{5.26}$$

The term magnification is a misnomer since the magnitude of M_T can certainly be less than one in which case the image is smaller than the object. We have $M_T = -1$ when the object and image distances are positive and equal, and that happens (5.17) only when $s_o = s_i = 2f$. This turns out to be (Problem 5.5) the configuration where the object and image are as close together as they can possibly get (namely a distance $4f$ apart). Table 5.3 summarizes a number of image

Table 5.3 Images of real objects formed by thin lenses.

Convex				
Object	Image			
Location	Type	Location	Orientation	Relative size
$\infty > s_o > 2f$	Real	$f < s_i < 2f$	Inverted	Minified
$s_o = 2f$	Real	$s_i = 2f$	Inverted	Same size
$f < s_o < 2f$	Real	$\infty > s_i > 2f$	Inverted	Magnified
$s_o = f$		$\pm\infty$		
$s_o < f$	Virtual	$\lvert s_i\rvert > s_o$	Erect	Magnified

Concave				
Object	Image			
Location	Type	Location	Orientation	Relative size
Anywhere	Virtual	$\lvert s_i\rvert < \lvert f\rvert$	Erect	Minified

configurations resulting from the juxtaposition of a thin lens and a real object.

Presumably the image of a three-dimensional object will itself occupy a three-dimensional region of space. The optical

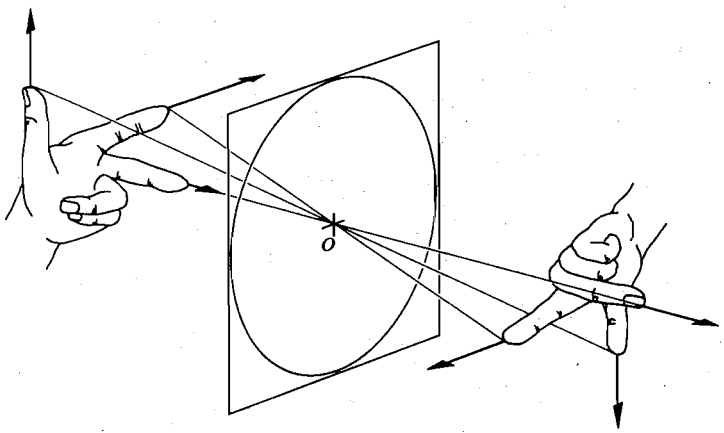


Fig. 5.25 Image orientation for a thin lens.

system can apparently affect both the transverse and longitudinal dimensions of the image. The *longitudinal magnification*, M_L, which relates to the axial direction, is defined as

$$M_L \equiv \frac{dx_i}{dx_o}. \tag{5.27}$$

This is the ratio of an infinitesimal axial length in the region of the image to the corresponding length in the region of the object. Differentiating Eq. (5.23) leads to

$$M_L = -\frac{f^2}{x_o^2} = -M_T^2 \tag{5.28}$$

for a thin lens in a single medium. Evidently, $M_L < 0$ which implies that a positive dx_o corresponds to a negative dx_i and vice versa. In other words, a finger pointing towards the lens is imaged pointing away from it (Fig. 5.25).

Form the image of a window on a sheet of paper using a simple convex lens. Assuming a lovely arboreal scene, image the distant trees on the screen. Now move the paper *away* from the lens so that it intersects a different region of the image space. The trees will fade while the nearby window itself comes into view.

iv) Thin-Lens Combinations

It is not our intent here to have the reader develop a proficiency with the subtle intricacies of modern lens design, but rather to bring him to a point where he can begin to appreciate, utilize and adapt those systems already available.

In constructing a new optical system, one generally begins by sketching out a rough arrangement using the quickest approximate calculations. Refinements are then added as the designer goes on to the prodigious and more exact ray-tracing techniques. Nowadays these computations are most often carried out by electronic digital computers. Even so, the simple thin-lens concept provides a highly useful basis for preliminary calculations in a broad range of situations.

No lens is actually a thin lens in the strict sense of having a thickness which approaches zero. Yet many simple lenses, for all practical purposes, function in a fashion equivalent to that of a thin lens. Spectacle lenses which, by the way, have been used at least since the thirteenth century are almost all in this category. When the radii of curvature are large and the lens diameter is small, the thickness will usually be small as well. A lens of this sort would generally have a large focal length compared to which the thickness would be quite small, e.g. many early telescope objectives fit that description perfectly.

We now propose to derive some expressions for parameters associated with thin-lens combinations. The approach here will be fairly simple, leaving the more elaborate traditional treatment for those tenacious enough to pursue the matter into the next chapter.

Suppose we have two thin positive lenses L_1 and L_2 separated by a distance d which is smaller than either focal length as in Fig. 5.26. The resulting image can be located graphically as follows. Overlooking the presence of L_2, the image formed exclusively by L_1 is constructed using rays 1 and 3. As usual, these pass through the lens' object and image foci, namely F_{o1} and F_{i1} respectively. The object is in a normal plane so that two rays determine its top and a perpendicular to the optical axis finds its bottom. Ray 2 is then constructed running backwards from P_1' through O_2. Inserting L_2 has no effect on ray 2 whereas 3 is refracted through the image focus F_{i2} of L_2. The intersection of rays 2 and 3 fixes the image, which in this particular case is real, minified and inverted.

A similar pair of lenses is illustrated in Fig. 5.27 where now the separation has been increased. Once again rays 1 and 3 through F_{i1} and F_{o1} fix the position of the intermediate image generated by L_1 alone. As before, ray 2 is drawn backward from O_2 to P_1' to S_1. The intersection of 2 and 3, as the latter is refracted through F_{i2}, locates the final image. This time it is real and erect. Notice that if the focal

More on Geometrical Optics

6

The preceding chapter, for the most part, dealt with paraxial theory as applied to thin spherical lens systems. The two predominant approximations were, rather obviously, that we had *thin* lenses and that first-order theory was sufficient for their analysis. Neither of these assumptions can be maintained throughout the design of a precision optical system but both, taken together, provide the basis for a first rough solution. This chapter will carry things a bit further by examining thick lenses and aberrations; even at that, it is only a beginning. The advent of computerized lens design requires a certain shift in emphasis—there is little need to do what a computer can do better. Moreover, the sheer wealth of existing material developed over centuries demands a bit of judicious pruning to avoid a plethora of pedantry.

6.1 THICK LENSES AND LENS SYSTEMS

Figure 6.1 depicts a thick lens, i.e., one whose thickness is by no means negligible. As we shall see, it could equally well be envisioned more generally as an optical system allowing of the possibility that it consists of a number of simple lenses, not merely one. The first and second focal points, or if you like, the object and image foci, F_o and F_i, can conveniently be measured from the two (outermost) vertices. In that case we have the familiar front and back focal lengths denoted by f.f.l. and b.f.l. When extended, the incident and emerged rays will meet at points, the locus of which forms a curved surface that may or may not reside within the lens. The surface, approximating a plane in the paraxial region, is termed the *principal plane* [see Section 6.3.1(ii)]. Points where the primary and secondary principal planes (as shown in Fig. 6.1) intersect the optical axis are known as the *first* and *second principal points*, H_1 and H_2 respectively. They provide a set of very useful references from which to measure several of the system parameters. We saw earlier (Fig. 5.19, p. 110) that a ray traversing the lens through its optical

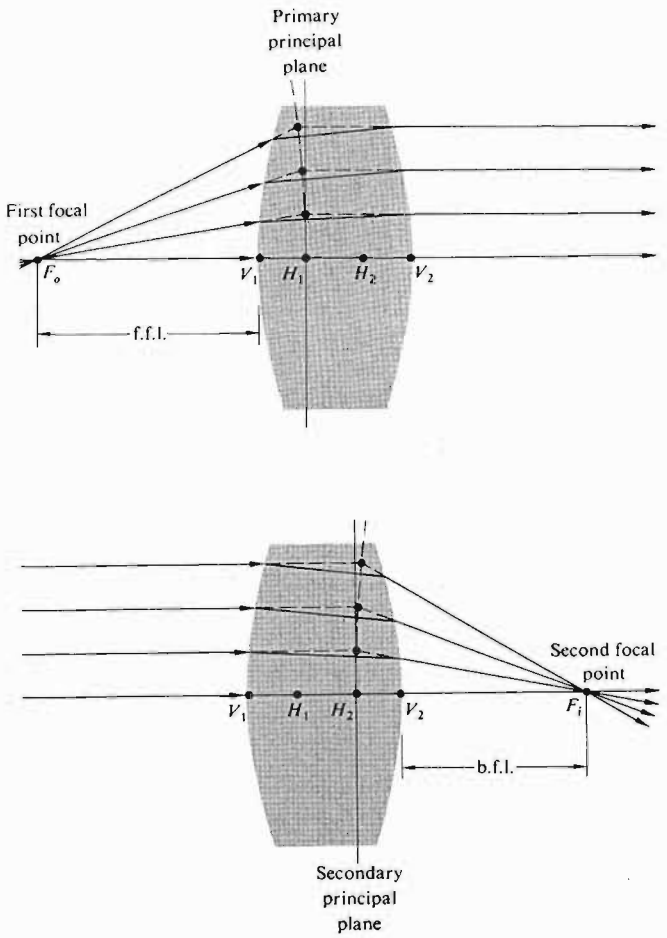


Fig. 6.1 A thick lens.

center emerges parallel to the incident direction. Extending both the incoming and outgoing rays until they cross the optical axis locates what are called the *nodal points*, N_1 and N_2 in Fig. 6.2. When the lens is surrounded on both sides by the same medium, generally air, the nodal and principal points will be coincident. The six points, two focal, two principal and two nodal, constitute the *cardinal points* of the system. As shown in Fig. 6.3 the principal planes can indeed lie completely outside of the lens system. Here although differently configured, each lens in either group has the same power. Observe that in the symmetrical lens the principal planes are, quite reasonably, symmetrically located. In the case of either the planar-concave or planar-convex

$$\mathscr{A} = \mathscr{R}_2\mathscr{R}_1 = \begin{bmatrix} 1 & -(\mathscr{D}_1 + \mathscr{D}_2) \\ 0 & 1 \end{bmatrix}. \tag{6.38}$$

But as we saw in Section 5.7.2, the power of a thin lens $\mathscr{D}$ is the sum of the powers of its surfaces. Hence

$$\mathscr{A} = \begin{bmatrix} 1 & -\mathscr{D} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1/f \\ 0 & 1 \end{bmatrix}. \tag{6.39}$$

In addition, for two thin lenses separated by a distance d, in air, the system matrix is

$$\mathscr{A} = \begin{bmatrix} 1 & -1/f_2 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ d & 1 \end{bmatrix}\begin{bmatrix} 1 & -1/f_1 \\ 0 & 1 \end{bmatrix}$$

or

$$\mathscr{A} = \begin{bmatrix} 1 - d/f_2 & -1/f_1 + d/f_1f_2 - 1/f_2 \\ d & -d/f_1 + 1 \end{bmatrix}.$$

Clearly then

$$-a_{12} = \frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} - \frac{d}{f_1f_2}$$

and from Eqs. (6.36) and (6.37)

$$\overline{O_1H_1} = fd/f_2, \qquad \overline{O_2H_2} = -fd/f_1,$$

all of which by now should be quite familiar. Note how easy it would be with this approach to find the focal length and principal points for a compound lens composed of three, four or more thin lenses.

6.3 ABERRATIONS

To be sure, we already know that first-order theory is no more than a good approximation—an exact ray trace or even measurements performed on a prototype system would certainly reveal inconsistencies with the corresponding paraxial description. Such departures from the idealized conditions of Gaussian optics are known as *aberrations*. There are two main classifications of these, namely *chromatic aberrations* (which arise from the fact that n is actually a function of frequency or color) and *monochromatic aberrations*. The latter occur even with light which is highly monochromatic, and they, in turn, fall into two subgroupings. There are monochromatic aberrations which deteriorate the image making it unclear, such as *spherical aberration*, *coma* and *astigmatism*. In addition, there are aberrations which deform the image, as for example, *Petzval field curvature* and *distortion*.

We have known all along that spherical surfaces in general would yield perfect imagery only in the paraxial region. What must now be determined is the kind and extent of the deviations which result simply from using those surfaces with finite apertures. By the judicious manipulation of a system's physical parameters (e.g. the powers, shapes, thicknesses, glass types and separations of the lenses as well as the locations of stops), these aberrations can indeed be minimized. In effect, one cancels out the most undesirable faults by a slight change in the shape of a lens here, or a shift in the position of a stop there (very much like trimming up a circuit with small variable capacitors, coils and pots). When it's all finished, the unwanted deformations of the wavefront incurred as it passes through one surface will hopefully be negated as it traverses some other surfaces further down the line.

Today there are elaborate computer programs for "automatically" doing this kind of analysis. Broadly speaking, you give the computer a quality factor (or merit function) of some sort to aim for, i.e., you essentially tell it how much of each aberration you are willing to tolerate. Then you give it a roughly designed system (e.g. some Tessar configuration) which, in the first approximation, meets the particular requirements. Along with that, you feed in whatever parameters must be held constant, such as a given f-number, focal length, or lens diameter, the field of view, or magnification. The computer will then trace several rays through the system and evaluate the image errors. Having been given leave to vary, say, the curvatures and axial separations of the elements, it will calculate the optimum effect of such changes on the quality factor, make them, and then reevaluate. After perhaps twenty or more iterations, usually taking a matter of minutes, it will have changed the initial configuration so that it now meets the specified limits on aberrations. The final lens design will still be a Tessar, but not the one you started with. The result is, if you will, an *optimum configuration* but probably not *the* optimum. We can be fairly certain that all aberrations cannot be made exactly zero in any real system comprised of spherical surfaces. Moreover, there is no presently known way to determine how close to zero we can actually come. A quality factor is somewhat like a crater-pocked surface in a multidimensional space. The computer will carry the design from one hole to the next until it finds one deep enough to meet the specifications. There it stops and presumably presents us with a perfectly satisfactory configuration. But there is no way to tell if that solution

corresponds to the deepest hole without sending the computer out again and again meandering along totally different routes.

We mention all of this so that the reader may appreciate the current state of the art. In a word, it is magnificent, but still incomplete; it is "automatic" but a bit myopic.

6.3.1 Monochromatic Aberrations

The paraxial treatment was based on the assumption that $\sin\varphi$, as in Fig. 5.8, could be represented satisfactorily by φ alone, i.e. the system was restricted to operating in an extremely narrow region about the optical axis. Obviously, if rays from the periphery of a lens are to be included in the formation of an image, the statement that $\sin\varphi \approx \varphi$ is somewhat unsatisfactory. Recall that we also occasionally wrote Snell's law simply as $n_i\theta_i = n_t\theta_t$ which again would be inappropriate. In any event, if the first two terms in the expansion

$$\sin\varphi = \varphi - \frac{\varphi^3}{3!} + \frac{\varphi^5}{5!} - \frac{\varphi^7}{7!} + \cdots \qquad [5.7]$$

are retained as an improved approximation, we have the so-called *third-order theory*. Departures from first-order theory which then result are embodied in the five *primary aberrations* (spherical aberration, coma, astigmatism, field curvature and distortion). These were first studied in detail by Ludwig von Seidel (1821–96) in the eighteen-fifties. Accordingly, they are frequently spoken of as the *Seidel aberrations*. In addition to the first two contributions, the series obviously contains many other terms, smaller to be sure, but still to be reckoned with. Thus, there are most certainly *higher-order aberrations*. The difference between the results of exact ray-tracing and the computed primary aberrations can therefore be thought of as the sum of all contributing higher-order aberrations. We shall restrict this discussion to the primary aberrations exclusively.

i) Sperical Aberration

Let's return for a moment to Section 5.2.2 (p.102) where we computed the conjugate points for a single refracting spherical interface. We found then, for the paraxial region, that

$$\frac{n_1}{s_o} + \frac{n_2}{s_i} = \frac{n_2 - n_1}{R}. \qquad [5.8]$$

If the approximations for ℓ_o and ℓ_i are improved a bit (Problem 6.11) we get the third-order expression:

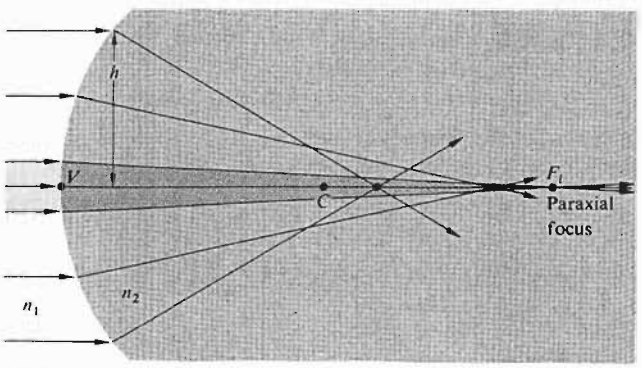


Fig. 6.12 Spherical aberration resulting from refraction at a single interface.

$$\frac{n_1}{s_o} + \frac{n_2}{s_i} = \frac{n_2 - n_1}{R} + h^2\left[\frac{n_1}{2s_o}\left(\frac{1}{s_o} + \frac{1}{R}\right)^2 + \frac{n_2}{2s_i}\left(\frac{1}{R} - \frac{1}{s_i}\right)^2\right]. \qquad (6.40)$$

The additional term, which varies approximately as h^2, is clearly a measure of the deviation from first-order theory. As shown in Fig. 6.12 rays striking the surface at greater distances above the axis (h) are focused nearer to the vertex. In brief, spherical aberration or SA corresponds to a dependence of focal length on aperture for nonparaxial rays. Similarly, for a converging lens, as in Fig. 6.13, the marginal rays will, in effect, be bent too much, being focused in front of the paraxial rays. The distance between the axial intersection of a ray and the paraxial focus, F_i, is known as the *longitudinal spherical aberration*, or L · SA, of that ray. In this case, the SA is *positive*. In contrast the marginal rays for a diverging lens will generally intersect the axis behind the paraxial focus and we say that its spherical aberration is therefore *negative*.

If a screen is placed at F_i in Fig. 6.13, the image of a star would appear as a bright central spot on the axis surrounded by a symmetrical halo delineated by the cone of marginal rays. For an extended image, SA would reduce the contrast and degrade the details. The height above the axis where a given ray strikes this screen is called the *transverse* (or *lateral*) *spherical aberration*, T · SA for short. Evidently, SA can be reduced by stopping down the aperture—but that reduces the amount of light entering the system as well. Notice that if the screen is moved to the position labeled Σ_{LC} the image blur will have its smallest diameter. This is known as the *circle of least confusion* and Σ_{LC} is generally the best place to observe the image. If a lens exhibits appreciable SA, it

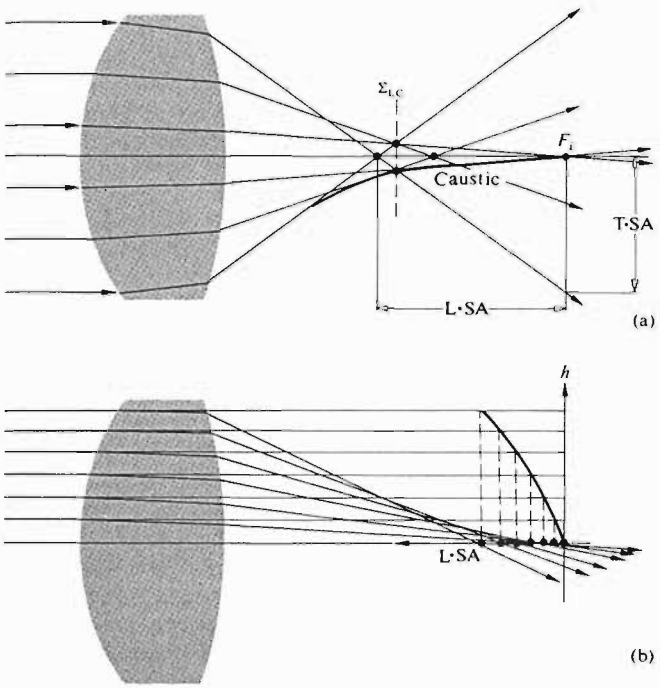


Fig. 6.13 Spherical aberration for a lens. The envelope of the refracted rays is called a caustic. The intersection of the marginal rays and the caustic locates $\sum_{LC}$.

will have to be refocused after it is stopped down because the position of Σ_{LC} will approach F_i as the aperture decreases.

The amount of spherical aberration, when the aperture and focal length are fixed, varies with both the object distance and the lens shape. For a converging lens, the nonparaxial rays are too strongly bent. Yet if we imagine the lens as roughly resembling two prisms joined at their bases, it is evident that *the incident ray will undergo a minimum deviation when it makes, more or less, the same angle as does the emerging ray* (Section 5.5.1). A striking example is illustrated in Fig. 6.14 where simply turning the lens around markedly reduces the SA. When the object is at infinity a simple concave or convex lens which has an almost, but not quite, flat rear side will suffer a minimum amount of spherical aberration. In the same way, if the object and image distances are to be equal ($s_o = s_i = 2f$) the lens should be equiconvex to minimize SA. A combination of a converging and a diverging lens (as in an achromatic doublet) can also be utilized to diminish spherical aberration.

Recall that the aspherical lenses of Section 5.2.1 were completely free of spherical aberration for a specific pair of conjugate points. Moreover, Huygens seems to have been the first to discover that two such axial points exist for spherical surfaces as well. These are shown in Fig. 6.15(a) which depicts rays issuing from P and leaving the surface as if they came from P'. It is left for a problem to show that the appropriate locations of P and P' are those indicated in the figure. Just as with the aspherics, lenses can be formed which have this same zero SA for the pair of points P and P'. One simply grinds another surface of radius $\overline{PA}$ centered on P to form either a positive- or negative-meniscus lens. The oil-immersion microscope objective uses this principle to great advantage. The object under study is positioned at P and surrounded by oil of index n_2 as in Fig. 6.16. P and P' are the proper conjugate points for zero SA for the first element while P' and P'' are those for the meniscus lens.

ii) Coma

Coma or *comatic aberration* is an image-degrading, monochromatic, primary aberration associated with an object point even a short distance from the axis. Its origins lie in the fact that the principal "planes" can actually be treated as planes only in the paraxial region. They are, in fact, principal curved surfaces (Fig. 6.1). In the absence of SA a parallel bundle of rays will focus at the axial point F_i, a distance b.f.l. from the rear vertex. Yet the effective focal lengths and therefore the

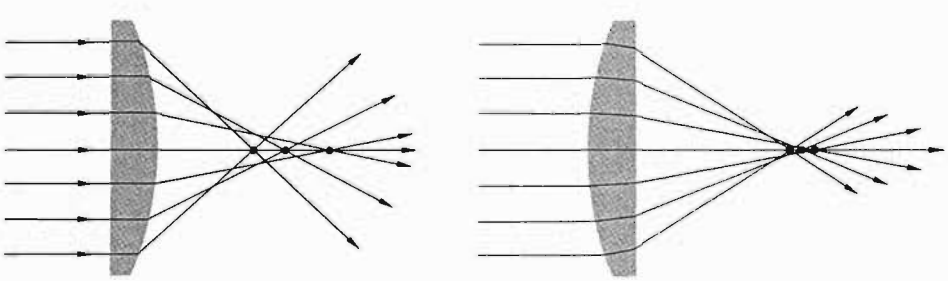


Fig. 6.14 SA for a planar-convex lens.

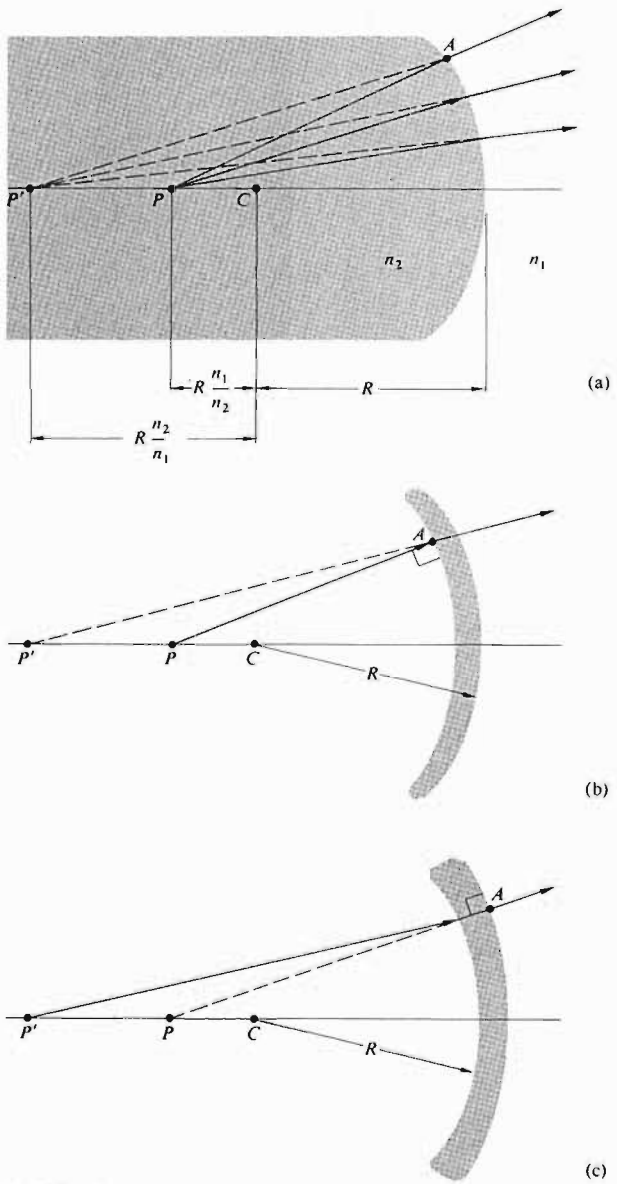


Fig. 6.15 Corresponding axial points for which SA is zero.

transverse magnifications will differ for rays traversing off-axis regions of the lens. When the image point is on the optical axis, this situation is of little consequence, but when the ray bundle is oblique and the image point is off-axis, coma will be

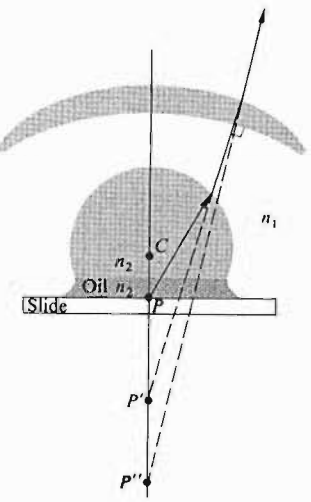


Fig. 6.16 An oil-immersion microscope objective.

evident. The dependence of M_T on h, the ray height at the lens, is evident in Fig. 6.17. Here meridional rays traversing the extremities of the lens arrive at the image plane closer to the axis than do the rays in the vicinity of the *principal ray* (i.e., the ray which passes through the principal points). In this instance, the least magnification is associated with the marginal rays which would form the smallest image—the coma is said to be negative. By comparison, the coma in Fig. 6.18 is positive because the marginal rays focus further from the axis. Several skew rays are drawn from an extra-axial object point S in Fig. 6.19 to illustrate the formation of the geometrical comatic image of a point. Observe that each circular cone of rays whose endpoints (1-2-3-4-1-2-3-4) form a ring on the lens is imaged in what H. Dennis Taylor called a *comatic circle* on Σ_i. This case corresponds to positive coma and so the larger the ring on the lens, the more distant will be its comatic circle from the axis. When the outer ring is the intersection of marginal rays, the distance from 0 to 1 in the image is the *tangential coma*, while the length from 0 to 3 on Σ_i is termed the *sagittal coma*. A little more than half of the energy in the image appears in the roughly triangular region between 0 and 3. The coma flare, which owes its name to its comet-like tail, is often thought to be the worst of all aberrations, primarily because of its asymmetric configuration.

Like SA, coma is dependent on the shape of the lens. Thus, a strongly concave positive-meniscus lens) with

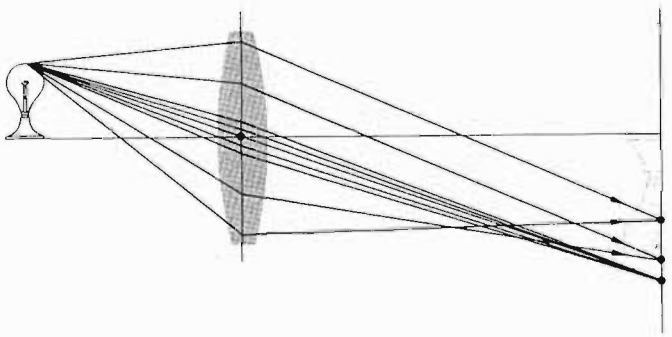
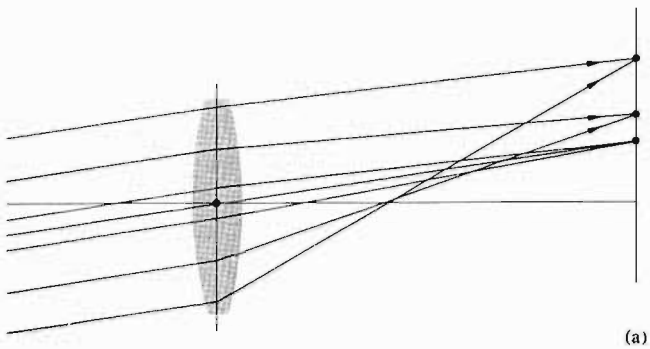


Fig. 6.17 Negative coma.

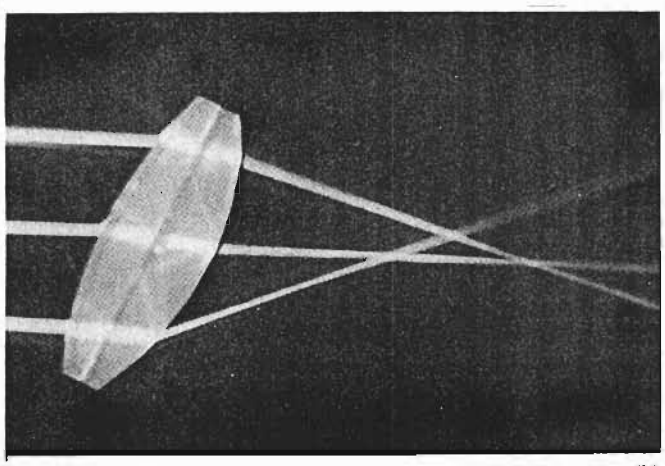
the object at infinity will have a large negative coma. Bending the lens so that it becomes planar-convex), then equi-convex), convex-planar (and finally convex-meniscus (, will change the coma from negative, to zero, to positive. The fact that it can be made exactly zero for a single lens with a given object distance is quite significant. The particular shape it would then have ($s_o = \infty$) is almost convex-planar and very nearly the configuration for minimum SA.

It is quite important to realize that *a lens which is well corrected for the case where one conjugate point is at infinity* ($s_o = \infty$) *may not perform satisfactorily when the object is nearby*. One would therefore do well, when using off-the-shelf lenses in a system operating at finite conjugates, to combine two infinite conjugate corrected lenses as in Fig. 6.20. In other words, since it is unlikely that a lens with the desired focal length, which is also corrected for the particular set of finite conjugates, can be obtained ready made, this back-to-back lens approach is an appealing alternative.

Coma can also be negated by using a stop at the proper location, as was discovered in 1812 by William Hyde Wollaston (1766–1828). The order of the list of primary aberrations (SA, coma, astigmatism, Petzval field curvature and distortion) is significant because any one of them, except SA and Petzval curvature, will be affected by the position of a stop, but only if one of the preceding aberrations is also present in the system. Thus while SA is independent of the location along the axis of a stop, coma will not be, as long as SA is present. This can better be appreciated by examining the representation in Fig. 6.21. With the stop at Σ_1, ray 3 is the chief ray, there is SA but no coma; i.e., the ray pairs meet on 3. If the stop is moved to Σ_2, the symmetry is upset, ray 4 becomes the chief ray and the rays on either side of it, such as 3 and 5, meet above it, not on it—there is then positive coma. With the stop at Σ_3, the rays 1 and 3 intersect below the chief ray, 2, and there is negative coma. In this way, controlled amounts of the aberration can be introduced into a compound lens in order to cancel coma in the system as a whole.



(a)



(b)

Fig. 6.18 Positive coma. [Photo by E.H.]

The *optical sine theorem* is an important relationship which must be introduced here even if space precludes its formal proof. It was discovered independently in 1873 by Abbe and Helmholtz although a different form of it was given ten years earlier by R. Clausius (of thermodynamics fame). In any event, it states that

$$n_o y_o \sin \alpha_o = n_i y_i \sin \alpha_i, \tag{6.41}$$

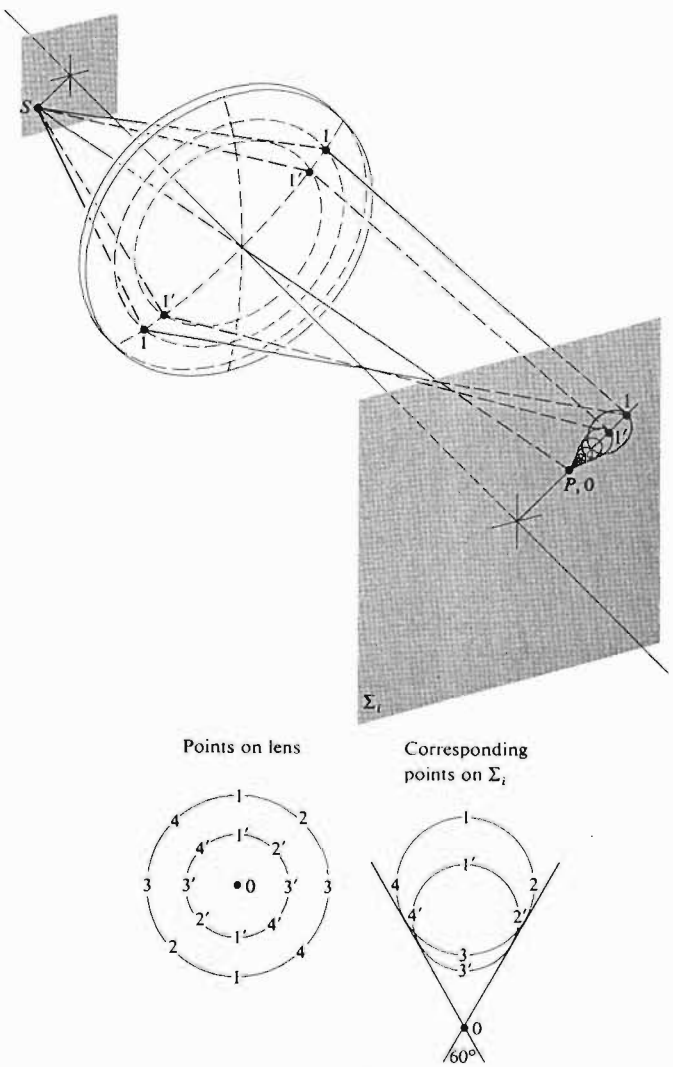


Fig. 6.19 The geometrical coma image of a point. The central region of the lens forms a point image at the vertex of the cone.

where n_o, y_o, α_o and n_i, y_i, α_i are the index, height and slope angle of a ray in object and image space, respectively at any aperture size* (Fig. 6.9). If coma is to be zero

* To be precise, the sine theorem is valid for all values of α_o only in the sagittal plane (from the Latin *sagitta* meaning arrow) which is discussed in the next section.

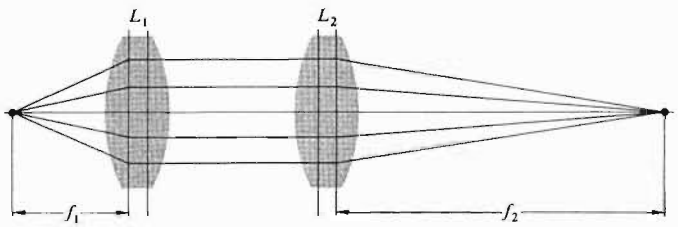


Fig. 6.20 A combination of two infinite conjugate lenses yielding a system operating at finite conjugates.

$$M_T = \frac{y_i}{y_o} \qquad [5.24]$$

must be constant for all rays. Suppose then that we send a marginal and a paraxial ray through the system. The former will comply with Eq. (6.41), the latter with its paraxial version (in which $\sin \alpha_o = \alpha_{op}$, $\sin \alpha_i = \alpha_{ip}$). Since M_T is to be constant over the entire lens, we equate the magnification for both marginal and paraxial rays to get

$$\frac{\sin \alpha_o}{\sin \alpha_i} = \frac{\alpha_{op}}{\alpha_{ip}} = \text{constant}, \qquad (6.42)$$

which is known as the *sine condition*. A necessary criterion for the absence of coma is that the system meet the sine condition. If there is no SA, compliancy with the sine condition will be both necessary and sufficient for zero coma.

It's an easy matter to observe coma. In fact, anyone who has focused sunlight with a simple positive lens has no doubt seen the effects of this aberration. A slight tilt of the lens, so that the nearly collimated rays from the sun make an angle with the optical axis, will cause the focused spot to flare out into the characteristic comet shape.

iii) *Astigmatism*

When an object point lies an appreciable distance from the optical axis the incident cone of rays will strike the lens asymmetrically, giving rise to a third primary aberration known as *astigmatism*. To facilitate its description, envision the meridional plane (also called the *tangential plane*) containing both the chief ray (i.e. the one passing through the center of the aperture) and the optical axis. The *sagittal plane* is then defined as the plane containing the chief ray which, in addition, is perpendicular to the meridional plane (Fig. 6.22). Unlike the latter which is unbroken from one end of a complicated lens system to the other, the sagittal plane generally changes slope as the chief ray is deviated at

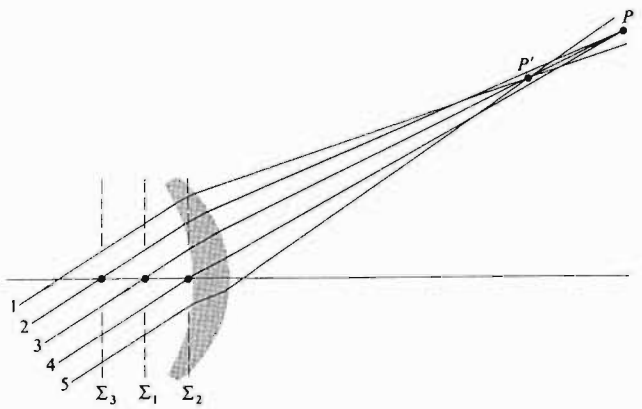


Fig. 6.21 The effect of stop location on coma.

the various elements. Hence to be accurate we should say that there are actually several sagittal planes, one attendant with each region within the system. Nevertheless, all skew rays from the object point lying in a sagittal plane are termed *sagittal rays*.

In the case of an axial object point, the cone of rays is symmetrical with respect to the spherical surfaces of a lens. There is no need to make a distinction between meridional and sagittal planes. The ray configurations in all planes containing the optical axis are identical. In the absence of spherical aberration, all the focal lengths are the same and consequently all rays arrive at a single focus. Contrastingly, the configuration of an oblique, parallel ray bundle will be different in the meridional and sagittal planes. As a result, the focal lengths in these planes will be different as well. In effect, here the meridional rays are tilted more with respect to the lens than are the sagittal rays and they have a shorter focal length. It can be shown,* using Fermat's principle, that the *focal length difference* depends effectively on the power of the lens (as opposed to the shape or index) and the angle at which the rays are inclined. This *astigmatic difference*, as it is often called, increases rapidly as the rays become more oblique, i.e. as the object point moves further off the axis and is, of course, zero on axis.

Having two distinct focal lengths, the incident conical bundle of rays takes on a considerably altered form after refraction (Fig. 6.23). The cross-section of the beam as it leaves the lens is initially circular but it gradually becomes elliptical with the major axis in the sagittal plane until at the

* See A. W. Barton, *A Text Book on Light*, p. 124.

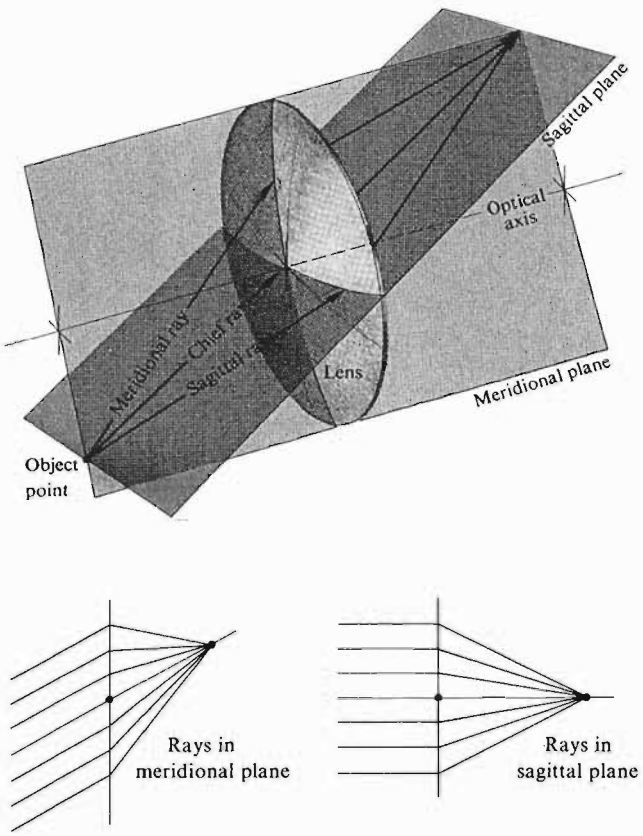


Fig. 6.22 The sagittal and meridional planes.

tangential or meridional focus F_T, the ellipse degenerates into a *line* (at least in third-order theory). All rays from the object point traverse this line which is known as the *primary image*. Beyond this point the beam's cross-section rapidly opens out until it is again circular. At that location the image is a circular blur known as the *circle of least confusion*. Moving further from the lens the beam's cross-section again deforms into a line called the *secondary image*. This time it's in the meridional plane at the *sagittal focus* F_S. Remember that all of this is assuming the absence of SA and coma.

Since the circle of least confusion increases in diameter as the astigmatic difference increases, i.e. as the object moves further off-axis, the image will deteriorate, losing definition around its edges. Observe that the secondary line image will change in orientation with changes in the object position but it will always point toward the optical axis, i.e. it will be

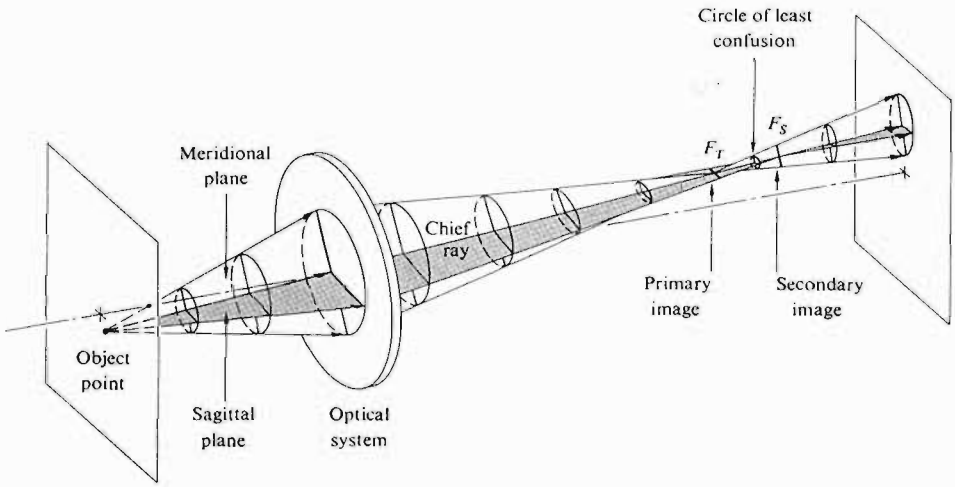


Fig. 6.23 Astigmatism.

radial. Similarly, the primary line image will vary in orientation but it will remain normal to the secondary image. This arrangement causes the interesting effect shown in Fig. 6.24 when the object is made up of radial and tangential elements. The primary and secondary images are, in effect, formed of transverse and radial dashes which increase in size with distance from the axis. In the latter case, the dashes point like arrows toward the center of the image—ergo, the name sagitta.

The existence of the sagittal and tangential foci can be verified directly with a fairly simple arrangement. Place a short focal length positive lens (say about 10 or 20 mm) in the beam of an He–Ne laser. Position another, somewhat longer focal length, positive test lens far enough away so that the now diverging beam fills that lens. A convenient object, to be located between the two lenses, is a piece of ordinary wire screening (or a transparency). Align it so the wires are horizontal (x) and vertical (y). If the test lens is rotated through roughly 45° about the vertical (with the x-, y-, z-axes fixed in the lens), astigmatism should be observable. The meridional is the xz-plane (z being the lens axis, now at about 45° to the laser axis) while the sagittal plane corresponds to the plane of y and the laser axis. As the wire mesh is moved toward the test lens, a point will be reached where the horizontal wires are in focus on a screen beyond the lens, while the vertical wires are not. This is the location of the sagittal focus. Each point on the object is imaged as a short line in the meridional (horizontal) plane, which accounts for the fact that only the horizontal wires are in focus. Moving the mesh slightly closer to the lens will bring the vertical lines into clarity while the horizontal ones are blurred. This is the tangential focus. Try rotating the mesh about the central laser axis while at either focus.

Note that unlike visual astigmatism which arose from an actual asymmetry in the surfaces of the optical system, the third-order aberration by that same name applies to spherically symmetric lenses.

Mirrors, with the singular exception of the plane mirror, suffer much the same monochromatic aberrations as do lenses. Thus while a paraboloidal mirror is free of SA for an infinitely distant axial object point, its off-axis imagery is quite poor due to astigmatism and coma. This strongly restricts its use to narrow field devices such as searchlights and astronomical telescopes. A concave spherical mirror shows SA, coma and astigmatism. Indeed one could draw a diagram just like Fig. 6.23 with the lens replaced by an obliquely illuminated spherical mirror. Incidentally, such a mirror displays appreciably less SA than would a simple convex lens of the same focal length.

iv) Field Curvature

Suppose that we have an optical system which is free of all of the aberrations thus far considered. There would then be a one-to-one correspondence between points on the object and image surfaces (i.e. stigmatic imagery). We mentioned earlier [Section 5.2.3(iii)] that a planar object normal to the axis will be imaged approximately as a plane only in the

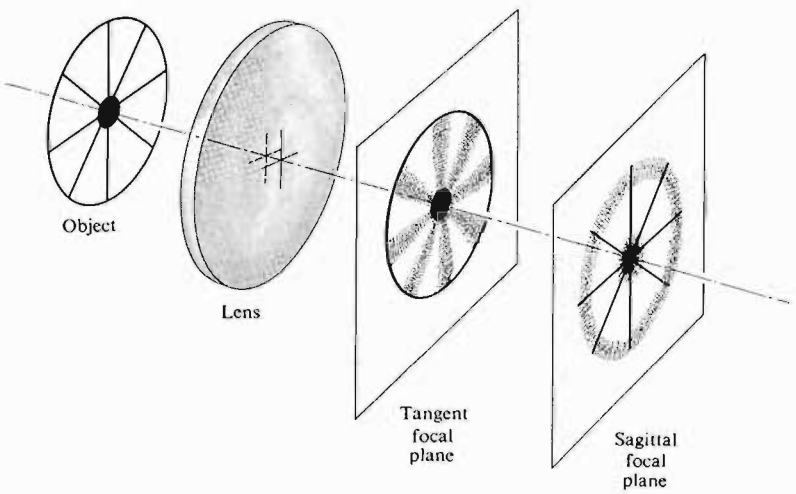


Fig. 6.24 Images in the tangent and sagittal focal planes.

paraxial region. At finite apertures the resulting curved stigmatic image surface is a manifestation of the primary aberration known as *Petzval field curvature* after the Hungarian mathematician Josef Max Petzval (1807–91). The effect can readily be appreciated by examining Figs. 5.22 (p. 111) and 6.25. A spherical object segment σ_o is imaged by the lens as a spherical segment σ_i, both centered at O. Flattening out σ_o into the plane σ_o' will cause each object point to move toward the lens along the concomitant chief ray, thus forming a paraboloidal *Petzval surface* Σ_P. While the Petzval surface for a positive lens curves *inward* toward the object plane, for a negative lens it curves *outward*, i.e. away from that plane. Evidently, a suitable combination of positive and negative lenses will negate field curvature. Indeed, the displacement Δx of an image point at height y_i on the Petzval surface from the paraxial image plane is given by

$$\Delta x = \frac{y_i^2}{2} \sum_{j=1}^{m} \frac{1}{n_j f_j}, \tag{6.43}$$

where n_j and f_j are the indices and focal lengths of the m thin lenses forming the system. This implies that the Petzval surface will be unaltered by changes in the positions or shapes of the lenses, or in the location of the stop, so long as the values of n_j and f_j are fixed. Notice that for the simple case of two thin lenses ($m = 2$) having any spacing, Δx *can be made zero* provided that

$$\frac{1}{n_1 f_1} + \frac{1}{n_2 f_2} = 0$$

or equivalently

$$n_1 f_1 + n_2 f_2 = 0. \tag{6.44}$$

This is the so-called *Petzval condition*. As an example of its use, suppose we combine two thin lenses, one positive, the other negative, such that $f_1 = -f_2$ and $n_1 = n_2$. Since

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} - \frac{d}{f_1 f_2}, \tag{6.8}$$

$$f = \frac{f_1^2}{d},$$

the system can satisfy the Petzval condition, have a flat field, and still have a finite positive focal length.

In visual instruments a certain amount of curvature can be tolerated because the eye can accommodate for it. Clearly in photographic lenses field curvature is most undesirable since it has the effect of rapidly blurring the off-axis image when the film plane is at F_i. An effective means of nullifying the inward curvature of a positive lens is to place a negative *field flattener* lens near the focal plane. This is often done in projection and photographic objectives when it is not otherwise practicable to meet the Petzval condition (Fig. 6.26). In this position the flattener will have little effect on other aberrations (take another look at Fig. 6.7).

Astigmatism is intimately related to field curvature. In the presence of the former aberration, there will be *two* paraboloidal image surfaces, the tangential, Σ_T, and the sagittal, Σ_S (as in Fig. 6.27). These are the loci of all the

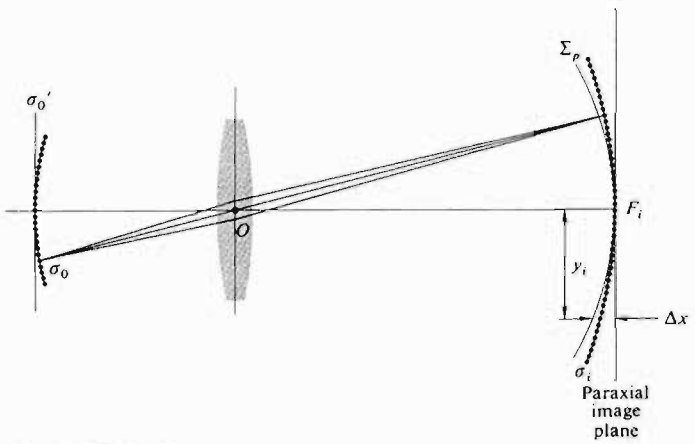


Fig. 6.25 Field curvature.

primary and secondary images, respectively, as the object point roams over the object plane. At a given height (y_i), a point on Σ_T always lies three times as far from Σ_P as does the corresponding point on Σ_S and both are on the same side of the Petzval surface (Fig. 6.27). When there is no astigmatism Σ_S and Σ_T coalesce on Σ_P. It is possible to alter the shapes of Σ_S and Σ_T by bending or relocating the lenses or by moving the stop. The configuration of Fig. 6.27(b) is known as an *artificially flattened* field. A stop in front of an inexpensive meniscus box camera lens is usually arranged to produce just this effect. The surface of least confusion, Σ_{LC}, is planar and the image there is tolerable, losing definition at the margins due to the astigmatism. That is to say, although their loci form Σ_{LC}, the circles of least confusion increase in diameter with distance off the axis. Modern good-quality photographic objectives are generally *anastigmats*, i.e. they are designed so that Σ_S and Σ_T cross each other yielding an additional off-axis angle of zero astigmatism. The Cooke Triplet, Tessar, Orthometer and Biotar (Fig. 5.100) are all anastigmats as is the relatively fast Zeiss Sonnar whose residual astigmatism is illustrated graphically in Fig. 6.28. Note the relatively flat field and small amount of astigmatism over most of the film plane.

Let's return briefly to the Schmidt camera of Fig. 5.95 (p. 157) since we are now in a better position to appreciate how it functions. With a stop at the center of curvature of the spherical mirror, all chief rays, which by definition pass through C, are incident normally on the mirror. Moreover each pencil of rays from a distant object point is symmetrical about its chief ray. In effect, each chief ray serves as an

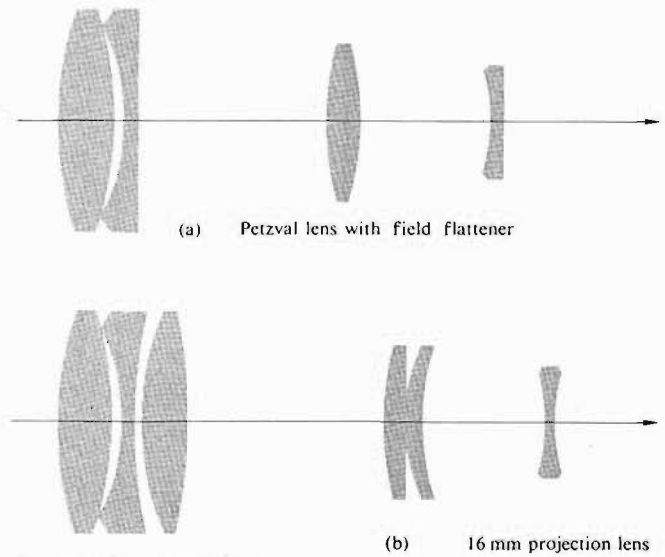


Fig. 6.26 The field flattener.

optical axis and so there are no off-axis points and in principle no coma or astigmatism. Instead of attempting to flatten the image surface, curvature is coped with by simply shaping the

v) Distortion

The last of the five primary, monochromatic aberrations is *distortion*. Its origin lies in the fact that the transverse magnification, M_T, may be a function of the off-axis image distance, y_i. Thus, that distance may differ from the one predicted by paraxial theory in which M_T is constant. In the absence of any of the others, this aberration is manifest in a misshaping of the image as a whole, even though each point is sharply focused. Consequently, when processed by an optical system suffering *positive* or *pin-cushion distortion*, a square array deforms as in Fig. 6.29(b). In that instance, each image point is displaced radially outward from the center with the most distant points moving the greatest amount, i.e., M_T increases with y_i. Similarly *negative* or *barrel distortion* corresponds to the situation where M_T decreases with the axial distance and, in effect, each point on the image moves radially inward toward the center [Fig. 6.29(c)]. Distortion can easily be seen by just looking through an aberrant lens at a piece of lined or graph paper. Fairly thin lenses will show essentially no distortion whereas ordinary positive or negative, thick, simple lenses will generally suffer positive or negative distortion, respectively.

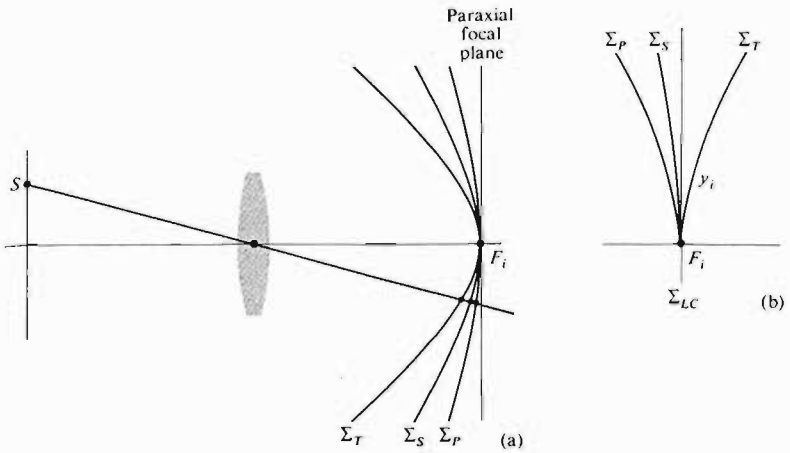
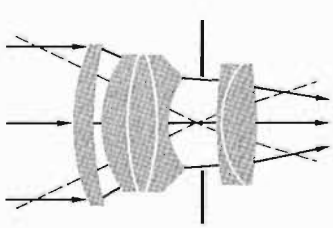


Fig. 6.27 The tangential, sagittal and Petzval image surfaces.

The introduction of a stop into a system of thin lenses is invariably accompanied by distortion, as indicated in Fig. 6.30. One exception to this is when the aperture stop is at the lens so that the chief ray is, in effect, the principal ray (i.e. it passes through the principal points, here coalesced at O). If the stop is in front of a positive lens, as in Fig. 6.30(b), the object distance measured along the chief ray will be greater than it was with the stop at the lens ($S_2A > S_2O$). Thus x_o will be greater and (5.26) M_T will be smaller—ergo, barrel distortion. In other words, M_T for an off-axis point will be less with a front stop in position than it would be without it. The difference is a measure of the aberration which, by the way, exists regardless of the size of the aperture. In the same way a rear stop [Fig. 6.30(c)] decreases x_o along the chief ray, (i.e., $S_2O > S_2B$), thereby increasing M_T and introducing pin-cushion distortion. *Interchanging the object and image thus has the effect of changing the sign of the distortion* for a given lens and stop. The aforementioned stop positions will produce the opposite effect when the lens is negative.

All of this rather suggests using a stop midway between identical lens elements. The distortion from the first lens



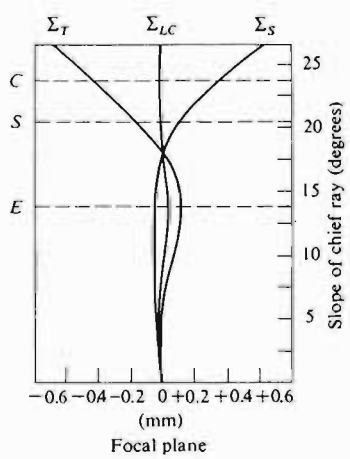


Fig. 6.28 A typical Sonnar. The markings C, S, and E denote the limits of the 35 mm film formate (field stop) i.e. corners, sides and edges. The Sonnar family lies between the double Gauss and the triplet.

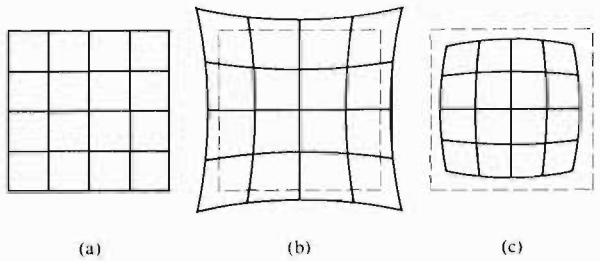


Fig. 6.29 Distortion.

would precisely cancel the contribution from the second. This approach has been used to advantage in the design of a number of photographic lenses (Fig. 5.100). To be sure, if the lens is perfectly symmetrical and operating as in Fig. 6.30(d) the object and image distances will be equal and hence $M_T = 1$. (Incidentally, coma and lateral color would then be identically zero as well.) This applies to (finite conjugate) copy lenses used, for example, to record data. Nonetheless, even when M_T is not one, making the system approximately symmetrical about a stop is a very common practice since it markedly reduces these several aberrations.

Distortion can arise in compound lens systems, as for example in the telephoto arrangement shown in Fig. 6.31. For a distant object point, the margin of the positive achromat serves as the aperture stop. In effect, the arrangement is like a negative lens with a front stop and so it displays positive or pin-cushion distortion.

Suppose a chief ray enters and emerges from an optical system in the same direction as e.g. in Fig. 6.30(d). The point at which the ray crosses the axis is the optical center of the system; but at the same time, since this is a chief ray, it is also the center of the aperture stop. This is the situation approached in Fig. 6.30(a) with the stop up against the thin lens. In both instances the incoming and outgoing segments of the chief ray are parallel and there is zero distortion, i.e. the system is *orthoscopic*. This also implies that the entrance and exit pupils will correspond to the principal planes (if the system is immersed in a single medium—see Fig. 6.2). Bear in mind that the chief ray is now a principal ray. *A thin-lens system will have zero distortion if its optical center is coincident with the center of the aperture stop.* By the way, in a pinhole camera, the rays connecting conjugate object and image points are straight and pass through the center of the aperture stop. The entering and emerging rays are obviously parallel (being one and the same) and there is no distortion.

6.3.2 Chromatic Aberrations

The five primary or Seidel aberrations have been considered in terms of monochromatic light. To be sure, if the source had a broad spectral bandwidth these aberrations would be

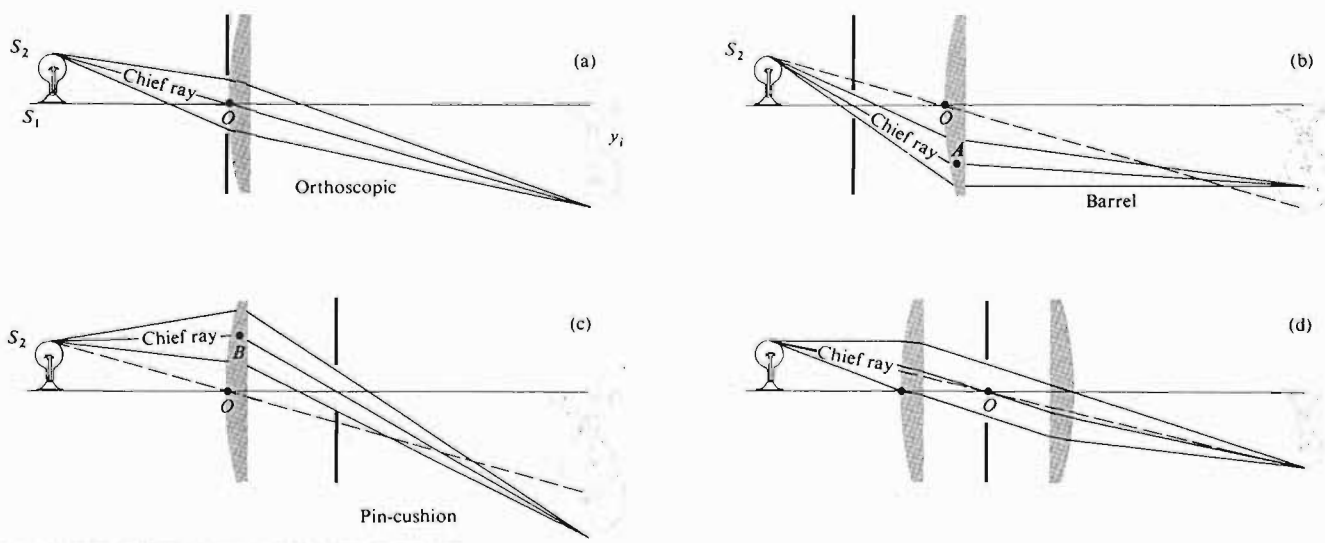


Fig. 6.30 The effect of stop location on distortion.