CHAPTER 4
OPTICAL ABERRATIONS

4.1 *Introduction*

We have hitherto made the assumption that a lens or a curved mirror is able to form a point image of a point object. This may be approximately true if the depth of the mirror or the thickness of the lens is small compared with other distances, and if the angle that all rays make with axis of the mirror or lens is small, and if we are using monochromatic light. Usually none of these conditions is satisfied exactly, and consequently the image formed by a lens or curved mirror suffers from several aberrations.

There are five *geometrical* aberrations, given the names

Spherical aberration
Astigmatism
Coma
Curvature of field
Distortion (pincushion or barrel distortion).

In addition, unless we are using monochromatic light, lenses (but not mirrors) exhibit chromatic aberration (longitudinal and transverse).

It may be possible to minimize some of these aberrations by careful choice of the radii of curvature of a lens system ("bending the lens"), although the condition for minimizing one aberration may be different from minimizing another. Consequently some sort of compromise must be reached, which may depend on which aberrations are important, and which are not so important, for a particular application.

4.2 *Spherical Aberration*

We'll begin by looking at the spherical aberration resulting from reflection from a spherical mirror. We have hitherto assumed that a parallel beam of light, after reflection from a spherical mirror, comes to a focus at a point, and that the distance of the focal point from the surface of the mirror is half the radius of curvature of the mirror, as in figure IV.1:

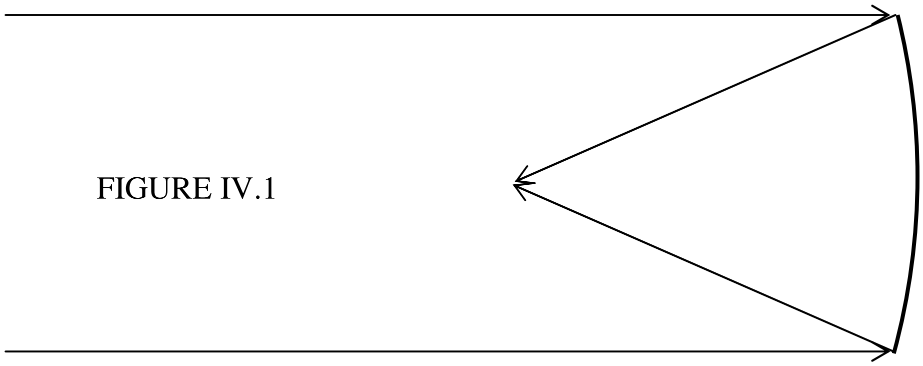


FIGURE IV.1

This is approximately true for a small aperture mirror ("aperture" meaning the ratio of the diameter to the focal length). This is not the case, however, for a large aperture mirror. In figure IV.2 I have drawn a hemispherical mirror. I assume that there is an incident beam of light (not drawn) coming in horizontally from the left, and I have drawn the rays after reflection from the mirror. (Some of the rays will be reflected a second time from the surface before eventually escaping, but I have not drawn the rays after a second reflection because they would only clutter up the diagram and are not pertinent in describing what I want to describe. You can see that the reflected rays are bounded by an envelope known as a *caustic curve*, shown as a dashed red curve in figure IV.2.

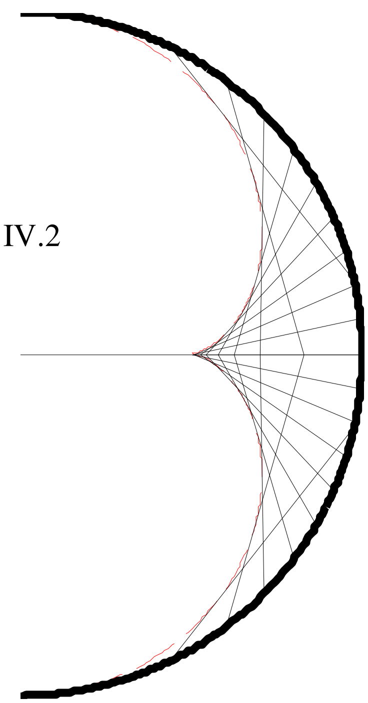


FIGURE IV.2

Can we find the equation to this caustic curve?

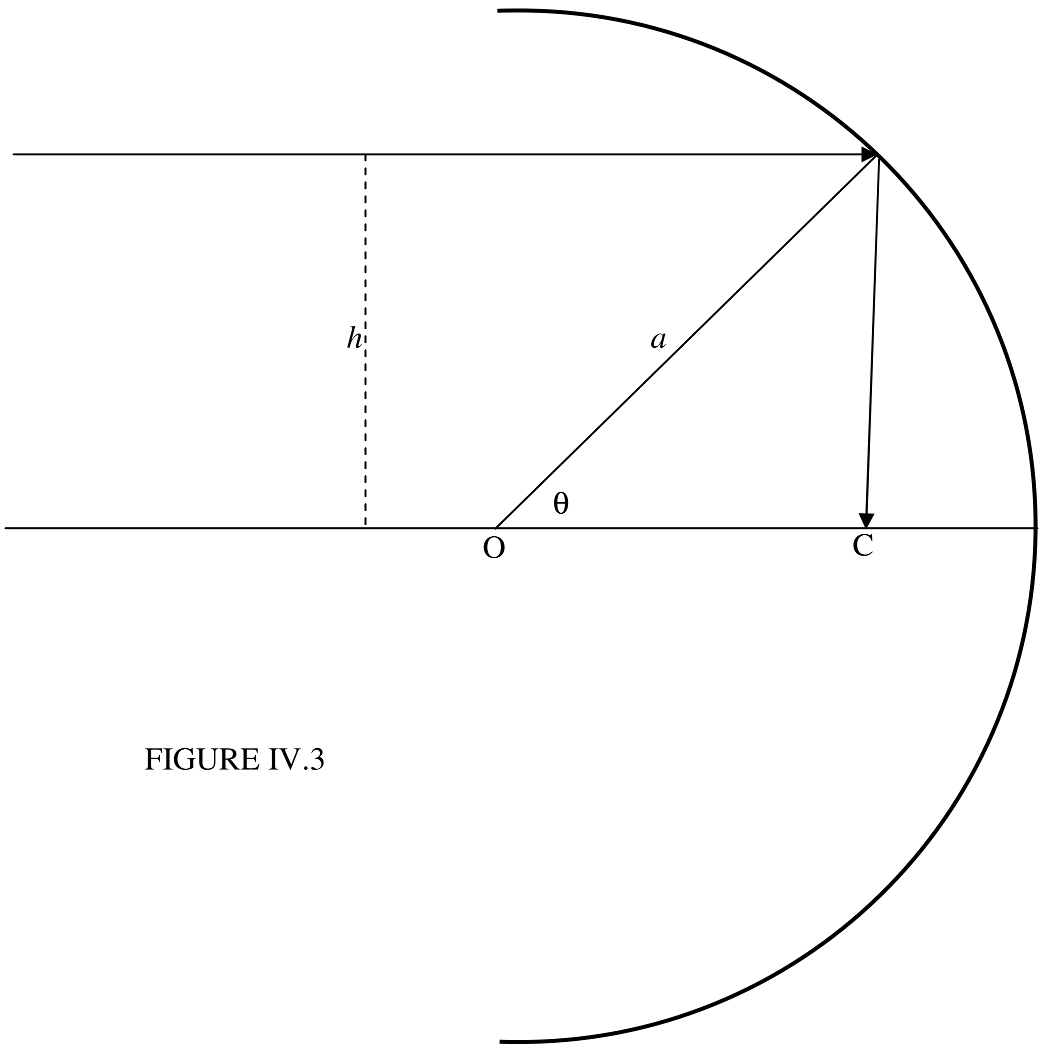


FIGURE IV.3

We'll take the centre of curvature of the mirror as origin O of coordinates, and suppose that the radius of curvature of the mirror is a. Let us consider the adventures of a ray of light coming in parallel to the horizontal (x) axis and at a height h from it. The equation to the incoming light ray is just $y = h$, and the equation to the mirror surface is $x^2 + y^2 = a^2$. A little bit of coordinate geometry will enable us to determine that the equation to the reflected ray is

$$y = \frac{h}{a^2 - 2h^2}\left[\left(2\sqrt{a^2 - h^2}\right)x - a^2\right], \qquad 4.2.1$$

and that it crosses the x-axis at a point C such that

$$\mathrm{OC} = \frac{a^2}{2\sqrt{a^2 - h^2}}. \qquad 4.2.2$$

It is also convenient to write these formulas in terms of the angle θ, which is given by $h = a\sin\theta$. After a little algebra and application of some trigonometric identities, we obtain

$$y = x\tan 2\theta - \frac{a\sin\theta}{\cos 2\theta} \qquad 4.2.3$$

for the equation to the reflected ray, and

$$\mathrm{OC} = \tfrac{1}{2}a\sec\theta. \qquad 4.2.4$$

We can write equation 4.2.3 as

$$f(x, y\ ;\ \theta) = x\tan 2\theta - \frac{a\sin\theta}{\cos 2\theta} - y = 0. \qquad 4.2.5$$

From our long-forgotten, yellowed and mildewy mathematics notes, we recall that to find the equation to the envelope of a family of curves of the form $f(x, y; \theta) = 0$, we have to eliminate the parameter θ from that equation and the equation $\frac{\partial f}{\partial\theta} = 0.$ After some more algebra and more application of trigonometric identities, we find that the latter equation comes to

$$x = a\cos\theta.(\tfrac{3}{2} - \cos^2\theta). \qquad 4.2.6$$

So, all we have to do is to eliminate the parameter θ from equations 4.2.3 and 4.2.6, and this would give us the x, y equation to the caustic curve. These two equations are, in fact, the parametric equations to the caustic curve. Now I don't know how easy it would be to eliminate θ. Since equation 4.2.6 is a cubic equation in cos θ, I suspect that it might not be particularly easy. But (as is often the case with two parametric equations to a curve) we can happily plot the curve numerically, without having to eliminate the parameter algebraically. Thus, in order to plot the red curve in figure IV.2, I varied θ from −90º to +90º, and calculated x from equation 4.2.6, and I then calculated y from equation 4.2.3.

To avoid spherical aberration, telescope mirrors can be made in a paraboloidal shape. It can be shown that an incident beam of light, coming in parallel to the axis of a paraboloidal mirror, after reflection will come to single focal point, namely at the focus of the parabola. A proof of this is given in Section 2.4 of Chapter 2 of my Celestial

Mechanics notes http://orca.phys.uvic.ca/~tatum/celmechs/celm2.pdf and is not repeated there. In that Chapter, it is also shown that, if a bucket of liquid is rotated about a vertical axis, the surface of the liquid will take up a paraboloidal shape, and mention is made there of two applications to the manufacture of paraboloidal mirrors. In one, a vat of molten glass is rotated, and is gradually cooled down until the glass solidifies into a paraboloidal shape. In the other, a container of mercury is rotated, the surface of the mercury taking up a paraboloidal shape, and this liquid paraboloid is then used as the main mirror of a reflecting telescope. While it can observe only close to the zenith, some excellent results have been obtained. I shan't repeat it here, but you might want to refer to the above-mentioned notes, since it is pertinent here.

This property (of light being reflected from the surface of a parabola to a single focal point) applies only to light coming in parallel to the axis of the paraboloid. Consequently paraboloidal telescope mirrors have only a rather narrow field of view. A *Schmidt* telescope uses a spherical mirror (hence a large field of view) and, to avoid spherical aberration, a corrector plate is mounted in front of the mirror. Typically the spherical mirror is at the "bottom end" of the telescope tube, and the corrector plate is at the "top end". The corrector plate causes light that is coming in parallel to the telescope tube, but some distance from the axis of the tube, to diverge slightly from the axis before reaching the spherical mirror. In this manner all of the incoming light, after reflection from the mirror, comes to a focus at a single point.

A lens also suffers from spherical aberration, of course, but it does not lend itself to such simple analysis as for a spherical mirror. One needs to perform detailed numerical ray-tracing to find the exact shape of the caustic curve for a lens. We showed, however, in Section 1.4 of Chapter 1, that refraction even at a plane surface produces spherical aberration.

One might wonder, given that a paraboloidal mirror when used on axis is free of spherical aberration, whether a lens made with paraboloidal surfaces, is also free of spherical aberration. Alas, that is not so.

One can, however, design a lens with spherical surfaces that minimize the spherical aberration, by suitable choice of the radii or curvature of the lens surfaces. This is called "bending the lens".

For example, figure IV.4 shows five lenses, in which I have written, beside each surface, its radius of curvature in cm. In what follows I assume that the lens is "thin" in the sense that its thickness is very small compared with any other distances under discussion. If the refractive index is 1.6, each of these lenses has a focal length of 20 cm.

You can characterize the shape of a lens by means of its *shape factor*

$$q = \frac{r_1 + r_2}{r_1 - r_2}\,. \qquad 4.2.7$$

In figure IV.4 I have written the shape factor above each lens.

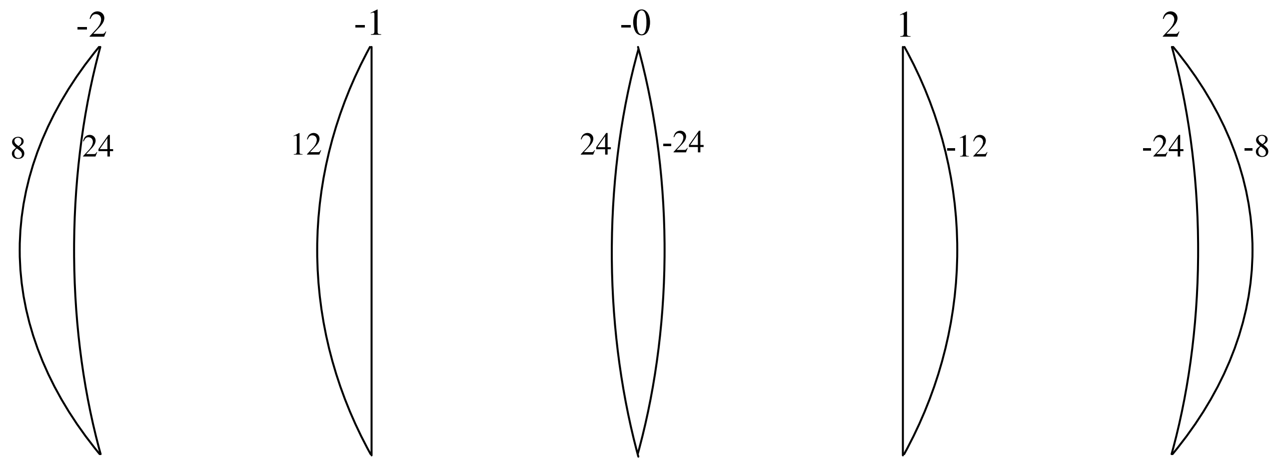


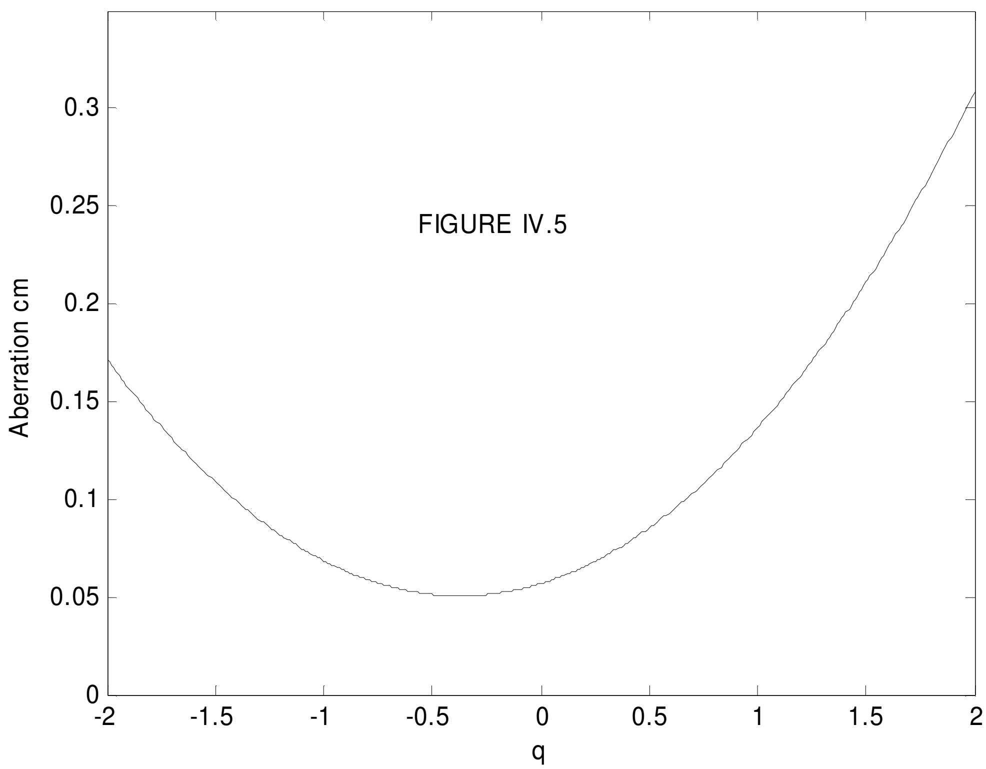
FIGURE IV.4

For light coming in horizontally near the axis, the focal length of each of these lenses is 20 cm. However, light coming in horizontally at some distance from the axis, after passage through the lens, falls a little short of 20 cm. We may characterize the spherical aberration by the amount it falls short. Assuming that the lenses are thin (compared with any other distances under consideration) I calculated the shortfall for a ray of light coming in from the left at a height of 1 cm from the axis. This is shown in figure IV.5, in which I have drawn the shortfall (labelled "Aberration" in the figure) versus shape factor q. It is seen that the aberration is least for a shape factor of about $q = -0.38$. The radii of curvatures of the lens must satisfy equation 4.2.7 as well as
$q = -0.38$

$$\frac{1}{f} = (n-1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right), \qquad 4.2.8$$

so that, for $f = 20$ cm and $q = -0.38$, the radii of curvature for least spherical aberration should be $r_1 = 17.4$ cm and $r_2 = -38.7$ cm.

Of course, you have to use the lens the right way round! If you turn it round, or if light is coming in from the right, the shape factor is +0.38, and the spherical aberration is not at a minimum. Mind you, the minimum is fairly shallow, so you can vary the shape factor a fair amount without grossly increasing the spherical aberration.



4.3 *Astigmatism*

In Greek, *stigma* means a mark - in particular the mark made by the prick of a pointed instrument. An ideal optical instrument produces an image of a point source, which is also a point. If the image is not a point, then it is astigmatic. However, the use of the word astigmatic to describe an image of a point source that is not also a point is restricted to the kind of optical aberration described in this section.

The easiest way to understand the phenomenon of astigmatism is to imagine a lens (or mirror) whose surfaces are not exactly spherical but for which the radius of curvature (and hence focal length) in one plane is different from the radius of curvature in a plane at right angles to the first.

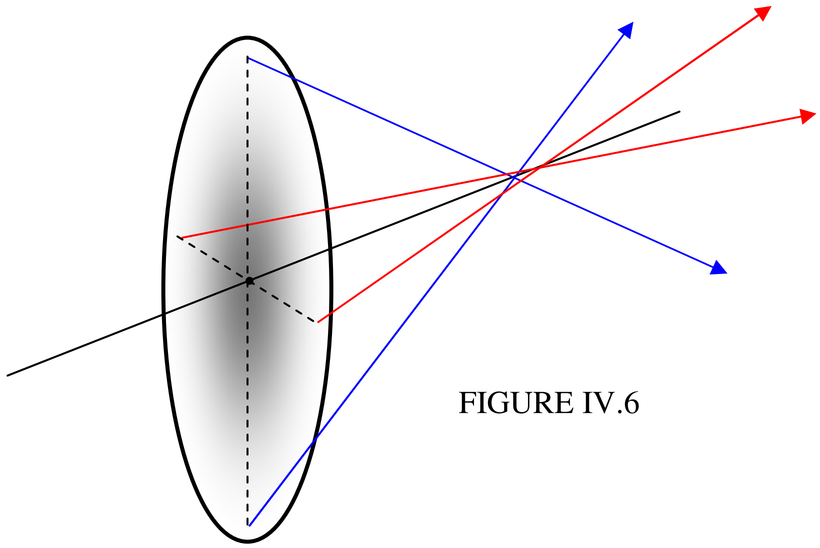


FIGURE IV.6

This is going to stretch my poor artistic abilities to the limit. The ellipse is intended to represent a lens seen somewhat from behind, at an angle. The black line is its optical axis. The lens is supposed to be illuminated from the left with a beam of light parallel to the optical axis. I have drawn two transmitted rays in the vertical plane by means of two blue arrows converging on to the optical axis at some distance from the lens. I have drawn two transmitted rays in the horizontal plane by means of two red arrows converging on to the optical axis at a slightly greater distance from the lens. (The colours of the arrows are not intended to means different colours of light. We'll suppose that the light is all monochromatic.) Evidently the rays from different points around the circumference of the lens come to a confused mess on the optical axis; the image is decidedly astigmatic. In fact at one point on the optical axis, a line image is formed; a little further along the axis, another line image, at right angles to the first, is formed. In figure IV.7, I repeat figure IV.6, but I add these two line images.

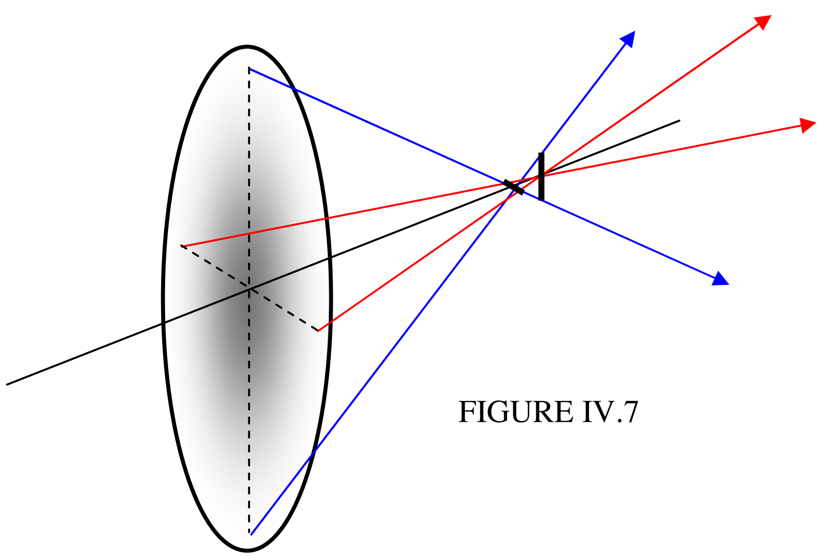
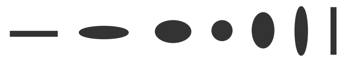


FIGURE IV.7

As you move along the optical axis, the image changes from a horizontal line to a vertical line in a sequence that looks something like this:



Somewhere about half way between the two linear images is the “circle of least confusion”.

I have explained the aberration of astigmatism by supposing that the lens has a different focal length in one plane than in the other. This may be the easiest way for an introductory explanation of the aberration. However, in practice it is unlikely that a lens has different focal lengths in two orthogonal planes; indeed it would be quite difficult to make such a lens.

In most cases astigmatism is caused, as we shall see, by using a perfectly good lens or mirror *off-axis*.

If you look at a star through a telescope, and if you move the eyepiece in and out as you look through it, you may see the star image going through a series of astigmatic images such as illustrated above. This is not usually caused by a bad lens, but is caused if the object glass (in a refracting telescope) or lens (in a reflecting telescope) is crooked in its cell, so that you are using it off-axis. Indeed, doing this little test is a good way of telling whether the object glass or the mirror is crooked in its cell.

Although different radii of curvature in different planes is not the usual cause of astigmatism, there is an exception - namely, the human eye. If the radii of curvature of the cornea, or of the lens, is different in different planes, then the image on the retina will be astigmatic even on-axis.

We saw in Chapter 1 that refraction at a plane surface produces *spherical aberration*. It is not always appreciated that refraction at a plane surface produces *astigmatism* when the surface is viewed at an angle. If you visit an aquarium and look into glass side of a tank at an angle, you will see that the fish look a little blurred because of this astigmatism.

In figure IV.8 I have drawn two rays from a point O at the bottom of a glass block, making angles of 20º and 30º with the normal to the upper surface.. With a refractive index of 1.6 the angles that the emerging rays make with the normal are 33º and 53º. I refer to the plane of the paper (or your computer screen) as the *tangential* plane. A vertical plane perpendicular to the plane of the paper is the *sagittal* plane. You will see that the two rays in the tangential plane diverge, after refraction, from a point T in the tangential plane. If we take the height of the glass block to be 1, we can calculate that the (x , y) coordinates of the point T in the tangential plane are (0.145 , 0.666).

To anticipate, the image at T is not a point; rather, it is a short horizontal line in the sagittal plane, perpendicular to the plane of the paper.

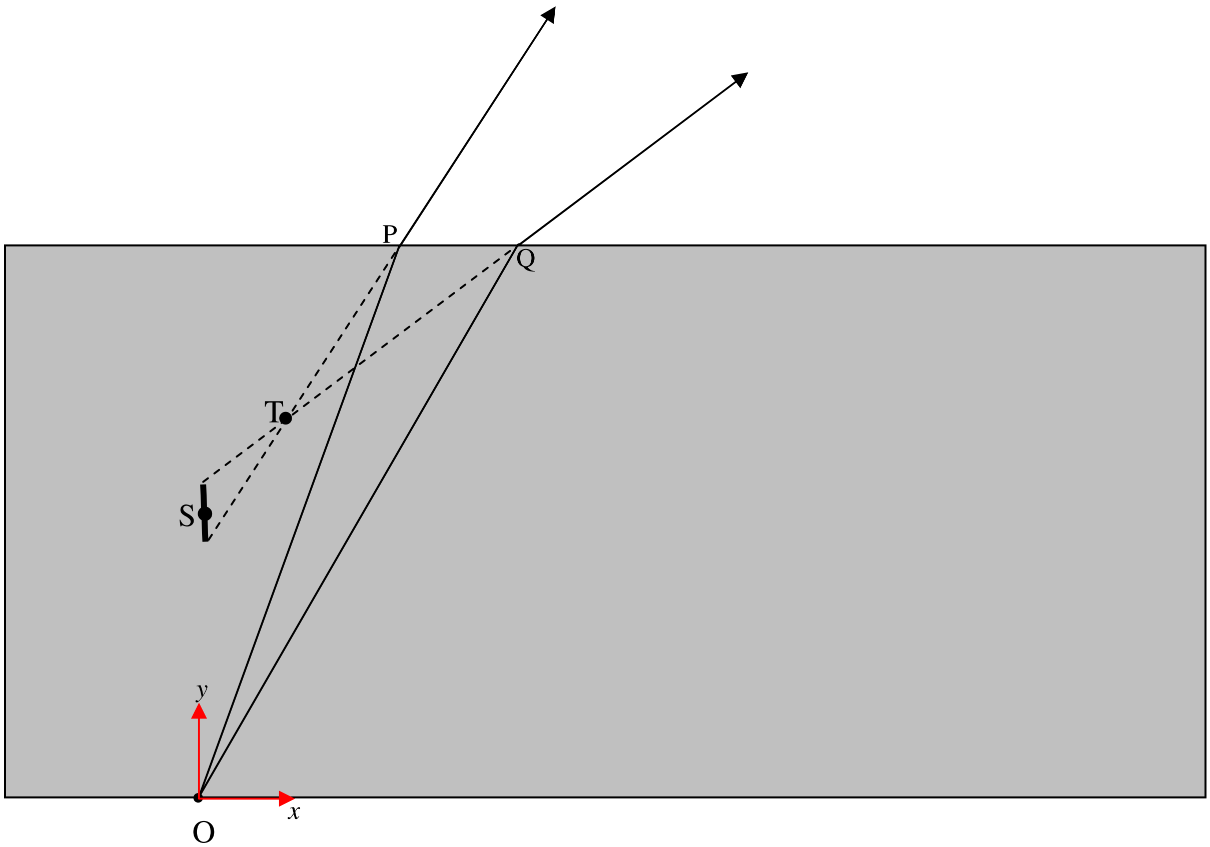


FIGURE IV.8

Now let's look at the glass block from above:

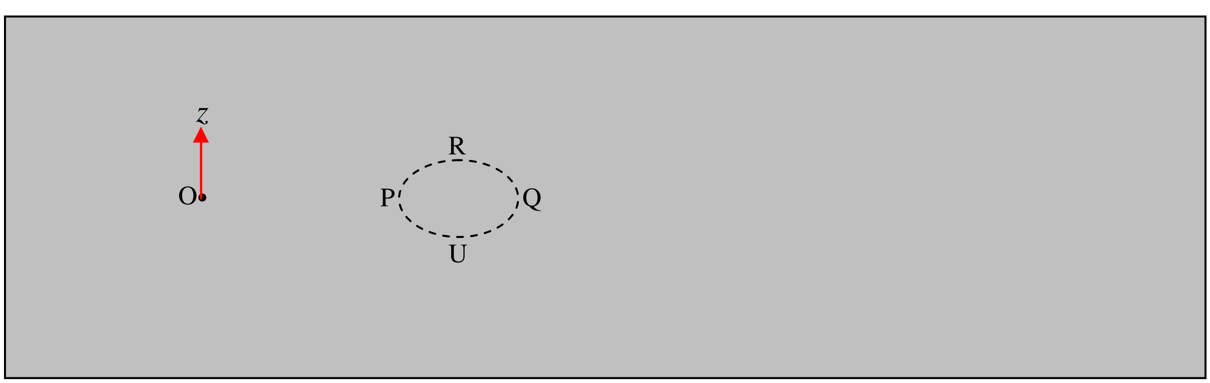


FIGURE IV.9

I have drawn (accurately, I hope, after some calculation) the ellipse where the cone of light coming from O intersects the upper surface of the block. The point P and Q are in the tangential plane, and light emerging from P and Q appears to diverge from T. The points R and U are in the sagittal plane. Tracing the rays OR and OU after emergence from the block doesn't look very easy, but it will probably be agreed that they do not diverge from T, as the rays in the tangential plane did. Indeed, after further thought, you'll probably see that the rays OR and OU after emergence will be diverging from a point S on the *y*-axis; that is to say, directly above O.

For reference, the coordinates of the several points on the drawing, if my calculations are correct, are:

O:	(0.000 000 000 ,	0.000 000 000 ,	0.000 000 000)
P:	(0.363 970 234 ,	1.000 000 000 ,	0.000 000 000)
Q:	(0.577 350 269 ,	1.000 000 000 ,	0.000 000 000)
R:	(0.470 660 252 ,	1.000 000 000 ,	0.069 344 256)
U:	(0.470 660 252 ,	1.000 000 000 ,	−0.069 344 256)
T:	(0.145 831 216 ,	0.666 360 298 ,	0.000 000 000)
S:	(0.000 000 000 ,	0.497 301 940 ,	0.000 000 000)

The angle of incidence of the ray at R is 25.442 358 40 degrees to the normal, and the angle of refraction is 43.421 850 83 degrees.

The net result of this is that there is a short linear "image" at T perpendicular to the tangential plane, and a short linear "image" at S perpendicular to the sagittal plane, and, somewhere in between, there is a circle of least confusion. One way of looking at the situation is to recognize that the wavefront of the emergent cone is nonspherical - its radii of curvature are different in the tangential and sagittal planes.

Thus refraction at a plane surface results in both spherical aberration and astigmatism. Refraction through a glass prism, as in a prism spectrograph, also produces astigmatism, and it can be shown that the astigmatism is least when the light passes through the prism symmetrically in the position of minimum deviation. This is one reason why prism spectrographs are normally used in the position of minimum deviation.

We have seen that a lens does not produce a point image of a point object on the axis of the lens, but the image is subject to spherical aberration. The spherical aberration is small if the aperture of the lens is small compared with its focal length and object and image distances, so that the angles that the various rays make with the optic axis are small enough that one can make the approximation $\sin\theta \approx \tan\theta \approx \theta$, and is small also if the shape of the lens is suitably designed as in the example in Section 4.2. For a point object on the axis, the image is free of astigmatism (presuming that the radii of curvature of the lens in the tangential and sagittal planes are equal). However, for a point object *off-axis*, in which the light passes through the lens at an oblique angle, the refracted cone gives rise to an astigmatic image in just the same way as for oblique refraction at a plane surface. This there will be a line "image" normal to the tangential plane, and, at a different distance, there will be another line normal to the sagittal plane , and a circle of least confusion between them. The further off-axis the object, the greater will be the distance between the tangential and sagittal lines. (The distance will be zero for a point object on axis.) Unlike the case for spherical aberration, the amount of astigmatism (the distance between T and S) is not greatly improved by changing the shape of the lens, and a third lens component is often used to correct for the astigmatism.

We mentioned, however, that astigmatism in the eye is generally caused by different tangential and sagittal curvatures of the cornea, and it is evident on axis as well as off axis. It may be corrected by a single lens, which is designed to have different tangential and sagittal curvatures. Such lenses are not easy to make, and they are generally fairly expensive.

4.4 *Coma*

Coma, like astigmatism, is another aberration that appears off axis, near the edge of an image field. If you look at a wide-field photograph of some stars taken with a photographic telescope, the stars near the centre of the field should be points, but, at the very edge of the photograph, if the telescope is less than perfect, the stars may appear like little comets, with a sharp nucleus, but each with a fuzzy tail directed away from the centre of the photograph. This aberration is called "coma". The word "coma", as well as the word "comet", comes from the Latin *coma*, meaning "hair", from a fanciful resemblance of a comet, or of a comatic image of a star, to the head of a girl with her long hair streaming out behind her.

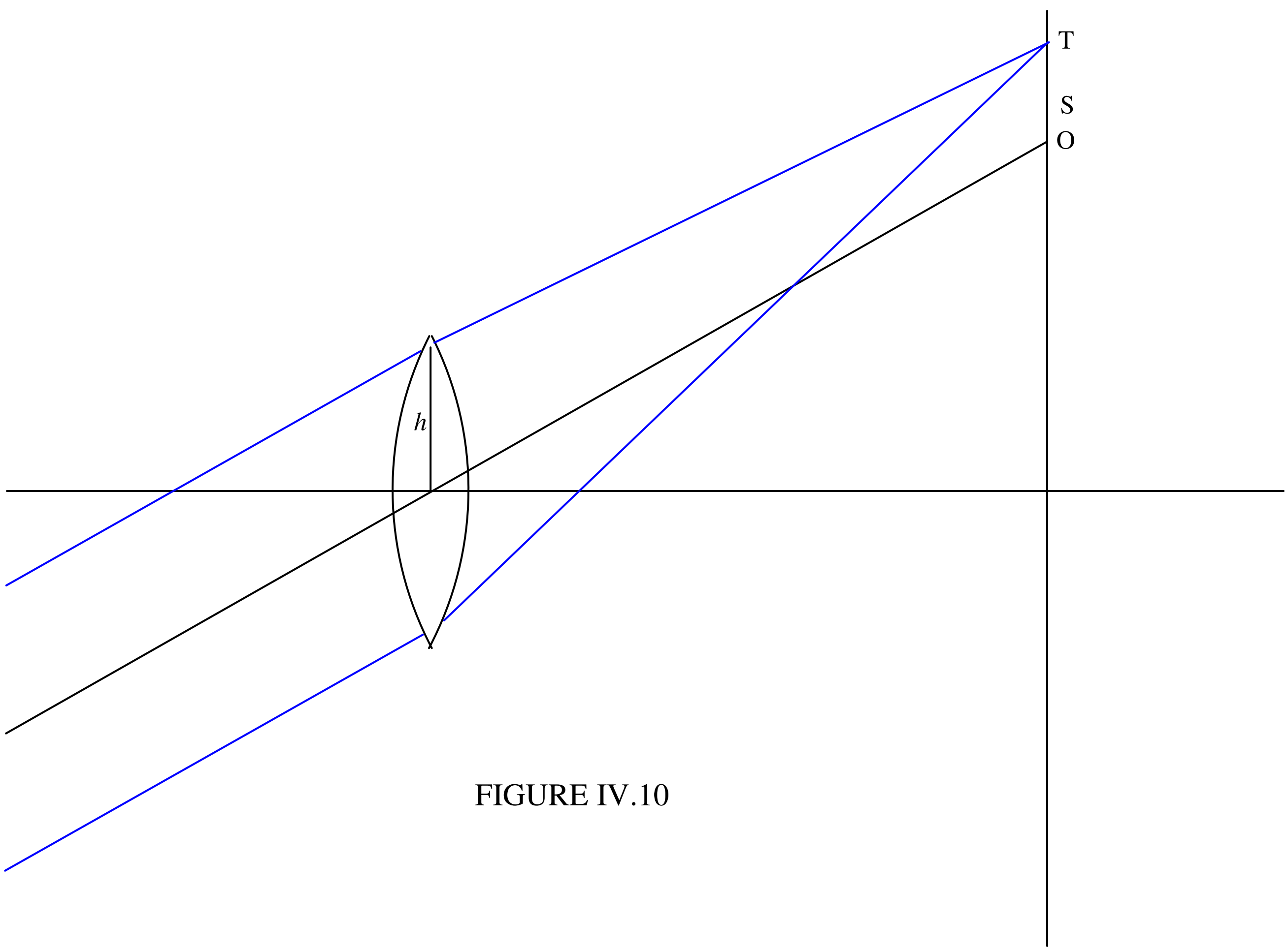


FIGURE IV.10

In figure IV.10, we see a parallel bunch of rays entering the lens obliquely from the left. The central ray, in black, goes straight through to a point O. Two rays in the tangential plane (i.e. the plane of the computer screen, or the paper, if you have printed it out) converge not to the point O, however, but to a point T as shown. If I could draw two rays equally far from the centre of the lens but in the sagittal plane (i.e. a vertical plane perpendicular to the plane of the paper), they would converge to a point S, about a third of the way between O and T.

If I could draw the rays entering the lens all around the zone of radius h on the lens, each pair of opposite rays would converge to a point on the *comatic circle*. See figure IV.11.

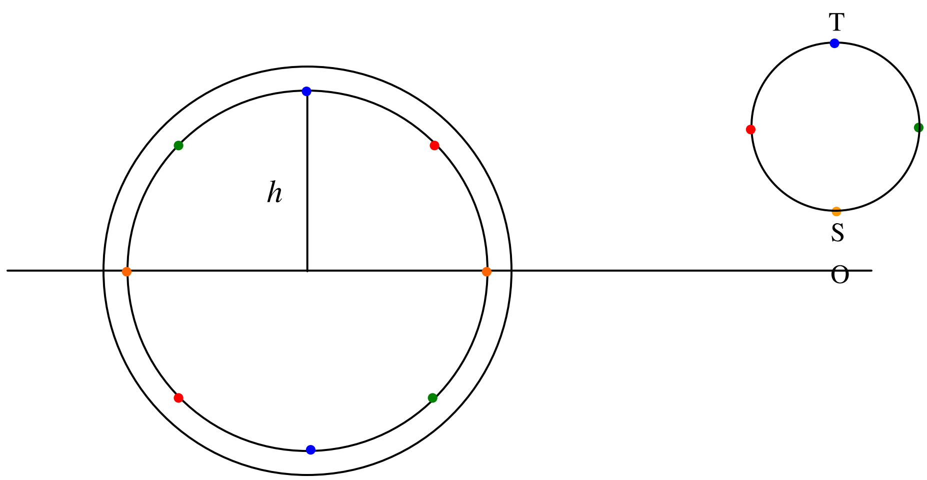
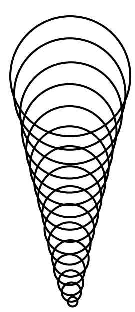


FIGURE IV.11

The radius and height of the comatic circle is different for each zone on the lens that produces it, with the result that the "image" appears as a superposition of all the comatic circles produced by all the zones on the lens, something like the drawing below. That, at least is a qualitative description of the phenomenon.



To go further is a bit of a specialist skill, so I'll leave it here. Suffice to say that the degree of coma and the degree of spherical aberration depend on the shape factor of the lens, and fortunately the shape that gives least spherical aberration is not very different from the shape that gives least coma.

The aberrations discussed so far are aberrations that result when the lens or mirror does not produce a point image of a point object. If, somehow, we manage to get rid of spherical aberration, astigmatism and coma, then a point object will result in a point image. But will that image be in the right place? There are two further aberrations that are concerned with where the image is formed. These aberrations are *curvature of field* and *distortion*.

4.5 *Curvature of Field*

Suppose we have a lens that we have managed to correct for (or at least to minimize) spherical aberration, astigmatism and coma, say by a combination of choosing the right shape of the lens and not going too far off-axis. (I.e. we might close the lens to an aperture of *f*/11 rather than opening it up to *f*/5.6.) Nothing that we know about refraction and lenses and mirrors tells us that that light coming in at different angles to the axis forms point images conveniently situated in a plane, as illustrated hopefully in figure IV.13.

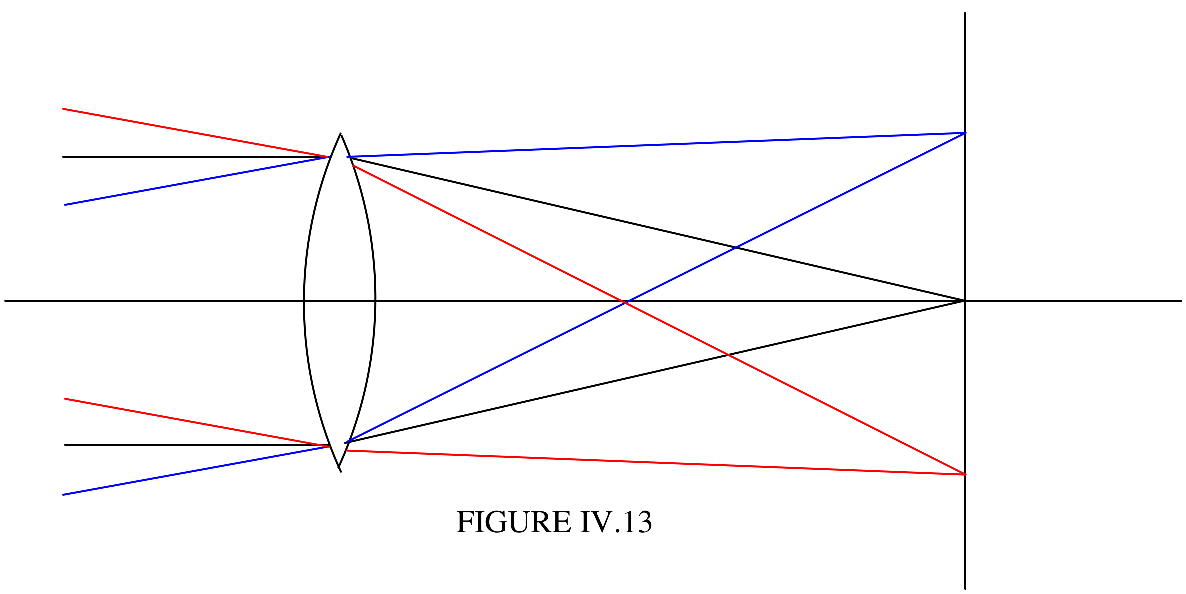


FIGURE IV.13

Alas, life is not as simple as that, and light doesn't generally come to a focus in a focal plane, but rather in a curved focal surface (sometimes called the *Petzval surface*), as in figure IV.14.

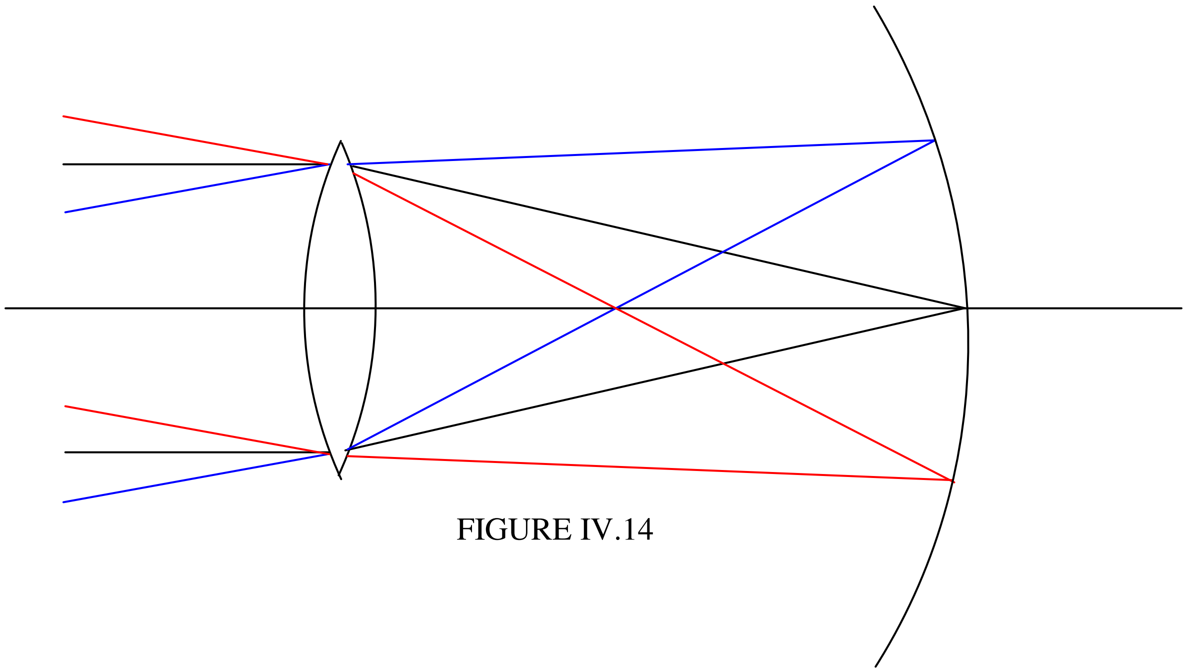


FIGURE IV.14

This doesn't matter a great deal in a telescope designed merely for looking through, since the eye can rapidly accommodate for slightly different image distance, but it obviously matters in a photographic telescope. One effective way of dealing with this problem, particularly if your detector is a flexible film, is to shape the filmholder so that the film fits along the Petzval surface. This is often done, for example, with Schmidt astronomical telescopes.

In designing a lens or lens system, the problems of astigmatism and curvature of field are often closely related. For example a meniscus lens tends to suffer from astigmatism, and there is a focal surface for the tangential image, and a focal surface for the sagittal image, and the tangential and sagittal surfaces curve in opposite senses. With luck, or more likely with some careful design, the surface (C) for the loci of the circles of least confusion is between the tangential (T) and sagittal (S) surfaces and is approximately planar (figure IV.15).

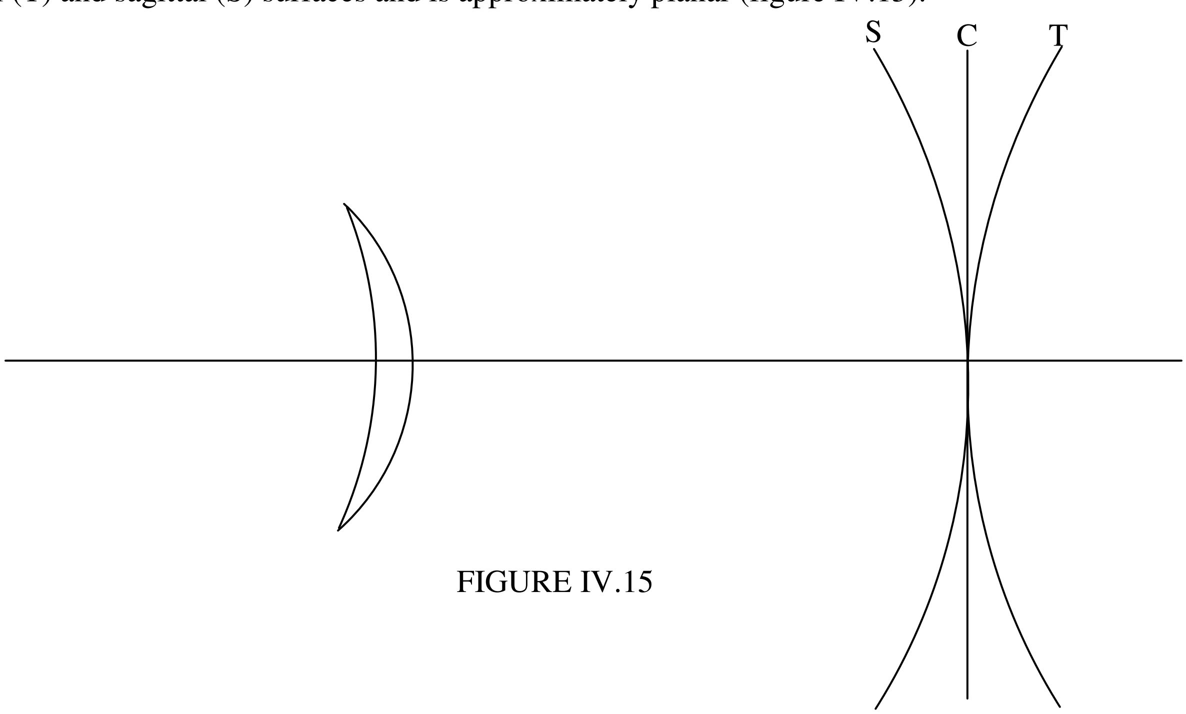


FIGURE IV.15

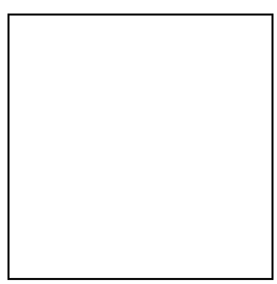
It has been shown that, if you have a doublet lens, made of two lenses, one a converging lens of focal length f_1 and refractive index n_1, and the other a diverging lens of focal length f_2 and refractive index n_2, curvature of field will be least if $\frac{1}{n_1 f_1} + \frac{1}{n_2 f_2} = 0$. For example if you have two glasses, of refractive indices $n_1 = 1.51$ and the other of refractive index $n_2 = 1.67$, and you want to make a doublet lens of focal length 100 cm, what should be the focal lengths of the two components of the doublet if you want to minimize curvature of field?

Answer: The lenses need to satisfy $\frac{1}{n_1 f_1} + \frac{1}{n_2 f_2} = 0$ and $\frac{1}{f_1} + \frac{1}{f_2} = \frac{1}{f}$. It's probably easier to work in terms of powers rather than focal lengths, so we have to solve $1.67P_1 + 1.51P_2 = 0$ and $P_1 + P_2 = 0.01$. This gives $P_1 = -0.094375 \text{ cm}^{-1}$ and $P_2 = +0.104375 \text{ cm}^{-1}$, or $f_1 = 10.60$ cm and $f_2 = 9.58$ cm. You will then have to design the lenses so that the faces of the two lenses that are in contact have the same radius of curvature, and we leave that to the reader.

For a similar problem concerning a doublet with minimum chromatic aberration, see Chapter 2, Section 2.10.

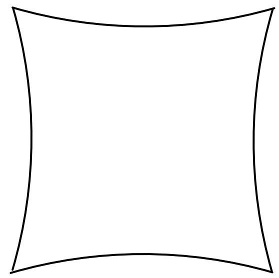
4.6 *Distortion*

Let us suppose that, by dint of great labour and overcoming many obstacles, we have finally designed a lens system that is free from chromatic aberration, spherical aberration, astigmatism, coma and curvature of field, or at least have minimized these aberrations or have come to a tolerable compromise for a particular purpose, can we at last relax? Unfortunately, no, we cannot. The magnification of an image is image distance divided by object distance, and image distance is different off-axis than on-axis, so the image magnification varies with distance from the axis. This means that the image of an object like this:



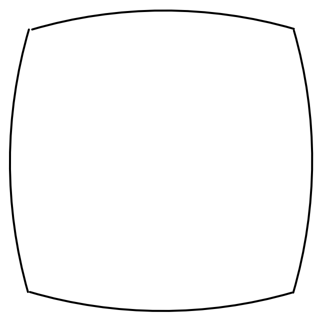
may look like this:

(pincushion
distortion)

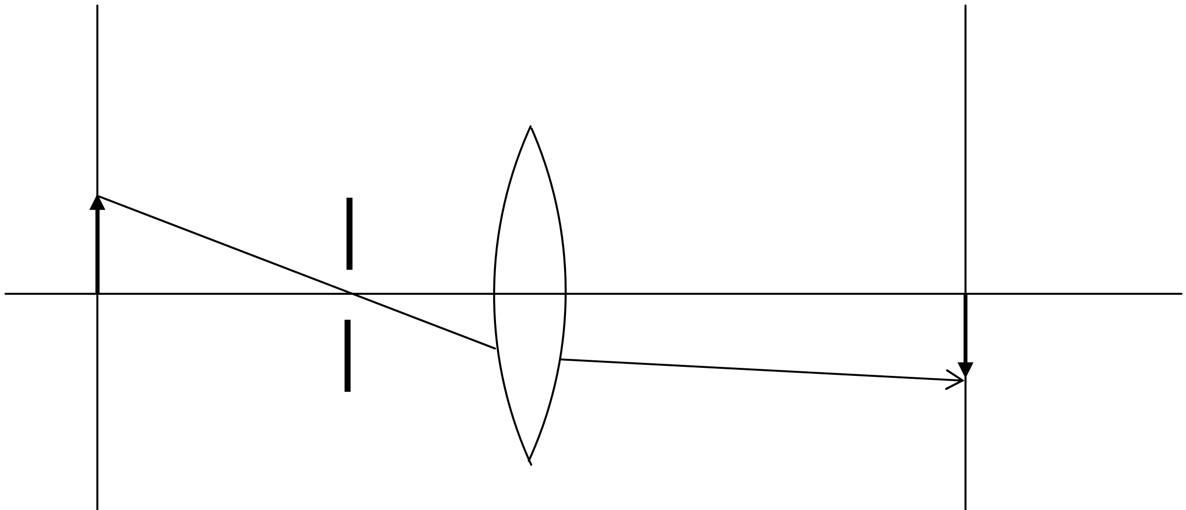


or like this:

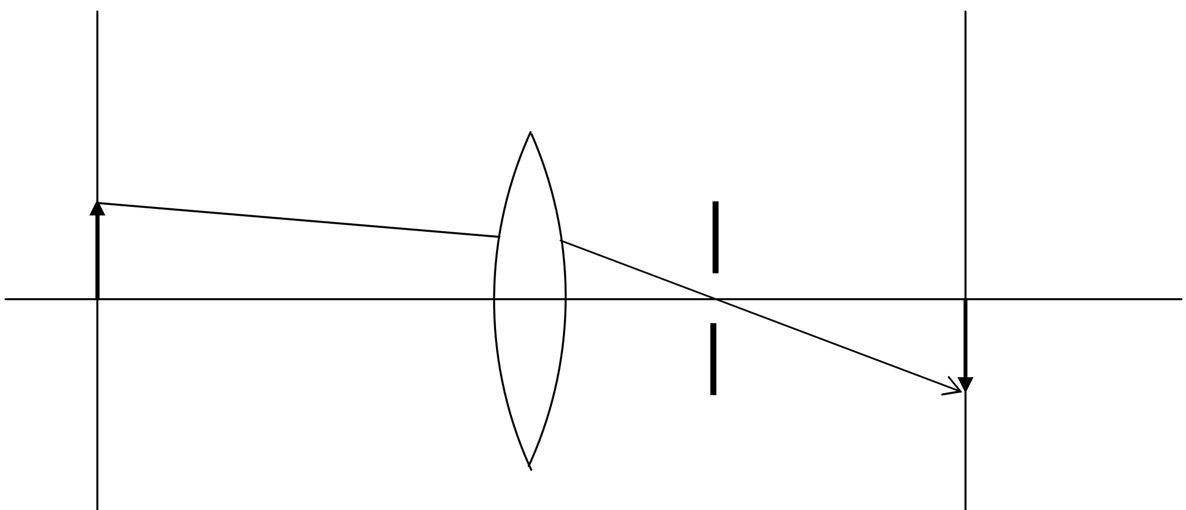
(barrel
distortion)



If the distortion is quite small, it may not be noticed in ordinary pictorial photography, but if one is using a photograph for precise positional measurements (for example, in astrometry) it is necessary to correct for the distortion. Often barrel distortion is introduced into a lens system if a stop is placed in front of a lens, while pincushion distortion results if a stop is placed behind a lens. The drawing below, in which I have exaggerated the situation by drawing a very small stop, may explain the reason why. I have placed the object at twice the focal distance from the lens, so that, on axis, the image and object distances are equal, and the magnification is unity. A symmetric air-spaced doublet with a stop half way between the two components minimizes distortion.



Off-axis, the image distance is less than the object distance, so the magnification is less off-axis than on-axis. Barrel distortion results.



Off-axis, the image distance is greater than the object distance, so the magnification is greater off-axis than on-axis. Pincushion distortion results.

FIGURE IV.16