

Geometrical Optics

27-1 Introduction

In this chapter we shall discuss some elementary applications of the ideas of the previous chapter to a number of practical devices, using the approximation called *geometrical optics*. This is a most useful approximation in the practical design of many optical systems and instruments. Geometrical optics is either very simple or else it is very complicated. By that we mean that we can either study it only superficially, so that we can design instruments roughly, using rules that are so simple that we hardly need deal with them here at all, since they are practically of high school level, or else, if we want to know about the small errors in lenses and similar details, the subject gets so complicated that it is too advanced to discuss here! If one has an actual, detailed problem in lens design, including analysis of aberrations, then he is advised to read about the subject or else simply to trace the rays through the various surfaces (which is what the book tells how to do), using the law of refraction from one side to the other, and to find out where they come out and see if they form a satisfactory image. People have said that this is too tedious, but today, with computing machines, it is the right way to do it. One can set up the problem and make the calculation for one ray after another very easily. So the subject is really ultimately quite simple, and involves no new principles. Furthermore, it turns out that the rules of either elementary or advanced optics are seldom characteristic of other fields, so that there is no special reason to follow the subject very far, with one important exception.

The most advanced and abstract theory of geometrical optics was worked out by Hamilton, and it turns out that this has very important applications in mechanics. It is actually even more important in mechanics than it is in optics, and so we leave Hamilton's theory for the subject of advanced analytical mechanics, which is studied in the senior year or in graduate school. So, appreciating that geometrical optics contributes very little, except for its own sake, we now go on to discuss the elementary properties of simple optical systems on the basis of the principles outlined in the last chapter.

In order to go on, we must have one geometrical formula, which is the following: if we have a triangle with a small altitude h and a long base d , then the diagonal s (we are going to need it to find the difference in time between two different routes) is longer than the base (Fig. 27-1). How much longer? The difference $\Delta = s - d$ can be found in a number of ways. One way is this. We see that $s^2 - d^2 = h^2$, or $(s - d)(s + d) = h^2$. But $s - d = \Delta$, and $s + d \sim 2s$. Thus

$$\Delta \sim h^2/2s. \tag{27.1}$$

This is all the geometry we need to discuss the formation of images by curved surfaces!

27-2 The focal length of a spherical surface

The first and simplest situation to discuss is a single refracting surface, separating two media with different indices of refraction (Fig. 27-2). We leave the case of arbitrary indices of refraction to the student, because *ideas* are always the most important thing, not the specific situation, and the problem is easy enough to do in any case. So we shall suppose that, on the left, the speed is 1 and on the right it is $1/n$, where n is the index of refraction. The light travels more slowly in the glass by a factor n .

27-1 Introduction

27-2 The focal length of a spherical surface

27-3 The focal length of a lens

27-4 Magnification

27-5 Compound lenses

27-6 Aberrations

27-7 Resolving power

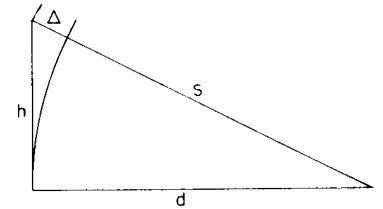


Figure 27-1

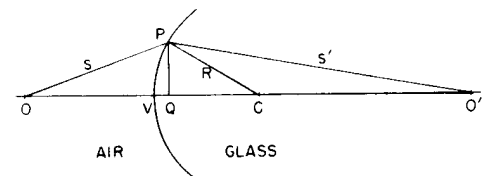


Fig. 27-2. Focusing by a single refracting surface.

Now suppose that we have a point at O , at a distance s from the front surface of the glass, and another point O' at a distance s' inside the glass, and we desire to arrange the curved surface in such a manner that every ray from O which hits the surface, at any point P , will be bent so as to proceed toward the point O' . For that to be true, we have to shape the surface in such a way that the time it takes for the light to go from O to P , that is, the distance OP divided by the speed of light (the speed here is unity), plus $n \cdot O'P$, which is the time it takes to go from P to O' , is equal to a constant independent of the point P . This condition supplies us with an equation for determining the surface. The answer is that the surface is a very complicated fourth-degree curve, and the student may entertain himself by trying to calculate it by analytic geometry. It is simpler to try a special case that corresponds to $s \rightarrow \infty$, because then the curve is a second-degree curve and is more recognizable. It is interesting to compare this curve with the parabolic curve we found for a focusing mirror when the light is coming from infinity.

So the proper surface cannot easily be made—to focus the light from one point to another requires a rather complicated surface. It turns out in practice that we do not try to make such complicated surfaces ordinarily, but instead we make a compromise. Instead of trying to get *all* the rays to come to a focus, we arrange it so that only the rays fairly close to the axis OO' come to a focus. The farther ones may deviate if they want to, unfortunately, because the ideal surface is complicated, and we use instead a spherical surface with the right curvature at the axis. It is so much easier to fabricate a sphere than other surfaces that it is profitable for us to find out what happens to rays striking a spherical surface, supposing that only the rays near the axis are going to be focused perfectly. Those rays which are near the axis are sometimes called *paraxial rays*, and what we are analyzing are the conditions for the focusing of paraxial rays. We shall discuss later the errors that are introduced by the fact that all rays are not always close to the axis.

Thus, supposing P is close to the axis, we drop a perpendicular PQ such that the height PQ is h . For a moment, we imagine that the surface is a plane passing through P . In that case, the time needed to go from O to P would exceed the time from O to Q , and also, the time from P to O' would exceed the time from Q to O' . But that is why the glass must be curved, because the total excess time must be compensated by the delay in passing from V to Q ! Now the *excess* time along route OP is $h^2/2s$, and the excess time on the other route is $nh^2/2s'$. This excess time, which must be matched by the delay in going along VQ , differs from what it would have been in a vacuum, because there is a medium present. In other words, the time to go from V to Q is not as if it were straight in the air, but it is slower by the factor n , so that the excess delay in this distance is then $(n - 1)VQ$. And now, how large is VQ ? If the point C is the center of the sphere and if its radius is R , we see by the same formula that the distance VQ is equal to $h^2/2R$. Therefore we discover that the law that connects the distances s and s' , and that gives us the radius of curvature R of the surface that we need, is

$$(h^2/2s) + (nh^2/2s') = (n - 1)h^2/2R \quad (27.2)$$

or

$$(1/s) + (n/s') = (n - 1)/R. \quad (27.3)$$

If we have a position O and another position O' , and want to focus light from O to O' , then we can calculate the required radius of curvature R of the surface by this formula.

Now it turns out, interestingly, that the same lens, with the same curvature R , will focus for other distances, namely, for any pair of distances such that the sum of the two reciprocals, one multiplied by n , is a constant. Thus a given lens will (so long as we limit ourselves to paraxial rays) focus not only from O to O' , but between an infinite number of other pairs of points, so long as those pairs of points bear the relationship that $1/s + n/s'$ is a constant, characteristic of the lens.

In particular, an interesting case is that in which $s \rightarrow \infty$. We can see from the formula that as one s increases, the other decreases. In other words, if point O

goes out, point O' comes in, and vice versa. As point O goes toward infinity, point O' keeps moving in until it reaches a certain distance, called the *focal length* f' , inside the material. If parallel rays come in, they will meet the axis at a distance f' . Likewise, we could imagine it the other way. (Remember the reciprocity rule: if light will go from O to O' , of course it will also go from O' to O .) Therefore, if we had a light source inside the glass, we might want to know where the focus is. In particular, if the light in the glass were at infinity (same problem) where would it come to a focus outside? This distance is called f . Of course, we can also put it the other way. If we had a light source at f and the light went through the surface, then it would go out as a parallel beam. We can easily find out what f and f' are:

$$n/f' = (n - 1)/R \quad \text{or} \quad f' = Rn/(n - 1), \quad (27.4)$$

$$1/f = (n - 1)/R \quad \text{or} \quad f = R/(n - 1). \quad (27.5)$$

We see an interesting thing: if we divide each focal length by the corresponding index of refraction we get the same result! This theorem, in fact, is general. It is true of any system of lenses, no matter how complicated, so it is interesting to remember. We did not prove here that it is general—we merely noted it for a single surface, but it happens to be true in general that the two focal lengths of a system are related in this way. Sometimes Eq. (27.3) is written in the form

$$1/s + n/s' = 1/f. \quad (27.6)$$

This is more useful than (27.3) because we can measure f more easily than we can measure the curvature and index of refraction of the lens: if we are not interested in designing a lens or in knowing how it got that way, but simply lift it off a shelf, the interesting quantity is f , not the n and the 1 and the R !

Now an interesting situation occurs if s becomes less than f . What happens then? If $s < f$, then $(1/s) > (1/f)$, and therefore s' is negative; our equation says that the light will focus only with a negative value of s' , whatever that means! It does mean something very interesting and very definite. It is still a useful formula, in other words, even when the numbers are negative. What it means is shown in Fig. 27-3. If we draw the rays which are diverging from O , they will be bent, it is true, at the surface, and they will not come to a focus, because O is so close in that they are “beyond parallel.” However, they diverge as if they had come from a point O' outside the glass. This is an apparent image, sometimes called a *virtual image*. The image O' in Fig. 27-2 is called a *real image*. If the light really comes to a point, it is a real image. But if the light *appears* to be coming from a point, a fictitious point different from the original point, it is a virtual image. So when s' comes out negative, it means that O' is on the other side of the surface, and everything is all right.

Now consider the interesting case where R is equal to infinity; then we have $(1/s) + (n/s') = 0$. In other words, $s' = -ns$, which means that if we look from a dense medium into a rare medium and see a point in the rare medium, it appears to be deeper by a factor n . Likewise, we can use the same equation backwards, so that if we look into a plane surface at an object that is at a certain distance inside the dense medium, it will appear as though the light is coming from not as far back (Fig. 27-4). When we look at the bottom of a swimming pool from above, it does not look as deep as it really is, by a factor $3/4$, which is the reciprocal of the index of refraction of water.

We could go on, of course, to discuss the spherical mirror. But if one appreciates the ideas involved, he should be able to work it out for himself. Therefore we leave it to the student to work out the formula for the spherical mirror, but we mention that it is well to adopt certain conventions concerning the distances involved:

- (1) The object distance s is positive if the point O is to the left of the surface.
- (2) The image distance s' is positive if the point O' is to the right of the surface.
- (3) The radius of curvature of the surface is positive if the center is to the right of the surface.

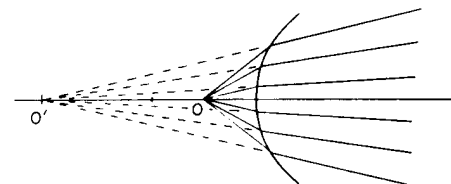


Fig. 27-3. A virtual image.

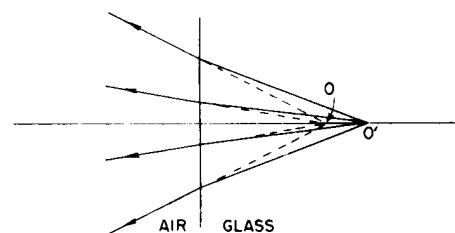


Fig. 27-4. A plane surface re-images the light from O' to O .

In Fig. 27-2, for example, s , s' , and R are all positive; in Fig. 27-3, s and R are positive, but s' is negative. If we had used a concave surface, our formula (27.3) would still give the correct result if we merely make R a negative quantity.

In working out the corresponding formula for a mirror, using the above conventions, you will find that if you put $n = -1$ throughout the formula (27.3) (as though the material behind the mirror had an index -1), the right formula for a mirror results!

Although the derivation of formula (27.3) is simple and elegant, using least time, one can of course work out the same formula using Snell's law, remembering that the angles are so small that the sines of angles can be replaced by the angles themselves.

27-3 The focal length of a lens

Now we go on to consider another situation, a very practical one. Most of the lenses that we use have two surfaces, not just one. How does this affect matters? Suppose that we have two surfaces of different curvature, with glass filling the space between them (Fig. 27-5). We want to study the problem of focusing from a point O to an alternate point O' . How can we do that? The answer is this: First, use formula (27.3) for the first surface, forgetting about the second surface. This will tell us that the light which was diverging from O will appear to be converging or diverging, depending on the sign, from some other point, say O' . Now we consider a new problem. We have a different surface, between glass and air, in which rays are converging toward a certain point O' . Where will they actually converge? We use the same formula again! We find that they converge at O'' . Thus, if necessary, we can go through 75 surfaces by just using the same formula in succession, from one to the next!

There are some rather high-class formulas that would save us considerable energy in the few times in our lives that we might have to chase the light through five surfaces, but it is easier just to chase it through five surfaces when the problem arises than it is to memorize a lot of formulas, because it may be we will never have to chase it through any surfaces at all!

In any case, the principle is that when we go through one surface we find a new position, a new focal point, and then take that point as the starting point for the next surface, and so on. In order to actually do this, since on the second surface we are going from n to 1 rather than from 1 to n , and since in many systems there is more than one kind of glass, so that there are indices n_1, n_2, \dots , we really need a generalization of formula (27.3) for a case where there are two different indices, n_1 and n_2 , rather than only n . Then it is not difficult to prove that the general form of (27.3) is

$$(n_1/s) + (n_2/s') = (n_2 - n_1)/R. \quad (27.7)$$

Particularly simple is the special case in which the two surfaces are very close together—so close that we may ignore small errors due to the thickness. If we draw the lens as shown in Fig. 27-6, we may ask this question: How must the lens be built so as to focus light from O to O' ? Suppose the light comes exactly to the edge of the lens, at point P . Then the excess time in going from O to O' is $(n_1 h^2/2s) + (n_1 h^2/2s')$, ignoring for a moment the presence of the thickness T of glass of index n_2 . Now, to make the time for the direct path equal to that for the path OPO' , we have to use a piece of glass whose thickness T at the center is such that the delay introduced in going through this thickness is enough to compensate for the excess time above. Therefore the thickness of the lens at the center must be given by the relationship

$$(n_1 h^2/2s) + (n_1 h^2/2s') = (n_2 - n_1)T. \quad (27.8)$$

We can also express T in terms of the radii R_1 and R_2 of the two surfaces. Paying attention to our convention (3), we thus find, for $R_1 < R_2$ (a convex lens),

$$T = (h^2/2R_1) - (h^2/2R_2). \quad (27.9)$$

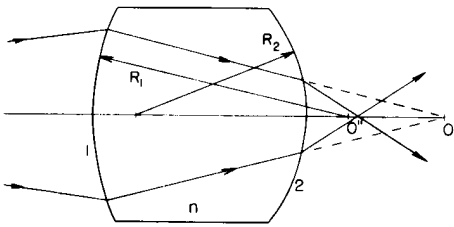


Fig. 27-5. Image formation by a two-surface lens.

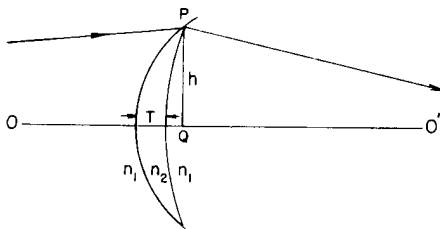


Fig. 27-6. A thin lens with two positive radii.

Therefore, we finally get

$$(n_1/s) + (n_1/s') = (n_2 - n_1)(1/R_1 - 1/R_2). \quad (27.10)$$

Now we note again that if one of the points is at infinity, the other will be at a point which we will call the focal length f . The focal length f is given by

$$1/f = (n - 1)(1/R_1 - 1/R_2), \quad (27.11)$$

where $n = n_2/n_1$.

Now, if we take the opposite case, where s goes to infinity, we see that s' is at the focal length f' . This time the focal lengths are equal. (This is another special case of the general rule that the ratio of the two focal lengths is the ratio of the indices of refraction in the two media in which the rays focus. In this particular optical system, the initial and final indices are the same, so the two focal lengths are equal.)

Forgetting for a moment about the actual formula for the focal length, if we bought a lens that somebody designed with certain radii of curvature and a certain index, we could measure the focal length, say, by seeing where a point at infinity focuses. Once we had the focal length, it would be better to write our equation in terms of the focal length directly, and the formula then is

$$(1/s) + (1/s') = 1/f. \quad (27.12)$$

Now let us see how the formula works and what it implies in different circumstances. First, it implies that if s or s' is infinite the other one is f . That means that parallel light focuses at a distance f , and this in effect *defines* f . Another interesting thing it says is that both points move in the same direction. If one moves to the right, the other does also. Another thing it says is that s and s' are equal if they are both equal to $2f$. In other words, if we want a symmetrical situation, we find that they will both focus at a distance $2f$.

27-4 Magnification

So far we have discussed the focusing action only for points on the axis. Now let us discuss also the imaging of objects not exactly on the axis, but a little bit off, so that we can understand the properties of *magnification*. When we set up a lens so as to focus light from a small filament onto a "point" on a screen, we notice that on the screen we get a "picture" of the same filament, except of a larger or smaller size than the true filament. This must mean that the light comes to a focus from *each point* of the filament. In order to understand this a little better, let us analyze the thin lens system shown schematically in Fig. 27-7. We know the following facts:

- (1) Any ray that comes in parallel on one side proceeds toward a certain particular point called the focus on the other side, at a distance f from the lens.
- (2) Any ray that arrives at the lens from the focus on one side comes out parallel to the axis on the other side.

This is all we need to establish formula (27.12) by geometry, as follows: Suppose we have an object at some distance x from the focus; let the height of the object be y . Then we know that one of the rays, namely PQ , will be bent so as to pass through the focus R on the other side. Now if the lens will focus point P at all, we can find out where if we find out where just one other ray goes, because the new focus will be where the two intersect again. We need only use our ingenuity to find the exact direction of *one* other ray. But we remember that a parallel ray goes through the focus and *vice versa*: a ray which goes through the focus will come out parallel! So we draw ray PT through U . (It is true that the actual rays which are doing the focusing may be much more limited than the two we have drawn, but they are harder to figure, so we make believe that we can make this ray.) Since it would come out parallel, we draw TS parallel to XW . The intersection S is the point we need. This will determine the correct place and the correct height. Let

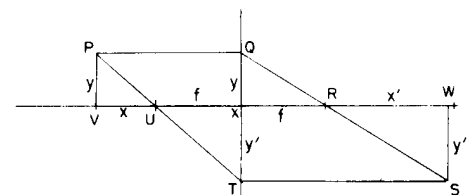


Fig. 27-7. The geometry of imaging by a thin lens.

us call the height y' and the distance from the focus, x' . Now we may derive a lens formula. Using the similar triangles PVU and TXU , we find

$$\frac{y'}{f} = \frac{y}{x}. \quad (27.13)$$

Similarly, from triangles SWR and QXR , we get

$$\frac{y'}{x'} = \frac{y}{f}. \quad (27.14)$$

Solving each for y'/y , we find that

$$\frac{y'}{y} = \frac{x'}{f} = \frac{f}{x}. \quad (27.15)$$

Equation (27.15) is the famous lens formula; in it is everything we need to know about lenses: It tells us the magnification, y'/y , in terms of the distances and the focal lengths. It also connects the two distances x and x' with f :

$$xx' = f^2, \quad (27.16)$$

which is a much neater form to work with than Eq. (27.12). We leave it to the student to demonstrate that if we call $s = x + f$ and $s' = x' + f$, Eq. (27.12) is the same as Eq. (27.16).

27-5 Compound lenses

Without actually deriving it, we shall briefly describe the general result when we have a number of lenses. If we have a system of several lenses, how can we possibly analyze it? That is easy. We start with some object and calculate where its image is for the first lens, using formula (27.16) or (27.12) or any other equivalent formula, or by drawing diagrams. So we find an image. Then we treat this image as the source for the next lens, and use the second lens with whatever its focal length is to again find an image. We simply chase the thing through the succession of lenses. That is all there is to it. It involves nothing new in principle, so we shall not go into it. However, there is a very interesting net result of the effects of any sequence of lenses on light that starts and ends up in the same medium, say air. Any optical instrument—a telescope or a microscope with any number of lenses and mirrors—has the following property: There exist two planes, called the *principal planes* of the system (these planes are often fairly close to the first surface of the first lens and the last surface of the last lens), which have the following properties: (1) If light comes into the system parallel from the first side, it comes out at a certain focus, at a distance from the *second* principal plane equal to the focal length, just as though the system were a thin lens situated at this plane. (2) If parallel light comes in the other way, it comes to a focus at the same distance f from the *first* principal plane, again as if a thin lens were situated there. (See Fig. 27-8.)

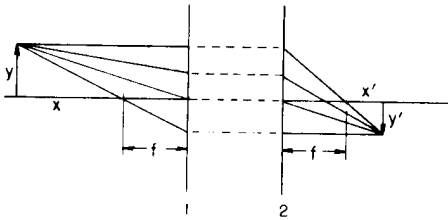


Fig. 27-8. Illustration of the principal planes of an optical system.

Of course, if we measure the distances x and x' , and y and y' as before, the formula (27.16) that we have written for the thin lens is absolutely general, provided that we measure the focal length from the principal planes and not from the center of the lens. It so happens that for a thin lens the principal planes are coincident. It is just as though we could take a thin lens, slice it down the middle, and separate it, and not notice that it was separated. Every ray that comes in pops out immediately on the other side of the second plane from the same point as it went into the first plane! The principal planes and the focal length may be found either by experiment or by calculation, and then the whole set of properties of the optical system are described. It is very interesting that the result is not complicated when we are all finished with such a big, complicated optical system.

27-6 Aberrations

Before we get too excited about how marvelous lenses are, we must hasten to add that there are also serious limitations, because of the fact that we have limited ourselves, strictly speaking, to paraxial rays, the rays near the axis. A real lens having a finite size will, in general, exhibit *aberrations*. For example, a ray that is on the axis, of course, goes through the focus; a ray that is very close to the axis will still come to the focus very well. But as we go farther out, the ray begins to deviate from the focus, perhaps by falling short, and a ray striking near the top edge comes down and misses the focus by quite a wide margin. So, instead of getting a point image, we get a smear. This effect is called *spherical aberration*, because it is a property of the spherical surfaces we use in place of the right shape. This could be remedied, for any specific object distance, by re-forming the shape of the lens surface, or perhaps by using several lenses arranged so that the aberrations of the individual lenses tend to cancel each other.

Lenses have another fault: light of different colors has different speeds, or different indices of refraction, in the glass, and therefore the focal length of a given lens is different for different colors. So if we image a white spot, the image will have colors, because when we focus for the red, the blue is out of focus, or vice versa. This property is called *chromatic aberration*.

There are still other faults. If the object is off the axis, then the focus really isn't perfect any more, when it gets far enough off the axis. The easiest way to verify this is to focus a lens and then tilt it so that the rays are coming in at a large angle from the axis. Then the image that is formed will usually be quite crude, and there may be no place where it focuses well. There are thus several kinds of errors in lenses that the optical designer tries to remedy by using many lenses to compensate each other's errors.

How careful do we have to be to eliminate aberrations? Is it possible to make an absolutely perfect optical system? Suppose we had built an optical system that is supposed to bring light exactly to a point. Now, arguing from the point of view of least time, can we find a condition on how perfect the system has to be? The system will have some kind of an entrance opening for the light. If we take the farthest ray from the axis that can come to the focus (if the system is perfect, of course), the times for all rays are exactly equal. But nothing is perfect, so the question is, how wrong can the time be for this ray and not be worth correcting any further? That depends on how perfect we want to make the image. But suppose we want to make the image as perfect as it possibly can be made. Then, of course, our impression is that we have to arrange that every ray takes as nearly the same time as possible. But it turns out that this is not true, that beyond a certain point we are trying to do something that is too fine, because the theory of geometrical optics does not work!

Remember that the principle of least time is not an accurate formulation, unlike the principle of conservation of energy or the principle of conservation of momentum. The principle of least time is only an *approximation*, and it is interesting to know how much error can be allowed and still not make any apparent difference. The answer is that if we have arranged that between the maximal ray—the worst ray, the ray that is farthest out—and the central ray, the difference in time is less than about the period that corresponds to one oscillation of the light, then there is no use improving it any further. Light is an oscillatory thing with a definite frequency that is related to the wavelength, and if we have arranged that the time difference for different rays is less than about a period, there is no use going any further.

27-7 Resolving power

Another interesting question—a very important technical question with all optical instruments—is how much *resolving power* they have. If we build a microscope, we want to see the objects that we are looking at. That means, for instance, that if we are looking at a bacterium with a spot on each end, we want to *see* that

there are two dots when we magnify them. One might think that all we have to do is to get enough magnification—we can always add another lens, and we can always magnify again and again, and with the cleverness of designers, all the spherical aberrations and chromatic aberrations can be cancelled out, and there is no reason why we cannot keep on magnifying the image. So the limitations of a microscope are not that it is impossible to build a lens that magnifies more than 2000 diameters. We can build a system of lenses that magnifies 10,000 diameters, but we *still* could not see two points that are too close together because of the limitations of geometrical optics, because of the fact that least time is not precise.

To discover the rule that determines how far apart two points have to be so that at the image they appear as separate points can be stated in a very beautiful way associated with the time it takes for different rays. Suppose that we disregard the aberrations now, and imagine that for a particular point P (Fig. 27-9) all the rays from object to image T take exactly the same time. (It is not true, because it is not a perfect system, but that is another problem.) Now take another nearby point, P' , and ask whether its image will be distinct from T . In other words, whether we can make out the difference between them. Of course, according to geometrical optics, there should be two point images, but what we see may be rather smeared and we may not be able to make out that there are two points. The condition that the second point is focused in a distinctly different place from the first one is that the two times for the extreme rays $P'ST$ and $P'RT$ on each side of the big opening of the lenses to go from one end to the other, must *not* be equal to the two possible object points to a given image point. Why? Because, if the times were equal, of course both would *focus* at the same point. So the times are not going to be equal. But by how much do they have to differ so that we can say that both do *not* come to a common focus, so that we can distinguish the two image points? The general rule for the resolution of any optical instrument is this: two different point sources can be resolved only if one source is focused at such a point that the times for the maximal rays from the other source to reach that point, as compared with its own true image point, differ by more than one period. It is necessary that the difference in time between the top ray and the bottom ray to the *wrong* focus shall exceed a certain amount, namely, approximately the period of oscillation of the light:

$$t_2 - t_1 > 1/\nu, \quad (27.17)$$

where ν is the frequency of the light (number of oscillations per second; also speed divided by wavelength). If the distance of separation of the two points is called D , and if the opening angle of the lens is called θ , then one can demonstrate that (27.17) is exactly equivalent to the statement that D must exceed $\lambda/n \sin \theta$, where n is the index of refraction at P and λ is the wavelength. The smallest things that we can see are therefore approximately the wavelength of light. A corresponding formula exists for telescopes, which tells us the smallest difference in angle between two stars that can just be distinguished.*

* The angle is about λ/D , where D is the lens diameter. Can you see why?

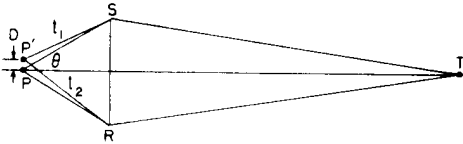


Fig. 27-9. The resolving power of an optical system.