Optics and Geometry
with Applications to Photography

Tom Davis
tomrdavis@earthlink.net
http://www.geometer.org/mathcircles
November 15, 2004

1 Useful approximations

This paper can be classified as applied mathematics since every equation we'll discuss has applications to useful physical objects, such as telescopes, cameras, prisms, eyeglasses and so on. As is the case with almost any application of mathematics to the real world, we will be making certain approximations to be able to solve the equations. For that reason we realize that our solutions will ususally be slightly in error, but that the errors can be made quite small. The main job of an engineer who is designing a real optical system is to create a physical situation that reduces the size of these errors to an acceptable level.

As an example, we'll talk about the "thin lens" approximation, where we will assume that the lens is infinitely thin. Obviously no real lens is infinitely thin, but as long as the thickness of the lens is small enough compared to the distances to the objects being viewed and imaged, this approximation may be good enough. We'll make further assumptions about the nature of light, about the perfection of our lens material, and so on. When it's important, we'll point out places where these assumptions may cause trouble in a real physical system.

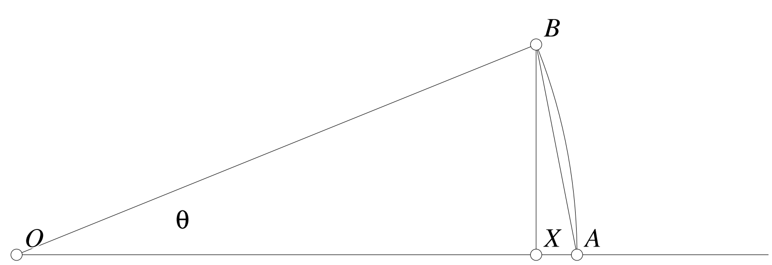

Figure 1: Trigonometric Approximations

We will make a couple of mathematical approximations from time to time as well. The most important ones concern the calculations of angles and of the trigonometic functions of those angles. We will always measure angles in radians rather than degrees.

Consider the angle $\angle AOB$ in Figure 1. If the length $OA = OB = 1$, then the length of the arc AB is the measure of the angle θ measured in radians. (If the length of OA happens not to be 1, then the measure of θ in radians is the length of the arc AB divided by the length of OA.) Similarly, if $OA = OB = 1$, then $\sin\theta = BX/OB = BX$, since $OB = 1$. It should be obvious from the figure, and it can be proven mathematically, that as the size of the angle θ gets smaller and smaller, the length BX is closer and closer to the length of the arc AB. In fact, we can write this mathmatically as:

$$\lim_{\theta\to 0}\frac{BX}{AB} = \lim_{\theta\to 0}\frac{\sin\theta}{\theta} = 1.$$

If we multiply through by θ, then we can see that for very small values of θ, $\sin\theta \approx \theta$, and for that reason, when we do deal with very small angles, we will often replace instances of $\sin\theta$ by θ in the calculations.

It should also be obvious from the figure (and can, again, be rigorously proven mathematically) that the lengh of the line segment AB and the length of the arc AB are also very similar. Thus if $OA = OB = 1$, we can approximate the angle θ as the length of the line segment AB, or, if OA happens not to be 1, we can approximate θ by AB/OB.

We will use the two approximations above repeatedly in what follows.

2 Images and Virtual Images

This paper will consider properties of lenses, where the word ”lens” is used in a general sense. We are going to examine how light is bent as it passes through different materials. Obviously, the most common example is the bending of light through the lenses of cameras, binoculars or microscopes that form images on film or on the retina of your eye.

The most important feature of lenses is that they are able to form ”images” of objects in the world. As an example, consider what happens when you use a camera to take a photo. Let’s consider one of the most boring possible photos: it is pitch-black in a room except that there is a point-sized light source and the camera is pointed at a that source. Your goal is to produce an image on the film that is pitch-black except for a single dot where the light from the point-source arrives on your film.

The tiny light sends off photons in all directions, so first imagine what would happen if there were no lens, but rather that the film were pointed at the light and exposed. Photons would arrive more or less uniformly across the surface and the whole area of the film would be uniformly exposed. With the lens, however, some collection of the photons will happen to hit the circular lens on your camera, and what we would like the lens to do is to bend all of those photons back so that they hit a single point on the film, as shown in Figure 2. The point O is the location of the “object”–in this case, the light source. The point I is the “image”–the place where the light from the object is focused. If the film happens to lie at point I, the point source will produce a point image, but if the film happens to be a little in front of or behind point I, the light from

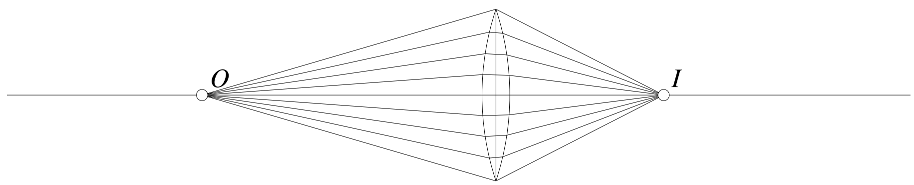

Figure 2: Object and Image

the source will produce a circle of light on the film–in other words, a blurred image. If the film is close to I, the blurring will be small, and if it's far away, there will be a lot of blur. When you focus a camera, you are basically changing the distance between the lens and the film so that the object you are interested in makes point sources of the object focus to point images on the film.

Figure 2 shows the effects of the lens on a single source of light; in reality, a perfect lens will focus an entire plane of point sources onto the film plane at the same time, so that an entire plane in the real world is in perfect focus on the film. For real lenses, however, nothing is perfect, so the world-plane that's in perfect focus will not, in fact, be a perfect plane, but will be some curved surface in 3-space. And the focus will never be exact, either, but with expensive lenses, the errors can be made to be quite small.

2.1 Virtual Images

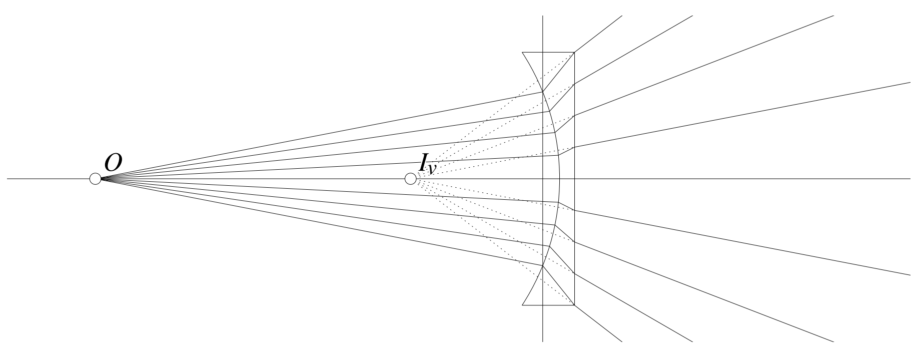

Figure 3: Object and Virtual Image

The lens illustrated in Figure 2 bends the light from the source so that it forms a real image where the photons diverging from the source meet again, hopefully on the film or on your retina. Other lenses, or locations of the object, like the one illustrated in

Figure 3, may cause the photons from the source to diverge instead of converge. In this case, the diverging photons will diverge forever, and will never meet, but if such a lens were perfect, those diverging lines would appear to be diverging from a single point, labeled I_v in the image. This is not a true, or real image, but the point is called a “virtual image”.

A lens that forms a virtual image may at first seem useless, but imagine that it is a part of a series of lenses. The next lens along would see the light from the source at O not as if it were coming from O, but as if it were diverging from the point I_v. The dotted lines in the figure simply indicate the direction of the diverging lines; the solid lines represent the true paths of the photons.

2.2 Cones of Light

If you examine either of the figures 2 or 3 you see a pattern: light is emitted in all directions from the object, but the part of interest is the collection of photons that strike the circular lens. This light fills a cone-shaped volume of space.

When the light in the cone strikes a lens, the light coming out may have changed direction somewhat but it now fits inside another cone-shaped piece of space. The cone may point in the opposite direction, as in figure 2 or it may point in the same direction but with a different angle, as in figure 3. Thus in a system of lenses, every time the light arrives at the next lens, it is in the form of a cone, either getting larger or smaller (or, if the beams happen to be parallel, it’s like an infinite cone or cylinder).

What every perfect lens does is to take light coming from one side in a cone-shaped volume and emit it on the other side in a different cone-shaped volume. If we know how to calculate the new cone shape from the old one, we can trace the light through lens after lens to find the resultant cone shape after passing through any number of lenses.

In fact, it’s not lenses that change the cone shapes; it’s lens surfaces, so to figure out how a simple two-sided lens works, we’ll merely need to figure out how the cone shape is changed by each of the two surfaces.

Since this is a short paper, we will only do one sample calculation for each example, but the formulas we derive will be very general in the following sense: We’ll assume that the object is on one side of the lens and the image is on the other and we’ll call those distances to the lens o and i for the object and image distances. If the image happens to lie on the same side as the object (as in the case of the virtual image in figure 3) the same equations will work, but they will give a negative value for i. Similarly, when we work out just how a lens ground with a certain radius r affects the light, we’ll obtain an equation, and if the lens surface happens to be concave instead of convex, the same equation will work, but you’ll need to put in a negative value for r to obtain the correct result.

2.3 Depth of Field

A camera with a perfect lens will focus every point on some focus plane in front of it to a point on the film plane that lies within the camera. *Only* points on the focus plane will be focused perfectly. Points in the world in front of or behind the plane of perfect focus will not form point images on the film, but rather will form small circles.

The best way to visualize this is to think of the photons that come through the lens and are bent back to a point as lying inside a cone of light. If the tip of the cone is at the film plane, we get perfect focus, but if the tip is a little in front of or behind the film plane, then photons from the object lie in a circle on the film. The farther the tip of the cone is from the film plane, the larger the circle. In a camera, this circle is sometimes called the “circle of confusion”.

Human eyes are not perfect either, and if the circle of confusion is small enough, we will still consider the focus to be perfect, so objects near the focus plane also form images that are almost in focus. Camera lens manufacturers define a concept called ”depth of field” which indicates the error allowable from the focal plane that still produces images that are good enough to seem perfect to humans.

Finally, notice that fat cones of light will allow less error than skinny cones, and the cones of light can be made skinnier by restricting the amount of the lens that is used to be a small part near the center. The “diaphragm” of the camera performs this function, so by having a smaller “aperture” through which the light can travel, a photographer can obtain a larger depth of field, which may or may not be good, depending on the goal of the photo.

Pinhole cameras, in fact, can be built that have no lens at all, but just use a tiny hole in place of a lens. If the hole could be made arbitrarily small, a pinhole camera could form arbitrarily good images. The problem is that light only approximately follows straight lines and is bent slightly at the edges of the diaphragm or hole in a pinhole camera due to a property called “diffraction”, so there is a limit to how good even a pinhole camera can be.

3 Index of Refraction and Snell’s Law

Photons of light travelling through a vacuum move in a straight line at the speed of light, which is about $300,000$ kilometers/second which physicists often denote by the letter c. Each photon can be thought of as a wave packet with some wavelength.

Light travelling through any other material, like water, plastic or glass, does not move with velocity c, but somewhat slower, and the ratio of c to the true speed of light in the material is called the index of refraction of that material. Here are the indices of refraction for a few common materials:

Material	Refractive index
Vacuum	1.0
Air	1.0003
Water	1.33
Glass (zinc crown)	1.51
Glass (heaviest flint)	1.89
Diamond	2.42

This means, for example, that light passing through a vacuum travels 1.33 times as fast as it does through water. There are many different types of glass which allows lens makers to construct lenses with very different properties. As we shall see shortly, the index of refraction is closely related to how much light bends when it enters or leaves a material, and the extremely high index of refraction for diamond offers some clue as to why those precious stones are so interesting to look at–it is *very* different from glass and produces very different visual effects.

3.1 Snell's Law

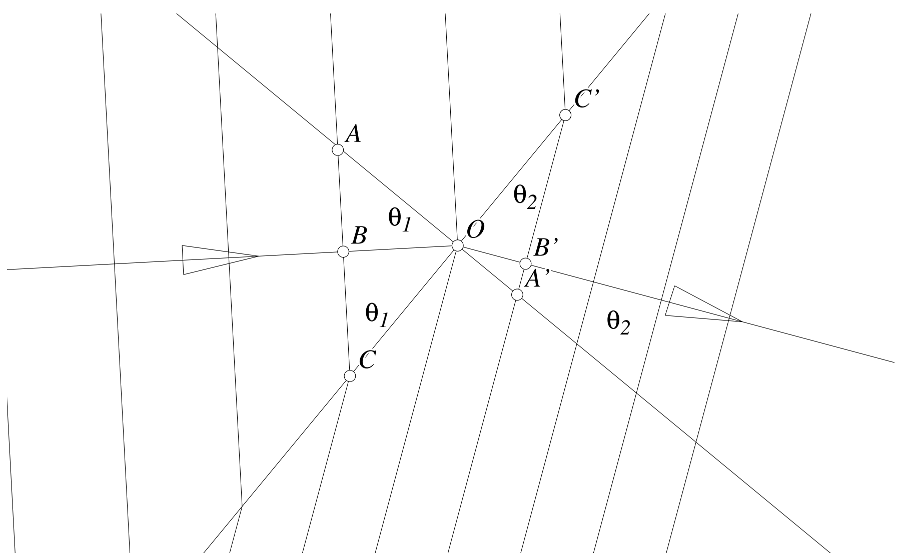

Figure 4: Snell's Law

Simply by knowing the relative speeds of light through the two materials and knowing that light is a wave phenomenon, we can calculate the amount by which it will be bent as it moves across a surface from one material to another. Consider a straight line interface between two different materials as illustrated in figure 4.

The line CC' is the interface between the materials. The two lines with arrows on them indicate that light is coming in from the left and leaving on the right at a slightly different angle. The line AA' is is called the surface normal and is perpendicular to the surface-surface interface and the regularly-spaced lines perpendicular to the lines OB and OB' indicate the crests of the waves of the moving light. The length $\lambda_1 = OB$, for example, is the wavelength of the light arriving from the left, and the length $\lambda_2 = OB'$ is the wavelength of the light leaving the interface to the right.

The angles of incidence and refraction, $\theta_1 = \angle BOA$ and $\theta_2 = \angle B'OA'$, respectively, are measured from the surface normal AA'.

It is easy to see that $\angle OCB = \theta_1$ and that $\angle OC'B' = \theta_2$. Since the waves are uniformly spaced, $OC = OC' = d$, and simple trigonometry gives us: $\sin\theta_1 = \lambda_1/d$ and $\sin\theta_2 = \lambda_2/d$. Combining these last two yields:

$$\frac{\sin\theta_1}{\lambda_1} = \frac{\sin\theta_2}{\lambda_2}. \tag{1}$$

Since it requires the same amount of time for each wave crest to pass in either material, $\lambda_1 n_1 = \lambda_2 n_2$. We can use this in combination with equation 1 to obtain Snell's Law:

$$n_1 \sin\theta_1 = n_2 \sin\theta_2. \tag{2}$$

3.2 Total Internal Reflection

Imagine that we have the same situation as in the first example in the section 3 where two materials meet along a straight line. If the light is coming from the side with a larger index of refraction, we recall that as it enters the less dense medium, the light speeds up and is bent away from the normal vector by a certain amount, depending on the relative sizes of the indices of refraction. For some incidence angle, this angle of refraction will equal 90° which means that the refracted light does not even enter the material and if light strikes at this angle or greater, no light enters the other material. This angle is called the critical angle.

Energy doesn't just disappear–it has to go somewhere–and what happens is that instead of passing through the surface, all of the light is simply reflected from the surface. This doesn't happen all of a sudden, however. What really occurs is that as the incidence angle gets shallower and shallower, more and more light is reflected.

If you shined a laser beam at a surface, for example, and started having it point at 90° to the interface, but then gradually tilted it, you would find that the more it was tilted, the closer the refracted beam would be to being parallel to the surface interface, and at the same time, the more light would be reflected rather than passing through. When the angle finally reached the critical angle, *all* of the laser light would be reflected.

At what angle does this occur? To find out, all we need to do is to set θ_2 to 90° in Snell's Law (equation 2):

$$\sin\theta_1 = n_2/n_1.$$

Since material 2 has a lower index of refraction, $n_2 < n_1$, so this has a solution. If the light is coming from the side of the interface with a higher index of refraction there can be no solution, since the sine function can never be larger than 1.

3.3 Prisms and Spectra

With real light, of course, the situation is a bit more complicted. It turns out that the speed of light in a material depends on the wavelength (which corresponds to the color) of the light. Higher energy photons travel faster, so the index of refraction for blue light is less than the index of refraction for red light, for example. (Blue light has higher energy photons than red light.)

Because of this variation of index of refraction with wavelength, if a beam of white light (which is a mixture of all wavelengths) hits the surface of a prism, each different color of light is bent by a different amount, and the result is a spectrum.

4 Focus through a Surface

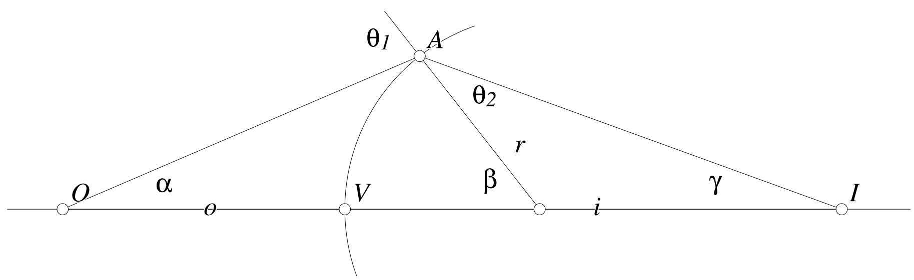

Figure 5: Focus distances

Let us consider the situation illustrated in Figure 5 where light is emitted from the point O (the object) and then passes through a spherical lens surface of radius r and is focused at the point I (the image). The length o is the distance from the object to the surface of the glass at point V and the length i is the distance from the surface of the glass to the point I where the image is focused.

In this example, we assume that everything to the left of the lens surface is one material (say air) and everything to the right is another (say glass), so in this situation, we assume that the image is actually formed inside a giant lens. Later we'll use the results here to show how normal lenses work where the focus is in the air on the other side of the lens. But the situation above is general, and all that we really require is that the index of refraction is the same for all points on the left of the glass surface and is the same on all points to the right. We'll let n_1 be the index of refraction on the left and n_2 be the index of refraction on the right.

Consider any ray of light emitted from point O that strikes the lens surface at point A. Due to the different indices of refraction on both sides of the lens surface, the light will be bent and may focus at a point I on the other side of the surface, forming the various angles and lengths indicated in Figure 5.

Simple Euclidean geometry gives the first two equations below and Snell's Law yields the third:

$$\begin{aligned}\theta_1 &= \alpha + \beta \\ \beta &= \theta_2 + \gamma \\ n_1 \sin\theta_1 &= n_2 \sin\theta_2\end{aligned}$$

Since we're going to assume that point A is quite close to point V so $\sin\theta_1 \approx \theta_1$ and $\sin\theta_2 \approx \theta_2$, we can replace the equation due to Snell's Law with:

$$n_1\theta_1 \approx n_2\theta_2.$$

With a little algebra, we can eliminate θ_1 and θ_2 from the above equations, and obtain:

$$\beta(n_2 - n_1) \approx \alpha n_1 + \gamma n_2. \tag{3}$$

Finally, we can denote by h the length of the segment AV and substitute the small-angle approximations: $\alpha \approx h/o$, $\beta \approx h/r$ and $\gamma \approx h/i$ into equation 3 where all the instances of h cancel to obtain:

$$\frac{n_2 - n_1}{r} \approx \frac{n_1}{o} + \frac{n_2}{i}. \tag{4}$$

Equation 4 relates the distance from the object to the image given the relative indices of refraction in the two media on both sides of the surface. Although we have not worked out all the cases, this equation is true for concave or convex lenses (where a concave lens would be indicated by a negative value of r), and where the distance of the image from the lens surface may be negative (on the same side of the surface as the object, meaning that it is a virtual image).

5 Lens Formulas

In section 2 we noted that as light passes through perfect lenses, the emitted photons always appear to be diverging toward a point or converging to a point. Thus if we consider the result of passing light through a series of lens surfaces, at each surface we can simply calculate how the diverging or converging light will be refocused by the new surface. Thus the calculation above is sufficient, for perfect lenses, to work out the charactertistics of any optical system composed of spherical lenses simply by tracing the light through one surface at a time.

In this section we will do exactly that–we will consider a single lens where light enters from air (or a vacuum), passes through a lens, both of whose surfaces are spherical, and enters air (or a vacuum) on the other side. We will do so by using the result from the previous section as we trace the path of light from a source, through a surface into glass, through the glass, and then out though the opposite lens surface.

In figure 6 we will assume that the object is at the point O and that the light passes through a lens having two surfaces of radius r_1 and r_2 centered at C_1 and C_2, respectively, and that the surfaces are a distance l apart.

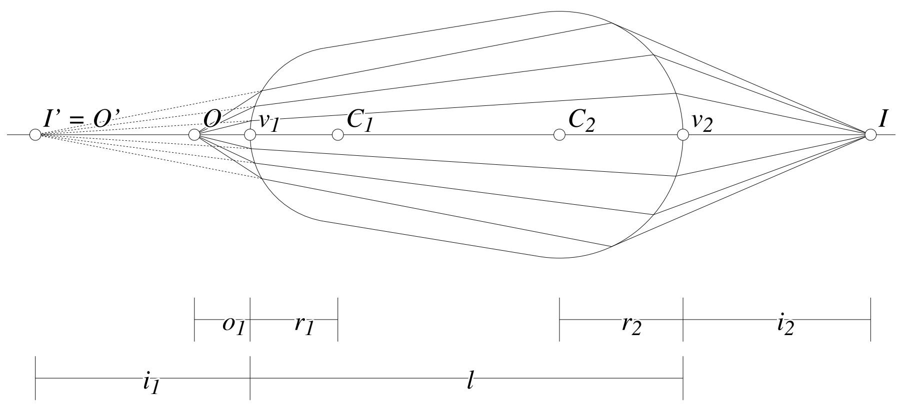

Figure 6: Fat Lens with Two Surfaces

Light from point O passes through the first surface which is a distance o_1 away and the beams are bent to form an image (or in the case of this figure, a virtual image) at the point I' which is at a distance of i_i from the original surface.

Now we just assume that the light is emitted from I' as if it were a new source, and see how that light will be affected as it passes through the other surface. The light will finally converge at a point I, and we want to find a relationship between the various distances and the radii of the two different lens surfaces.

To simplify things slightly, we’ll assume that the lens is glass and has an index of refraction n and that the material outside the lens is a vacuum with index of refraction 1. (Air has an index of refraction of about 1.0003 so if the lens is used in air instead of in a vacuum, the errors will be very slight.)

Using the formula from section 4, and under the assumption that the first image is virtual (and therefore that I' is to the left of v_1) we obtain[1]:

$$\frac{1}{o_1} - \frac{n}{i_i} = \frac{n-1}{r_1}. \tag{5}$$

If we analyze light assumed to be emanating from I' and passing through the second surface (so its distance to that surface is $i_1 + l$) using the same method, we obtain a second equation:

$$\frac{n}{i_1 + l} + \frac{1}{i_2} = \frac{1-n}{r_2}. \tag{6}$$

Combining equations 5 and 6, and letting $l = 0$ (this is the “thin lens” approximation),

[1] It works out fine if the first surface forms a real image to the right of the lens surface, since that will flip the sign on the second fraction below, but will measure i_1 in the opposite direction, effectively negating it.

we obtain:

$$\frac{1}{o_1} + \frac{1}{i_2} = (n-1)\Big(\frac{1}{r_1} - \frac{1}{r_2}\Big). \tag{7}$$

The "focal length" of a lens is defined to be the distance of the focus from the lens assuming that the incoming light is parallel, which is equivalent to stating that the source is infinitely far away. If we set o_1 to ∞ in equation 7 we obtain what is known as the "lens maker's equation", where f is the focal length of the lens:

$$\frac{1}{f} = (n-1)\Big(\frac{1}{r_1} - \frac{1}{r_2}\Big). \tag{8}$$

Thus if you know the two radii of the lens surfaces and the index of refraction of the glass you're using, you can calculate to fairly good accuracy the focal length of the lens produced. In the equation above, we make the assumption that the radii r_1 and r_2 are measured from centers that lie to the right of the surface. If the center lies to the left, the radius will be negative. For the double-convex lens in figure 6, for example, r_1 is positive and r_2 is negative (so in this case both fractions to the right of the equality signs will be positive).

If you are not a lens maker, but just purchase a lens with fixed surface radii and glass with a fixed index of refraction then the term on the right of both equations 7 and 8 is a constant, and equation 8 tells us that the constant is 1 divided by the focal length of the lens. Keeping this in mind, we have derived what is known as the "thin lens equation" where i and o are the distances of object and image on opposite sides of an infinitely thin lens:

$$\frac{1}{o} + \frac{1}{i} = \frac{1}{f}.$$

6 Consequences of the Thin Lens Equation

Equation 5 contains a huge amount of information. Light emitted from a focus point will leave the opposite side of the lens in parallel. Light arriving in parallel will focus at a distance f from the lens on the other side. Light emitted from a point $f/2$ from one side of the lens will focus at the same distance on the other side. As the object gets closer to the lens, the image moves away from the lens until the object is closer than the focal length in which case the lens fails to converge the light and a virtual image is formed on the same side of the lens as the object (and hence has a negative distance to the lens).

An interesting feature of lenses is their magnification. The magnification is simply the ratio of the size of the image to the size of the original object. If the image is twice as big as the object, we say that the magnification is 2. Given a lens with focal length f, how can we calculate the magnification of an object placed at a distance o from that lens?

See figure 7. From the thin lens equation, we know that the only way to obtain a real image (as opposed to a virtual image) is to have the object (and hence image) farther

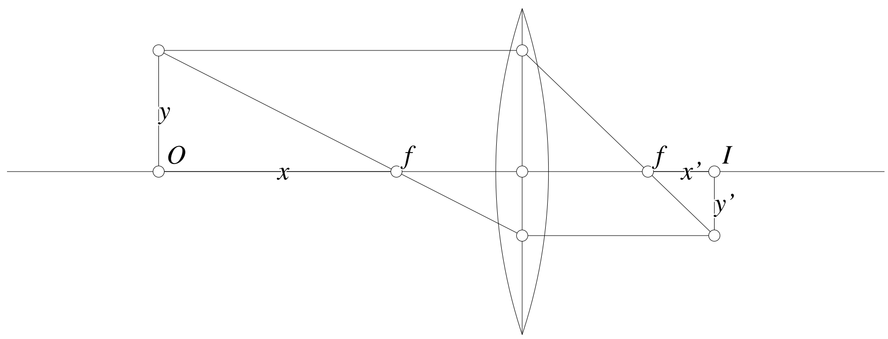

Figure 7: Magnification

away from the lens than the focal length. Suppose the object has height y and is x beyond the focal length. This will form an image of height y' a distance x' beyond the focal length on the other side of the lens.

To measure the magnification of this setup, y'/y, we need to figure out how big y' is, compared to y. If we can trace any two rays from the top of the object to the image, we can determine the height y' of the image. But we know that rays arriving parallel to the axis of the lens must pass through the focus, and rays passing through the focus of the lens must leave parallel to the axis, and those two rays are drawn in figure 7. Simple geometry of similar triangles yields the equations: $y'/x' = y/f$ and $y/x = y'/f$. Thus the magnification y'/y is given by:

$$\frac{y'}{y} = \frac{f}{x} = \frac{x'}{f}.$$

We can cross-multiply the final equality above to obtain $xx' = f^2$.

7 Real Lenses

Real lenses are not infinitely thin, and have plenty of other characteristics that make them not quite obey the equations we have derived in this article. In any real lens or even optical system composed of a series of lenses, there are some deviations from perfect images, and the general term used to describe this sort of error is “lens abberation”. Lens abberations come in various types; for example, “chromatic abberations”, “spherical abberations”, “astigmatism”, “field curvature”, and so on. Lens makers usually try to reduce all of these at the same time by means of clever design.

The first sort of error encountered is that real lenses generally have spherical surfaces which is not the perfect shape, but is a shape that is incredibly simple to produce very accurately. It is possible today to purchase non-spherical lenses for specific applicatons, but they tend to be quite expensive. The reason that spherical lenses are so easy to produce is that if you rub two materials together until they slide smoothly in any direction, the surface between them has to be either a plane or a portion of a sphere. Lenses can be ground by placing an abrasive material between two pieces of glass and rubbing the glass pieces together in every direction.

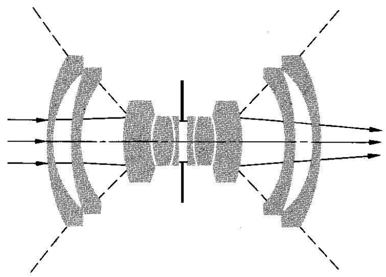

Figure 8: Optical System

What you will find in any real optical system where it is important to reduce the abberrations is that instead of using a single lens, a series of lenses is used, where the series is arranged in such a way that the abberations in one lens tend to be cancelled out by another. The glass in the different lenses can be made of different materials having different indices of refraction, the lenses may be of different shapes, and so on. You have probably seen pictures of such lens systems like the one illustrated in figure 8 in advertisements for camera lenses, et cetera.

The first thing to notice for such real systems is that even if you assumed that the lenses were infinitely thin, the combination certainly is not. Thus your formulas for the thin lens approximation will not work well if you have to work with optical systems instead of with single lenses.

The following approximation *is* approximately true, however. There are two planes in any such system perpendicular to the axis of the lenses called the principal planes, and if you imagine the space between them as being squashed down to zero width, the resulting system would behave like a single thin lens. This is true for *any* system of lenses aligned along their axis.

What this means is that if you're not too worried about distortion, you can attach together any sequence of lenses, and it will behave like a single lens. Photographers take advantage of this in many ways: you can screw a so-called "diopter lens" on the front of your normal lens to turn it into a closeup lens. You can push the lens away from the camera body with a hollow tube (called an "extension tube") that will allow it to focus much closer. You can put in a tele-extender between the camera body and a telephoto lens to multiply its focal length by some given amount (and thus turn a telephoto lens into a super-telephoto lens, for example). You can reverse a lens and shoot photos through it backwards, which typically increases its ability to take closeup shots. And really, you are only limited by your imagination. Of course, only experiment will tell whether the amount of distortion introduced is acceptable. The lens company goes to a great deal of trouble to make sure the lens has low distortion in normal use, but they probably don't do the calculations assuming that the lens is turned around, or sits three inches farther from the camera than the lens mount would indicate.

7.1 Designing Lens Systems

Rather than do sophisticated mathematical calculations that work out the details of the abberations of different lenses, modern lens design is more like trial and error. A set of lenses is used as a starting point, and then a computer simply traces rays of light through the lenses to evaluate the final image. If it's good enough, that's the design. If it's not good enough, a tiny change is made and the whole ray tracing exercise is repeated to see if the new system is better than the old. If not, make a different correction and if so, start from the new arrangement and try to improve it.