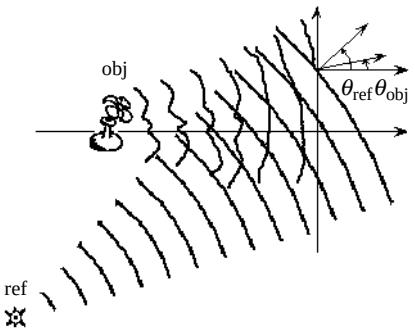
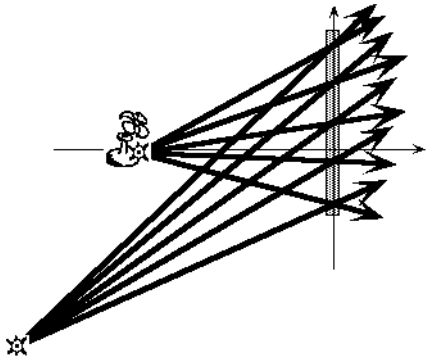
Chapter 8: A Ray-Tracing Analysis of Holography

Introduction

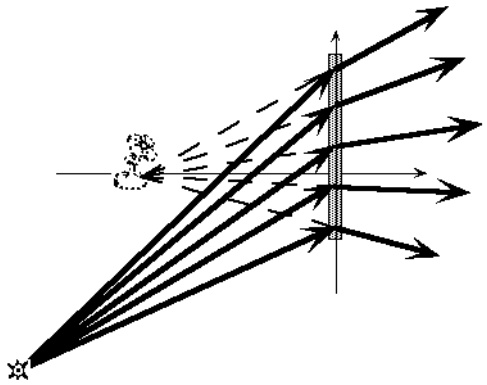
Rather than tackling a generalized and global proof of the wavefront reconstruction properties of holograms, we can instead look at the recording and reconstruction of wavefronts at every small area of a hologram, using just the simple ideas of two-beam interference and diffraction by periodic structures, with the coordinate system continually re-centered on the small region of momentary interest. In this approach we say either that the object beam is locally planar—because its radius of curvature is so much larger than the diameter of the region of interest—or that the object beam can be considered as the sum of a number of plane waves from point sources far from the hologram plane, and we consider them one at a time. Either way, we use a single plane wave as a localized "stand-in" for a more complex 3-D image-bearing object wave. Likewise, the reference wave is considered to be locally planar; it is usually a long-radius spherical wave, so this is a very good approximation.



As a sketching shorthand, we will indicate the diverging spherical wave by a cluster of arrows perpendicular to the surface of the wavefront. If we examine each arrow carefully, we might find that it has wavefronts within it that are too small to see. Thus these arrows are <u>not</u> the ***k***-vectors, or the "rays," or even the "ray bundles" you sometimes see mentioned in optics texts, although they resemble them all and are parallel to them ("rays," for example, are the line trajectories of imaginary light particles). The arrows are instead "mini-wave-beams" of a new sort, which let us draw accurate wave-optical pictures in familiar ray-optical ways. What they leave out is that the wavefronts within the arrows have a particular and fixed phase relationship; this doesn't usually matter for imaging calculations. Whatever we <u>ought</u> to call them, though, we will probably be careless and call them "rays" anyway, and it might even be useful to think of them as an extended variety of a generalized ray. Enough semantics!

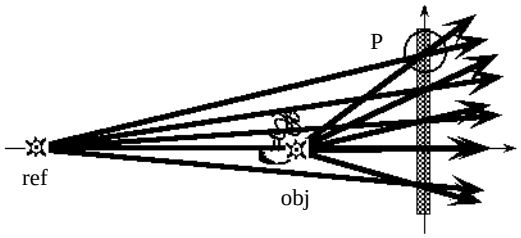


Suffice to say that we consider each point of the source to be emitting a diverging fan of rays, as does the reference point source. Where the object and reference rays cross, we compute the spatial frequency of the interference fringe pattern within their (not insignificant) width according to the grating equation, as they are both plane waves (locally). That pattern exposes the hologram plate, which is then processed to produce a modulating structure (or grating) with the same spatial frequency. A ray from the point source of illumination then strikes that grating, and becomes diffracted into several orders. The rays from any one order, such as the m=+1 order, can be traced back from several different locations on the plate, and their intersection will define the apparent location of a single source that produces them all, the "virtual" image of the point. If all goes well, holographically speaking, that location will be at the location in 3-D space of the original object point. We do the same for all points on the object (arguing by linearity that the hologram will hold all the little gratings independently of each other), and trace out the virtual image in 3-D space point by point. There are some subtleties here though: how does the eye know where, along the possible line of locations for the part of the ray that it receives, the intersection is? A mystery of visual perception perhaps, but at least consistent with simple triangulation (there is much more to spatial perception than triangulation, however!).



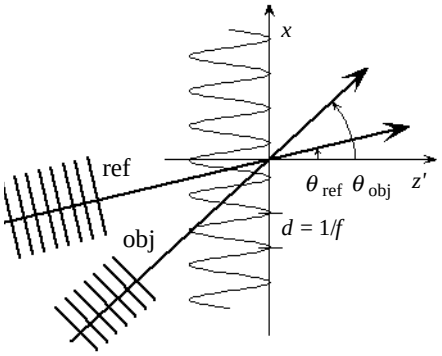
Mathematical Ray-Tracing:

The problem has now been reduced to keeping track of the fates of a few plane waves during interference and diffraction, something we are now well set up to handle. Consider this sketch of an "in-line" or "Gabor" hologram (this is the kind of hologram that Dennis Gabor was experimenting with in 1947 when he invented holography), the first type we will analyze in detail. We examine the beams or rays crossing a point, P, above the *z*-axis, where we construct a local coordinate system with axes *x'* and *z*'. An object point is on the *z*-axis at some distance, and the reference beam source is farther away, still on the *z*-axis. We examine the area around P, some distance up the *x*-axis, where the beams take on their local angle values, θ_{obj} and θ_{ref}. Where they overlap, a pattern of spatial frequency *f* is generated, where



$$f = \frac{\sin\theta_{\text{obj}} - \sin\theta_{\text{ref}}}{\lambda_1} \ . \qquad (1)$$

This becomes the spatial frequency of the grating created at P by exposure and processing of the holographic plate (that is, the plate doesn't expand or contract).



The plate is then illuminated by a point source at some other distance, producing the local illumination angle θ_{ill} . The output angles are therefore given by

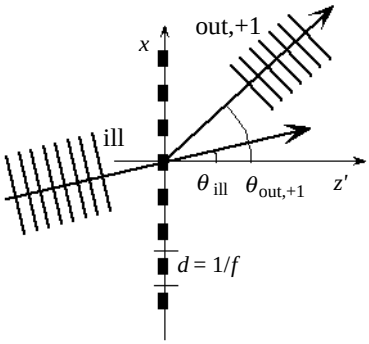
$$\sin\theta_{\text{out},m} = m\,\lambda_2\ f + \sin\theta_{\text{ill}} \ . \qquad (2)$$

Combining the relevant equations yields

$$\sin\theta_{\text{out},m} = m\frac{\lambda_2}{\lambda_1}\left(\sin\theta_{\text{obj}} - \sin\theta_{\text{ref}}\right) + \sin\theta_{\text{ill}} \ . \qquad (3)$$

This is our general raytracing equation, applied at every (*x*,*y*, 0) location on the hologram plate.

Now, considering the *m*=+1 term, it is clear that if λ_2=λ_1 and θ_{ill}=θ_{ref} (as it would be if the illumination source location were the same as the reference source location), then $\theta_{\text{out},+1}$=θ_{obj}. If this is true at every (*x*,*y*) point on the hologram surface, then the angle of the wavefront will have been reproduced everywhere on the hologram surface, and so will the wavefront itself (give or take an overall constant). If the reconstruction conditions are changed from "perfect," then the output angle follows the general relationship,



$$\sin\theta_{\text{out},+1} = \frac{\lambda_2}{\lambda_1}\left(\sin\theta_{\text{obj}} - \sin\theta_{\text{ref}}\right) + \sin\theta_{\text{ill}} \ , \qquad (4)$$

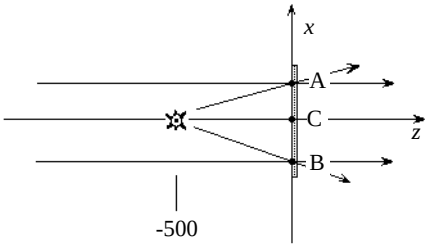
and the location of the image has to be determined by more careful numerical triangulation (as we do in the next section). In general, the rays of any order might not all diverge from (or converge to) an exact or single common point; in such a case the focus is said to be "aberrated" and its location is not well defined. For now, we will assume that the point is well-defined, and all wavefronts will be spherical.

Equation Three represents the ray-tracing equivalent of a general statement of holography. Indeed, it can be considered as a reduced version of the general phase equation, namely the relationship between the first-order *x*-derivatives of the wavefront phases. Unfortunately, it is limited to P locations in the *y*=0 plane (the *x*-*z* plane), and thus is not a fully three-dimensional statement. We will discuss fully-three-dimensional raytracing near the end of this chapter, and find that it takes us well beyond "shop math!" Fortunately, most of the relationships that we care about in this course are limited to the *x*-*z* plane, or very close to it, and we can prove them using the simpler 2-D form, which is Eq. 3. As a last resort, we can appeal to elaborations on the phase equation, which at least involve only scalar variables.

Numerical Example:

As an example, let's consider a specific case of recording a single point in the "in-line" or Gabor hologram geometry: the object is a point on the *x*-axis at

z_{obj}=-500 mm—that is, at location (x,y,z)=(0,0,-500), as sketched in the margin. The reference beam is a point source an infinite distance away, so that it produces a plane wave at the hologram surface, and all its rays are horizontal, parallel to the z-axis. We can assume that the intensities of the two beams are equal at the plate (although the reference beam would typically be 5 to 30 times stronger when recording an actual image hologram).



Getting even more concrete, let's try to trace at least two rays through the hologram in order to do some image location by triangulation. If we do one ray, we will get another "for free" in this case, and there is a third that is also available nearly "for free," so we will soon have even more rays than we really need. Let's consider first the ray-tracing location at point A, which is 50 mm above the z-axis. There, the object beam's angle is

$$\theta_{obj} = \tan^{-1}\left(\frac{50}{500}\right) = 5.71^\circ \ , \qquad (6)$$

and the reference beam's angle is 0°. Assuming that we are using a He-Ne laser, the spatial frequency of the grating formed at point A is therefore

$$f_A = \frac{\sin\theta_{obj} - \sin\theta_{ref}}{\lambda_1} = \frac{0.1 - 0}{633\times10^{-6}} = 157 \text{ cy/mm. .} \qquad (7)$$

By symmetry, the spatial frequency at point B, which is 10 mm below the z-axis, is also 157 cy/mm (note that there is a sign reversal, and that we are taking the magnitude of the result to be the spatial frequency).

The third "landmark" point of known spatial frequency is found by casting a line between the reference point source and the object point, and extending it to the hologram plane. In this example that location is at (x,y,z) = 0, 0, 0, the point we are calling C. As seen from this location, the two waves are "in-line," and the angle between them is zero; thus the spatial frequency here is also zero! We will often refer to this as the "zero-frequency-point" or "ZFP" in our analyses. For reasons that become clearer in a moment, it is also sometimes called the "hinge point" of the hologram.

Now, imagining that the hologram has been properly exposed and processed so that it has high contrast gratings everywhere, let's consider what happens when we illuminate the hologram again with the reference beam. Consider first the location A: the output angles are given by

$$\sin\theta_{out,m} = m\,\lambda_2\ f + \sin\theta_{ill} = m\cdot 633\times10^{-6}\cdot\ 157 + 0\ . \qquad (8)$$

And the values work out to be, for m between –2 and +2:

at A:

m	-2	-1	0	+1	+2
θ_{out}	-11.48°	-5.71°	0°	+5.71°	+11.48°

To calculate the angles at B, we have to be a little careful about which we call the m=+1 order. Although the spatial frequency is the same at B as at A, straightforward application of Eq. 1 would give a negative frequency. We have to apply the magnitude bars to get the same positive f. But in ray tracing we have to avoid the magnitude bars, and accept a negative frequency if we are to find the rays corresponding to the same m crossing at the same point. Note: this is not a problem for electrical engineers, who often deal with negative frequencies! The calculation at location B then produces the same results as at A, but with reversed signs (this follows from symmetry, without calculation in this case).

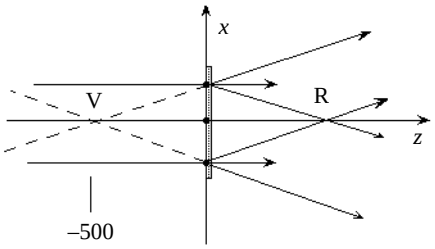
at B:

m	-2	-1	0	+1	+2
θ_{out}	+11.48°	+5.71°	0°	-5.71°	-11.48°

Point Image Locations - Real & Virtual

To find the locations of the focus points that represent the "images" produced by this hologram, let's look first at the easiest term, the m=+1 term. This produces rays propagating in the same direction as the original rays from the object at 5.71° away from the z-axis. Their "back-casted rays" cross each other, and the z-axis (recall that the ZFP ray will travel straight along the z-axis!) at z=-500 mm, so that a "virtual' image of a point will appear there, in the same location that the object point occupied. In this "perfect reconstruction" case, that location will be obtained no matter where we choose the ray-tracing locations A and B.

The m=-1 rays, on the other hand, are headed *toward* the z-axis at the same angle, 5.71°. Without much effort, we can predict that they will intersect at z=+500 mm, producing an on-axis focus in front of the hologram, a "real" image a distance in front of the hologram equal to the distance of the virtual image behind. It can be focused onto a card or ground glass as a bright point or (with much care) viewed directly with the eyes.



The angles of the two second-order rays are approximately twice as large as those of the two first-order rays, and produce virtual and real images that are roughly half the distance from the hologram. These follow from the definition of the trigonometric tangent, where h is the height of the ray-tracing location (plus and minus 50 mm in this example),

$$\frac{h}{-z_m} = \tan\theta_m \,, \qquad (9)$$

so that from the A and B locations we get these z-values:

m	-2	-1	0	+1	+2
z_m	+246 mm	+500 mm	∞	-500 mm	-246 mm

Approximations: A full expression of the image distance is given by cascading these relationships to yield

$$\frac{1}{z_m} = \frac{1}{h}\tan\left(\sin^{-1}\left(m\sin\left(\tan^{-1}\frac{h}{z_{\text{obj}}}\right)\right)\right). \qquad (10)$$

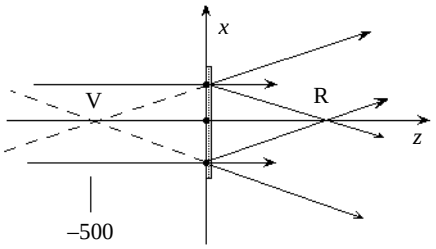
We note that an expansion of the trig terms of this expression yields an interesting approximation for $1/z_m$:

$$\frac{1}{z_m} \approx \frac{m}{z_{\text{obj}}} + \frac{h^2}{z_{\text{obj}}^3}\left(\frac{m\left(m^2-1\right)}{2}\right). \qquad (11)$$

If we let h decrease by a factor of 5, to 10 mm, then the second term decreases by a factor of 25, becoming negligible. We can describe this as a *paraxial* case, which involves only rays making small angles to the axis, so then $\tan\theta = \sin\theta = \theta$ (in radians), and which stay near enough to the axis so that h is small compared to the object and image distances. In this case the approximation reduces to

$$z_m = \frac{z_{\text{obj}}}{m} \,. \qquad (12)$$

However, in the general case we will have to do a more careful job of raytracing, yielding the deviations from the simple prediction that are shown in the table above.

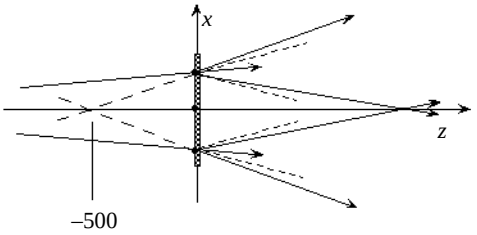


Illumination Wavelength Effects

If the wavelength of the illumination light, λ_2, is changed, then the angles of diffraction will change, and so will the image locations. Assuming that the new wavelength is 550 nm, for example, we find by calculation that the angles at A and the corresponding image locations become:

m	-2	-1	0	+1	+2
θ_{out}	–9.96°	–4.96°	0°	+4.96°	+9.96°
z_m	+284 mm	+576 mm	∞	-576 mm	-284 mm

Note that, because the green light is deflected "less radically" than the red, the images are formed farther out on the positive and negative z-axis. The higher-order images are still formed closer in than the first-order images, but are also farther out than the red images were.



The calculation equation is now given by

$$\frac{1}{z_m} = \frac{1}{h}\tan\left(\sin^{-1}\left(m\frac{\lambda_2}{\lambda_1}\sin\left(\tan^{-1}\frac{h}{z_{\text{obj}}}\right)\right)\right), \qquad (13)$$

where λ_1 and λ_2 are the recording and reconstruction wavelengths, respectively. The expansion then becomes

$$\frac{1}{z_m} \approx \frac{m\frac{\lambda_2}{\lambda_1}}{z_{\text{obj}}} + \frac{h^2}{z_{\text{obj}}^3}\left(\frac{m\frac{\lambda_2}{\lambda_1}\left(\left(m\frac{\lambda_2}{\lambda_1}\right)^2 - 1\right)}{2}\right). \qquad (14)$$

The paraxial approximation form then becomes

$$z_m = \frac{z_{\text{obj}}}{m}\frac{\lambda_1}{\lambda_2}. \qquad (15)$$

Source Distance Effects

Now let's consider the effects of moving the illumination source closer to the hologram. Leaving the wavelength at 550 nm (green), let's put the illumination at five meters from the plate, at $(x,z) = (0, -5000)$. Now the illumination angle at A is 0.57°, which will rotate all of the diffracted beams by roughly that amount (greatly exaggerated in the sketch). As a consequence, the m=+1 rays will be traveling at a larger angle to the z-axis, and the resulting virtual image will move in toward the hologram, while the m=-1 rays will travel at a smaller angle to the z-axis, and cross it farther out from the hologram. Plugging into the calculator again, we find the relevant numbers to be:

m	-2	-1	0	+1	+2
θ_{out}	-9.38°	-4.38°	0.57°	+5.54°	+10.54°
z_m	+303 mm	+652 mm	∞	-516 mm	-269 mm

Note that the first-order real image moves outward by almost 75 mm, and the first-order virtual image moves inward by only 60 mm.

The relevant cascaded mathematical expression is now

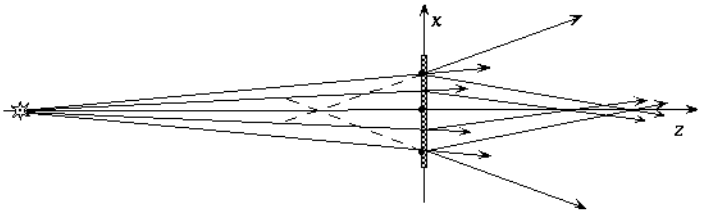
$$\frac{1}{z_m} = \frac{1}{h}\tan\left(\sin^{-1}\left(m\frac{\lambda_2}{\lambda_1}\left(\sin\left(\tan^{-1}\frac{h}{z_{\text{obj}}}\right) - \sin\left(\tan^{-1}\frac{h}{z_{\text{ref}}}\right)\right) + \sin\left(\tan^{-1}\left(\frac{h}{z_{\text{ill}}}\right)\right)\right)\right), \qquad (16)$$

and the corresponding expansion becomes

$$\frac{1}{z_m} = m\frac{\lambda_2}{\lambda_1}\left(\frac{1}{z_{\text{obj}}} - \frac{1}{z_{\text{ref}}}\right) + \frac{1}{z_{\text{ill}}}$$
$$+ \frac{h^2}{2}\left(\left(m\frac{\lambda_2}{\lambda_1}\left(\frac{1}{z_{\text{obj}}} - \frac{1}{z_{\text{ref}}}\right) + \frac{1}{z_{\text{ill}}}\right)^3 - \left(m\frac{\lambda_2}{\lambda_1}\left(\frac{1}{z_{\text{obj}}^3} - \frac{1}{z_{\text{ref}}^3}\right) + \frac{1}{z_{\text{ill}}^3}\right)\right) \qquad (17)$$
$$+ O[h^4] \,,$$

of which the first of the three lines, which is the paraxial approximation, suffices for the first-order images and matches the calculations very well if they are made for a ray height of only 10 mm or so.

Aberrations of Holograms - Spherical Aberration



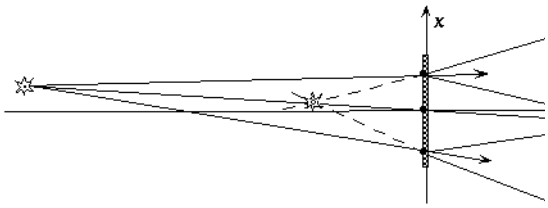
We have been pursuing this example to demonstrate exact ray tracing, and to develop some approximations to its results. But a few secondary issues have also come to light. We note that if we calculate z_m for a variety of ray heights, h, that rays from the edges of the hologram do not always cross the axis at the same z_m as do rays from near the axis, the paraxial rays. The resulting degradation of the focus of the image is called an *aberration* of the wavefront, describing its departure from perfectly spherical behavior. This particular type of aberration is directly analogous to *spherical aberration* in simple glass lenses (a consequence of the spherical shape of their surfaces), and so has been given the same name in spite of there being no spheres involved.

Aberrations of Holograms - Chromatic Aberration

While we are at it, we might as well point out that if the hologram is illuminated with light of several different wavelengths, each wavelength will be focused to an image at a different distance, and the overall focus will be degraded. Glass lenses produce a similar effect, due to the prism-like nature of their edges, and the result is called *chromatic aberration*. It is a much stronger effect in holograms than in lenses, and not so readily correctable (as we shall see!).

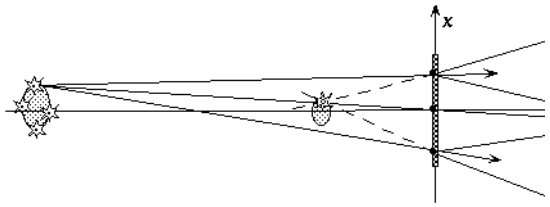
Source Angle Effects

Here, we consider the effect of moving the illuminating point source to one side of the *z*-axis, say upward by 50 mm. Now, the rays illuminating points A, B, and C all change angle by about the same amount, and with the same sign. The fan of diffracted output rays also rotates by roughly that angle at each location, and their intersections necessarily rotate also, so that the images are moved away from the *z*-axis.



Source Size Effects

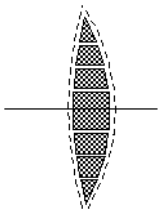
We can now imagine that if the illumination point moved continuously from the *z*-axis to a point 50 mm above, that the various images would move continuously from their original locations to the locations described in the above section. And, if an array of several sources were placed on a line connecting those two locations (each incoherent with all the others), then an array of several images would appear on a line between the two extreme images. Thus we can already begin to see how a spatially incoherent source can produce blur in an image.



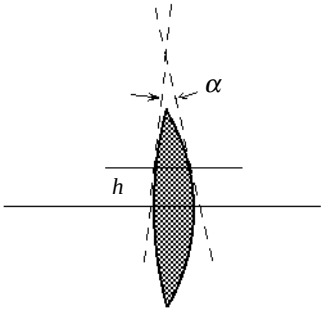
Comparison of Paraxial Hologram and Lens Optics

The expansion of the expression for the numerically ray-traced image location, Eq. 17, has as its first line of terms the *paraxial approximation* of holographic ray-tracing. This relationship is

$$\frac{1}{z_m} = m\frac{\lambda_2}{\lambda_1}\left(\frac{1}{z_{\text{obj}}} - \frac{1}{z_{\text{ref}}}\right) + \frac{1}{z_{\text{ill}}}\,. \qquad (18)$$



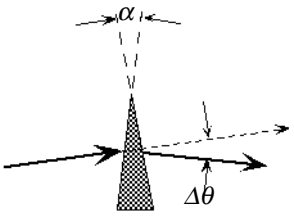
In the cases we have been describing, all of the z_{abc} are negative quantities, representing locations to the left of the x-axis, except for the z_m for negative m in most cases. The point of this section is that this formula is identical in form to the equation that describes focusing by a refracting or glass lens if the focal length is suitably described. The analogy between the elements of a hologram and conventional lenses is a very powerful one to those who are familiar with optical components. Assuming that this might not be the case here, we will demonstrate a few of the simpler principles along the way.



Definition of a Glass Lens

A normal glass lens (the same ideas will apply to plastic and liquid lenses) is defined by two spherical surfaces that cause the lens to be either thicker at the center than the edges (for a so-called *positive* lens), or thinner at the center (for a so-called *negative* lens). We denote the radii of curvature of the surfaces by R_1 and R_2, which are positive if they are convex to the right (or concave to the left). Thus, in the sketches here, R_1 is negative and R_2 is positive. Depending on which comes first, the lens might be positive or negative. The thickness at the edge or center doesn't affect the lens focusing in the thin-lens approximation, only the curvatures do, or rather the change in slopes of the surfaces as a function of height above the z-axis.

Such a lens can be approximated by a pile of prisms, as shown here. The central parallel-sided block of the prism doesn't deflect rays, only the refraction at the tilted surfaces at the edges does—which leads to a Fresnel-lens-like representation.



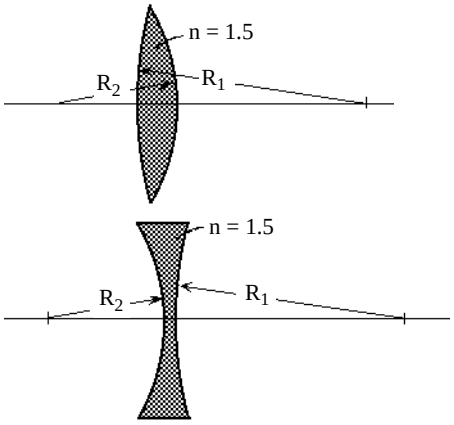
For a real lens, the surface slopes change continuously. The thickness at any height, $t(x,y)$, is given by (using the same approximations for spherical surfaces as for spherical wavefronts)

$$t(x,y) = t_0 + \frac{1}{2}\left(\frac{1}{R_1} - \frac{1}{R_2}\right)\left(x^2 + y^2\right), \qquad (19)$$

where t_0 is the center thickness. The downward-pointing angle, α, between the surfaces at height h (that is, at $(x,y)=(h,0)$) is given by

$$\begin{aligned}\alpha(h) &= -\left.\frac{\partial t(x,y)}{\partial x}\right|_{x=h,y=0} \\ &= -\left(\frac{1}{R_1} - \frac{1}{R_2}\right)h\,,\end{aligned} \qquad (20)$$

which increases linearly with height for a positive lens.

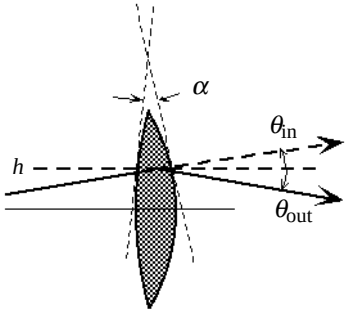


Ray Deflection by a Lens

We saw in Problem Set #0 that ray deflection by a prism has to be determined fairly carefully, due to the non-linearities of the sine functions in Snell's Law. However, for rays roughly perpendicular to the surfaces, where $\sin\theta = \theta$, a simple rule for the angle of deflection, $\Delta\theta$, as a function of the apex angle, α, and the index of refraction, n, can be used:

$$\Delta\theta = (n-1)\alpha\,. \qquad (21)$$

Image Distances: the Focus Law



Now, if we adopt the same illumination source convention as above, a source ray striking the lens at height h will be incident at an angle θ_{ill} given by

$$\theta_{ill} = \tan^{-1}\left(\frac{h}{R_{ill}}\right). \tag{22}$$

The output angle, θ_{out}, will be given by θ_{ill}-α, and will appear to be coming from a location z_{image} given by, within the paraxial approximation,

$$\begin{aligned}\frac{1}{z_{image}} &= -\frac{\tan\theta_{out}}{h} \approx -\frac{\theta_{ill}-\Delta\theta}{h} \\ &= -\frac{\dfrac{-h}{z_{ill}} + (n-1)\left(\dfrac{1}{R_1}-\dfrac{1}{R_2}\right)h}{h} \\ &= \frac{1}{z_{ill}} - (n-1)\left(\frac{1}{R_1}-\frac{1}{R_2}\right),\end{aligned} \tag{23}$$

which is completely independent of h!

Now, if we let $z_{ill} \to \infty$, the image will be formed at a positive z_{image} (recall that R_1 is negative), at a distance that we will call the "focal length" of the lens, or FL. This is the distance at which an image of the sun will be formed by a burning glass, for example. Note that the focal length of a "positive lens" is a positive number, so that the focus is formed to the right of the lens—a real image of the illumination source. The focal length of the lens is given by the so-called "lens-maker's formula" as

$$\frac{1}{FL} = (n-1)\left(\frac{1}{-R_1}+\frac{1}{R_2}\right). \tag{24}$$

z_{ill}	z_{image}
–∞	FL
–10 FL	1.1 FL
–2 FL	2 FL
–1.1 FL	10 FL
–FL	+∞
–FL/2	–FL

Note that a variety of combinations of curvatures can produce the same focal length lens; they differ only in their higher order optics—aberrations and so forth—so that dish-shaped or "meniscus" lenses are usually used for reading glasses, and flat-on-one-side or plano-convex lenses are used for collimators. If you read about this topic in other optics books, beware of differences in definitions of the signs of curvatures that might change some signs in the corresponding results.

Substituting the focal length into Eq. 23 then gives the focusing equation well known in all of optics:

$$\frac{1}{z_{image}} = \frac{1}{FL} + \frac{1}{z_{ill}}\ . \tag{25}$$

Some combinations of illumination and image distances are shown for reference.

Comparison to Holographic Focusing

Now lets re-examine the focusing law for the paraxial ray-trace of holograms and dwell on a few similarities:

$$\begin{aligned}\frac{1}{z_m} &= m\frac{\lambda_2}{\lambda_1}\left(\frac{1}{z_{obj}}-\frac{1}{z_{ref}}\right)+\frac{1}{z_{ill}} \\ &= \frac{1}{FL_{holo}}+\frac{1}{z_{ill}},\end{aligned} \tag{26}$$

where we define the focal length of the hologram, FL_{holo}, as

$$\frac{1}{FL_{\text{holo}}} = m\frac{\lambda_2}{\lambda_1}\left(\frac{1}{z_{\text{obj}}} - \frac{1}{z_{\text{ref}}}\right). \qquad (27)$$

That is, each order of diffraction by the hologram corresponds to focusing by a different glass lens, where the focal lengths of the lenses are both positive and negative and the focal lengths of higher orders are integer fractions of the first-order focal length. Further, the focal length is inversely proportional to the wavelength of the light used for reconstruction, so that red light is focused closer to the hologram than blue light in each order.

The plus and minus first-order holographic lenses always have the same diffraction efficiency, and always appear together, occupying the same location and providing the effects of positive and negative lenses of equal and opposite focal length. It is though two differently-shaped pieces of glass occupied the same physical space! This lens-pair model of a simple holographic lens will arise again and again in our discussions to come.

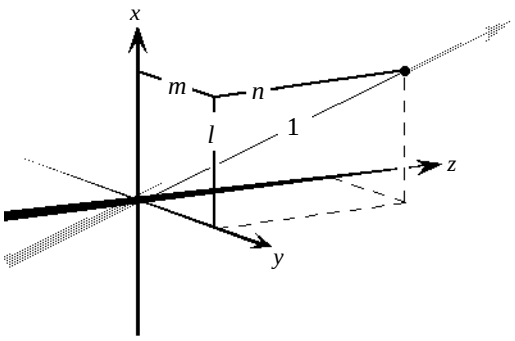
Depending on your own insight into conventional or refractive optics, the use of the lens analogy may or may not be useful, but it seems comforting to know that the results of diffraction by these simple holograms have at least a small resemblance to centuries-old optical principles!

Three-Dimensional Ray Tracing

The extension to angles out of the *x*-*z* plane has been shown by Welford[1], among others, to be (adapted to our notation)

$$\boldsymbol{n}\times\left(\boldsymbol{r}_{\text{out},m} - \boldsymbol{r}_{\text{ill}}\right) = m\frac{\lambda_2}{\lambda_1}\boldsymbol{n}\times\left(\boldsymbol{r}_{\text{obj}} - \boldsymbol{r}_{\text{ref}}\right), \qquad (28)$$

where $\boldsymbol{n}$ is a unit vector perpendicular to the hologram surface at the raytracing location, and the $\boldsymbol{r}_{\text{abc}}$ are four unit vectors in the directions of the corresponding obj, ref, ill, and out,*m* rays. The $\times$ denotes the vector cross product. Clearly this takes our discussion into vector algebra proofs for which "shop math" hasn't yet prepared us. However, the equation can be broken down into components that do resemble the equations we have been working with.



Let the individual ray unit vectors be represented by their components in the *x*-, *y*-, and *z*-directions, which are the cosines of the angles of the ray $\boldsymbol{r}_{\text{abc}}$ with the *x*-, *y*-, and *z*-axes, respectively, which we denote as ℓ_{abc}, m_{abc}, and n_{abc} (the reader will have to keep the distinction between "*m*, the order number," and "m_{abc}, the direction cosine" clearly in mind here). The ray unit vector could then be given by $\left(\ell_{\text{abc}}\,\mathbf{x},\ m_{\text{abc}}\,\mathbf{y}, n_{\text{abc}}\,\mathbf{z}\right)$, where **x**, **y**, and **z** are unit vectors in the corresponding directions. However, we shall analyze Eq. 28 for the *x*- and *y*-components separately to give

$$\begin{aligned}
\ell_{\text{out}} &= m\frac{\lambda_1}{\lambda_2}\left(\ell_{\text{obj}} - \ell_{\text{ref}}\right) + \ell_{\text{ill}}\,, \\
m_{\text{out}} &= m\frac{\lambda_1}{\lambda_2}\left(m_{\text{obj}} - m_{\text{ref}}\right) + m_{\text{ill}}\,, \text{ and} \qquad (29) \\
n_{\text{out}} &= \sqrt{1-\ell_{\text{out}}^2 - m_{\text{out}}^2}\quad .
\end{aligned}$$

If m_{abc} is constrained to be zero, so that the abc-ray lies in the *x*-*z* plane, then ℓ_{abc}, which is always the cosine of the angle between the ray and the *x*-axis, becomes equal to the sine of the angle between the ray and the *z*-axis, the θ_{abc} as we have been defining it. If this is true for the obj, ref, and ill rays (and thus for all the out,*m* rays), then the equations we have been using are actually just half of the components of the full three-dimensional ray-tracing analysis. Our simplified approach could be extended whenever desired to handle the other relevant components also. However, we will continue as we have been doing, pointing out this interesting connection only in passing.

Conclusion

Raytracing is highly accurate but computationally intensive and barren of physical insight. Simple approximations yield workable formulae that are handy for the purposes of designing systems and testing ideas. Analogies with conventional refracting optics can be drawn, although holograms are seen to be many lenses in one. Fully-three-dimensional ray tracing is seen to be possible with fairly straightforward extensions from techniques we have limited to the *x*-*z* plane.

[R]eferences:

1. W.T. Welford, "A Vector Raytracing Equation for Hologram Lenses of Arbitrary Shape," in *Optics Communications*, **14-3**, pp. 322-323 (July 1975)