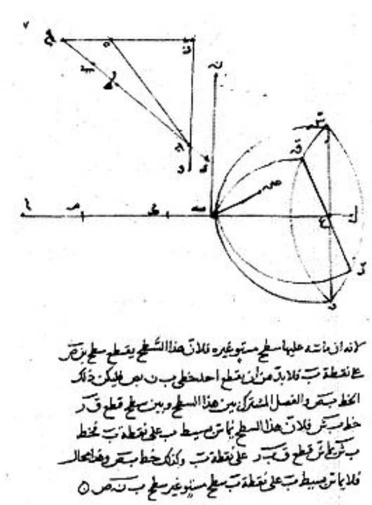
Tutorial: Geometrical Optics and Ray Tracing

Chris Brown

May 2, 2011



1 Abstract

Ray tracing in simplified models of optical systems illustrates the advantages of linear approximations and the idea of component transfer functions being composed into the transfer function of a complex system. Ray-casting, or rendering, is the workhorse of computer graphics: we give a brief introduction. This tutorial is meant to be suitable for freshmen and was originally given with Matlab as the assumed computational platform.

It is partnered with some computational exercises with sample code in a separate tutorial [5].

2 Ray Tracing: What Happens in a Microscope?

2.1 Mathematical Models of Lenses and Rays

"Now consider a spherical chicken..."
– hoary punch line

Most of this material is from the first four (out of 25) chapters in [2], a typical sophomore-level optics textbook.

Light is an electromagnetic wave with many fascinating properties, like also being an elementary particle. So much for Chapters 1 and 2 of [2].

Refraction is the bending of light by an optical medium; *reflection* is the bouncing-off of light from a surface. *Geometrical optics ([2], Ch.3)* discusses a special case of physical optics that tosses overboard all the wave (and quantum in general) attributes of light. Christiaan Huygens (Dutch, b. 1629) had an ether-based theory of light that was a sort of geometric optical approximation in that it failed for diffraction but did derive the law of refraction. Pierre Fermat (French, b. 1601) used a minimum-principle argument (minimizing light's travel time) to prove the same law. His principle is also not quite right, since there turn out to be time-*maximizing* light paths and cases where all paths are equivalent in time (between the foci of an ellipsoidal mirror, for instance).

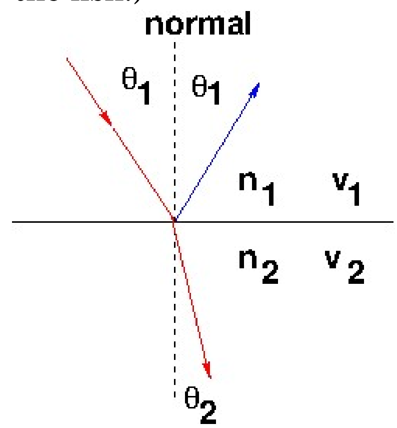
The law of refraction is named, for no very good reason, Snell's law. For one thing, his name wasn't Snell: it was Willebrord Snellius (Dutch, b. 1580). For another, the law has been known a long time, You can bet the first human ancestor whose spear went over the fish knew about it. One Ibn Sahl (Persia, b. ca. 940 AD) discovered it (his manuscript is the frontispiece). It was investigated by Ptolemy (Greek, b. ca. 90 AD), and Witelo in the middle ages, but they didn't have the formalism to write it down, and used tables instead. So what is it, anyway?

The law of refraction (Snell's Law): When a ray of light is refracted at an interface between two uniform media, the transmitted ray remains in the plane of incidence and the sine of the angle of refraction is directly proportional to the sine of the angle of incidence.

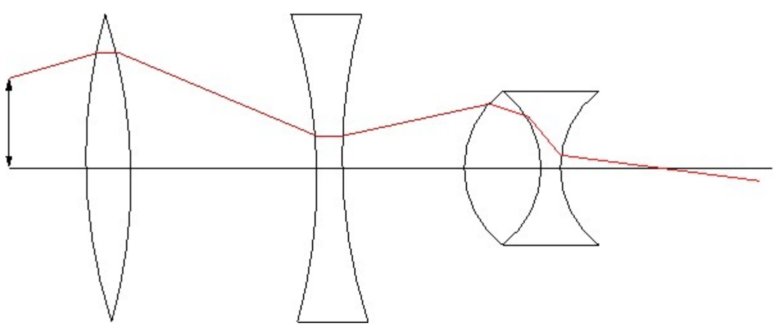
There's also a **law of reflection:** When a ray of light is reflected at an interface dividing two uniform media, the reflected ray remains in the plane of incidence, and the angle of reflection equals the angle of incidence. The plane of incidence is defined by the incident ray and the surface normal vector at the point of incidence.

Snell's Law is usually written:
$n_1 \sin\theta_1 = n_2 \sin\theta_2$
with n_1, n_2 the *refractive indices* of the two media. The refractive index n of a medium is the ratio of v, the speed of light in that medium, to c, the speed of light in a vacuum (299,792,458 m/s). That is, $n = \frac{c}{v}$. Slower light speed in the second medium means more refraction. Air has refractive index 1.000293, water 1.333, glass around 1.5, and diamond 2.42.

Here's a picture of reflection and refraction. The vertical dotted line represents the direction perpendicular to the horizontal surface we're seeing edge-on as a horizontal line. Reflection sends the ray coming in at angle θ_1 out (in the same plane, here the plane of the paper) at the same angle. Refraction bends the light at the medium's interface (also in in the same plane) if $n_1 \neq n_2$, which would imply the light speeds v_1, v_2 also differ. (This diagram shows why that spear went over the fish.)



Now using these two laws we can analyze systems of lenses and mirrors like the all-lens system below. In this and following diagrams, a light ray originates enters the system from the left at some angle and distance from the lenses central axis and our job is to trace its path.

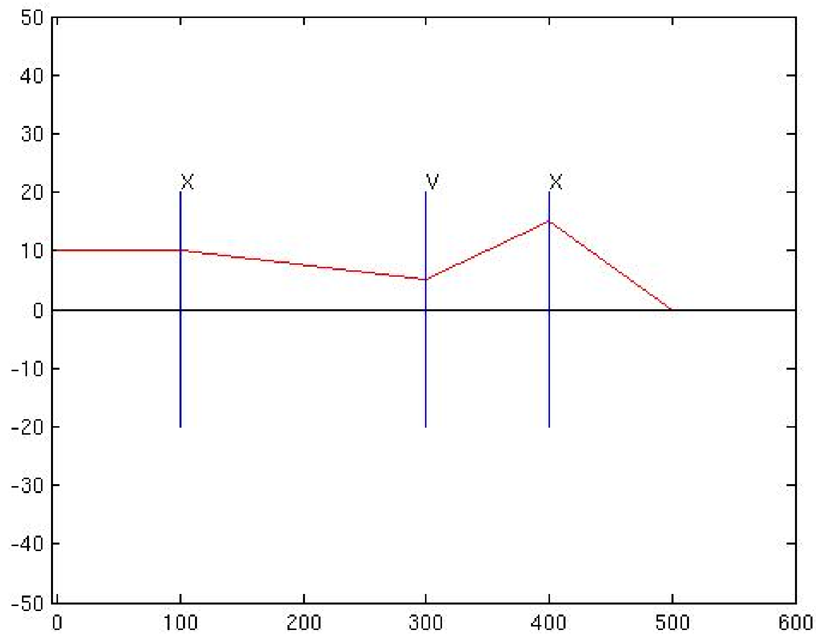


Ideally, we'd like to use Snell's Law in our geometrical optics approximation to reality. But we have a problem. Those sines are *nonlinear* functions. We won't be able to find solutions by the simple-minded linear methods we know and love. Well, then! Let's just heave some inconvenient reality over the side: Let's consider only *paraxial* rays, those that never make large angles with the optical axis. In fact, only those for which we can approximate, *with angles measured in radians, not degrees (!!)*,
$\sin\theta = \tan\theta = \theta$.

That sure cleans things up: now the paraxial Snell's Law is
$n_1\theta_1 = n_2\theta_2$

The paraxial assumption is an example of the "small angle approximation" so dear to physics and engineering. It is a *linear approximation*, which describes a given function (in some locality) as a linear function. Here, for example $\sin(x) \cong x$ near 0 (the graph of $sinx$ takes off from 0 headed up at 45 degrees, exactly like the graph of x, and stays "close" to x for "small" x.) In the future you'll often expand functions in infinite (Taylor) series and ignore all but the first term to get your approximation. It happens that $\sin(x) = x - x^3/3! + x^5/5! - ...$: near 0, and for small x, a good approximation is just to use the first term.

Now notice in the figure above that different rays are going to travel different distances through the lenses, which are fatter and thinner. So just what happens depends on just where and at what angle the ray enters and leaves the lens. Well, *that's* really inconvenient... But wait! We know what to do! We just jettison that annoying bit of reality, too. The *thin lens* assumption is: "consider an infinitely thin lens" with all the refractive power and none of the annoying size of an actual object. A diagram might look like this, with X for convex and V for concave lenses:



Using pretty simple algebra (but pretty complicated diagrams) [2] derives elegant formulae for useful, practical properties (e.g. you've probably heard of "focal length") that help in understanding and creating lenses. They're derived from the basic geometry (say radii of curvature, distance) and basic physics (Snell's law, indices of refraction). The results are usable parameters that describe spherical mirrors, refraction at spherical surfaces, thin lenses, etc.

An important thin lens concept is its *focal length f*, the (signed) distance to the image it forms of an object at infinity. Convex and concave lenses have positive and negative focal lengths, respectively. The *lensmaker's equation* predicts the focal length of a lens in terms of its refractive index, that of the medium it is in, and its radii of curvature.

In air (As usual, ignoring the inconvenient fact that its refractive index is only approximately 1), one version of the basic *thin-lens equation* states that an object at distance s is imaged at point s' by a lens of focal length f if
$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f}$.
The *power P* of a lens measures how strongly it bends light, and is defined as $1/f$.

In an $x - y$ plane (like our lens diagrams), We measure angles (in radians) that rays deviate from the optical system (horizontal, x) axis, so rays parallel to the x-axis are at 0 radians. Rays from the left rising in the y direction have positive angle ($\theta > 0$). A lens bends

an incoming ray of angle θ, according to our approximations, to give it a new angle:
$\theta' = \theta - yP$
if the ray impinges on the lens at height y. Thus
$\theta' = \theta - (\frac{y}{f})$

That's our quickie summary of [2], Ch. 3. Section titles to the contrary notwithstanding, we certainly do **not** have the tools to describe what really goes on in a real microscope. Instead of a spherical chicken we have a theoretical, paraxial, flat microscope in a vacuum that allows some (approximate) reasoning. Snideness aside, creating solvable linear approximations is a crucial skill you'll be developing in the next three years.

2.2 The Method

Often linear systems are described by some combination (summation, multiplication, convolution,..., or other linear operators) of transformations produced by the system's parts. The light of our rays is considered to move left to right. We describe a ray by a column vector of its instantaneous height and angle of flight (toward the right): $[y, \theta]^T$ We'll see that paraxial optical elements can be described by 2×2 matrix transformations (*ray transfer functions*) representing what happens to a ray at an optical element. A cascade of such transformations (say a sequence of lenses) combines the matrices in order by multiplication, so the product of the transfer matrices describes the optical system as a *system transfer function*. This material is covered in [2] Ch. 4, for example.

To diagram the system as above, we just make a 2-D plot of y versus x, with the x axis being the optical axis. Sounds like a job for Matlab. An object being imaged is considered to be in an *input plane* at the left of the diagram (below, it's always at $x = 0$), and the *output plane* is wherever we want to compute the final height of the ray. Things we might want to know about the system include (there are many more) its *front* and *back focal length*. The latter is just the effective focal length of the entire system, and the former is similarly the focus point of angle-0 rays coming in from the right, headed left. The *axial image point* of an object point nearer than infinity on the optic axis is where on the optic axis all its rays are focused (yes, they all go through one point). The *image plane* is the plane (line in a 2-D diagram) through the axial image point perpendicular to the optic axis. The *linear magnification* of the system is $\frac{y_f}{y_0}$, the relative height of a ray at the input and output planes.

Consider a ray $[y_0, \theta_0]^T$ whose front (imagine it's the track of small "light particle") is moving (translating) through a homogeneous medium. Angles are measured in radians, remember. If we follow the ray for a distance L along the x axis as it maintains its direction θ_0, its new description (by high school geometry) is $[y_0 + L \tan \theta_0), \theta_0]^T$. The paraxial assumption is that $tan(x) = sin(x) = x$, so making that substitution we get a linear *translation* transformation of the ray's description:

$$\begin{bmatrix} y_1 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} 1 & L \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_0 \\ \theta_0 \end{bmatrix}$$

That is, its angle doesn't change but its (our imaginary light particle's) distance from the center axis changed by $L \tan \theta_0$. This simple 2×2 matrix is how we describe the change in the ray as it moves through some uniform medium for an axial distance.

Remember this equation from the end of the last section? It describes the effect on the ray of hitting a thin lens:
$\theta' = \theta - (\frac{1}{f})y$. So for a thin lens, it's easy to put the *lens equation* into matrix form:

$$\begin{bmatrix} y_1 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1/f & 1 \end{bmatrix} \begin{bmatrix} y_0 \\ \theta_0 \end{bmatrix}$$

That is, the ray is bent – it suddenly takes a new direction.

Given the work in [2] Ch. 3, it's also easy to come up with simple transfer matrices for spherical or refraction interfaces, spherical mirrors, and even thick lenses. We won't use those, but they're simple.

Express a paraxial system with elements described by 2×2 matrices, say $M1, M2, M3, M4$ in order from left to right. Then for ray r,
$M1 * r$ is the ray after the first element, where $*$ is matrix multiplication,
$M2 * M1 * r$ after the second, since the 2nd acts on the output of the 1st,
$M3 * M2 * M1 * r$ after the third, and
$S = M4 * M3 * M2 * M1$ (S is also a 2×2 matrix) is the complete transformation and is called the *system (transfer) matrix*. The ray winds up being $S * r$. Plotting its segments (between elements) in an (x, y) plane and adding vertical lines where the optical elements are gives ray-tracing plots like those on pages 4 and 7.

This ray-transfer technique is used to find other basic geometric properties describing an optical system, such as its nodal points and first and second principal planes (things optics students care about). Not only that, but the four elements of the transfer matrix have intuitive semantics, which can be illuminated by considering the physical meaning of setting each of them to zero. For now, though, we're done.

2.3 Examples

A *magnifying (burning) glass* is a simple convex lens. Look through it, it magnifies; shine the sun through it and it concentrates rays at the distance of its focal length. We can describe it as a $S = M3M2M1$ system operating on a ray r. $M1$ and $M3$ are translations and $M2$ is our convex lens. Let's make r arrive at the lens parallel to the axis (thus $\theta = 0$), simulating the parallel rays of a distant sun, and 5 cm. away from the central axis:

$$r = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$$

The sun is a long way off but we can start our ray, which is moving parallel to the axis, at some closer point, namely at $x = 0$. Let's say the first lens is 10cm away to the right. This means $M1$ is a translation transformation:

$$M1 = \begin{bmatrix} 1 & 10 \\ 0 & 1 \end{bmatrix}$$

Say our lens has a focal length of 40 cm:

$$M2 = \begin{bmatrix} 1 & 0 \\ -(1/40) & 1 \end{bmatrix}$$

And let's check and see if the bent ray r' actually winds up on the axis (distance 0) at a distance equal to the focal length — that would make sense:

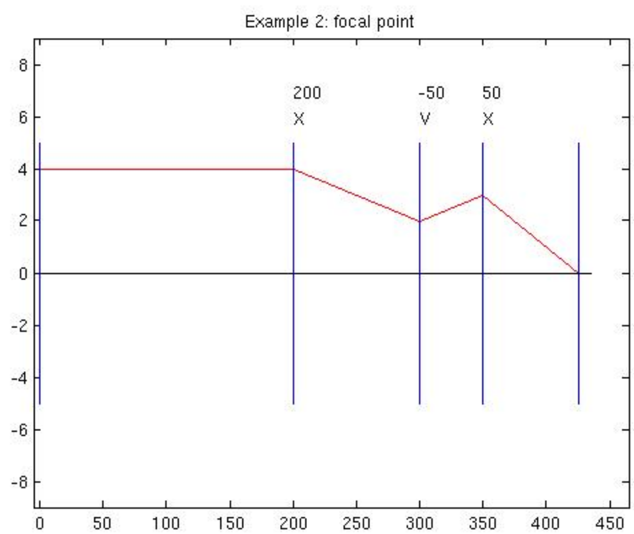
$$M3 = \begin{bmatrix} 1 & 40 \\ 0 & 1 \end{bmatrix}$$

$$S = M3M2M1 = \begin{bmatrix} 0 & 40 \\ -0.025 & 0.75 \end{bmatrix}$$

$$r' = Sr = M3M2M1 * r = \begin{bmatrix} 0 \\ -0.125 \end{bmatrix}$$

So 40 cm out, this ray is at the optic axis, which make sense. If r hits the lens closer to its center (the axis), we'd expect r' still to be at the axis 40 cm out, and we'd expect it to have been bent less: you can check that if $r = [3, 0]^T$, $r' = [0, -0.075]^T$; bent less and intersecting all other axis-parallel incoming rays at the focal length.

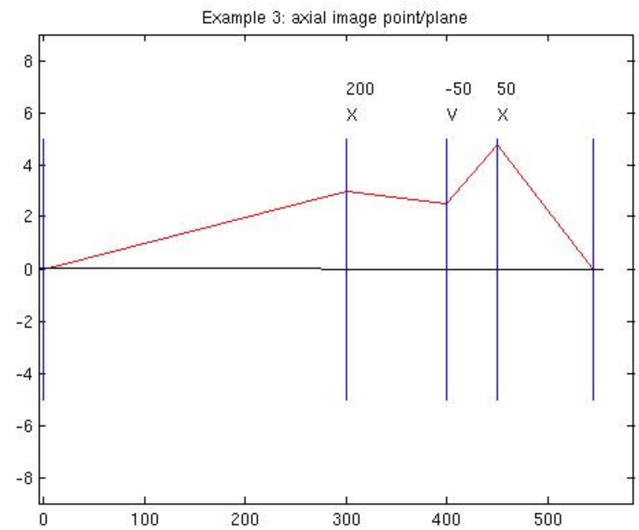
Let's find some important properties (here, distances) for an example optical system. Consider three lenses of focal length 200, -50, and 50 mm. The first and second are separated by 100 mm. the second and third by 50 mm. The input plane is 200 mm. in front of the first lens. What is the (back, effective) focal length of the system?



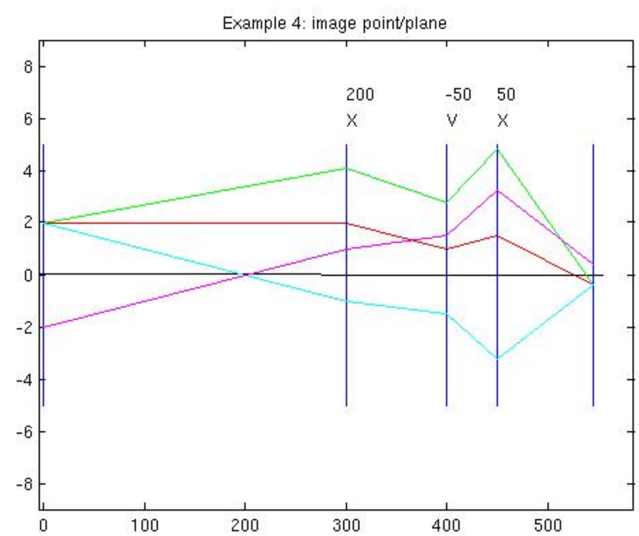
We send a ray parallel to the axis into the 6-element system (translation, lens, translation, lens, translation, lens). The last ray we get is the one emerging from the last lens at X=350. It turns out to be $(3, -.04)^T$. We're after the focal length $g = y/\tan\theta$, (the base of the final triangle in the figure from X= 350 to X= 425). With the small angle assumption, the tangent of the angle is the angle, so this ray intercepts the axis at $g = y/\theta$, or 75mm. We can then add a final, seventh, transfer of this distance, which should bring the ray down to the axis at the focal point, and that final system produced the figure above.

The *axial image point* of a system is where the image of a ray starting out at the origin crosses the optic axis. It's like the focal point, and calculated the same way, only using a different initial ray. It turns out that all the system's image points from an object at the origin will fall in the plane at that distance, so that is where an in-focus image will be formed.

For fun let's use a slightly different 3-lens system: lenses are spaced out by 300, 100, and 50, and have focal lengths 200, -50, and 50. Shooting out a ray from the origin at angle .01 and calculating the axial image distance (it's 95 mm. out from the third lens) we can compose a final translation after the 3rd lens of 95 mm to get this plot:



With this same system, let's shoot three rays from some non-zero y height at three different angles: they should intersect at the same axial image plane as above. Nicely, all the rays from the same object point wind up at the same image point (coincidence?!) and we can use the y height at the image plane to get the ratio object-height/image-height, or linear magnification, as .2 (the image height is upside-down).



3 Ray-Casting and Systems of Linear Equations

Ray-casting is the central, inner-loop operation in rendering computer graphics representations into images. The idea sounds simple: given a point of view and direction of gaze (ray), compute the color and intensity of the light in that direction. In *geometrical optics* the wave nature of light is ignored, and to follow light-paths (rays) through optical systems we only need geometry, some algebra, and a few physical laws.

A basic case is: given a 3-D point of origin for the ray and its direction, where does it intersect a given plane in space? No different from the "line-intersect plane" problem, and in optics sometimes called "intersecting a ray with a plane mirror", because mirrors are the most common use of optical "flats".

As usual, represent points p, x, r etc. in 3-D by $(x, y, z)^T$ column vectors. Let's represent directions α, β etc. in 3-D by $(x, y, z)^T$ *unit vectors*. Directions form a family that lives on a sphere of unit radius centered on the origin. Every direction out from the origin corresponds to a point on the sphere, and also to a vector whose head is on this *Gaussian Sphere*.

The standard way to intersect a vector with a surface in space is to stretch the vector out in its direction until it intersects the surface. The resulting length is all we need, since we can then compute the intersection point. So: Any point on a ray can be written
Ray: $r = r_0 + d\alpha$,
with d the length, r_0 the ray's origin, and α its direction. This vector equation represents three linear equations in x, y, z.

The plane equation is linear, and a 3-D version of the familiar line equation:
Plane: $Ax + By + Cz + D = 0$

Written like this, $[A, B, C]^T$ can be interpreted as a vector specifying a direction — in fact the direction normal to the plane. Scaling the whole equation so that this direction is a unit vector gives the new, scaled value of D a meaning: it's the perpendicular distance from the plane to the origin.

We're going to simplify our work right off by forcing our infinite mirror, or plane, to pass through the origin, so we can describe it with a linear equation: one strictly in (x, y, z) (i.e. $D = 0$: no pesky constant).
Plane through Origin: $Ax + By + Cz = 0$, or
$(A/C)x + (B/C)y + z = 0$.

Aside: Ray-casting actually wants to be (is best) described with *projective geometry*, a fascinating and pretty branch of mathematics. A (pinhole) camera projects the scene onto the sensor by ray-casting, and yields the familiar perspective effects like parallel railroad tracks seeming to come together out in the distance. This is a *non-linear* transformation: linear transformations preserve parallelism. Luckily, there is a neat trick for doing the needed calculations linearly (strictly matrix operations) only in a simple four-dimensional version of space (*homogeneous coordinates*). For a more general ray casting project, still pretty easy, see

`http://www.cs.rochester.edu/u/brown/173/exercises/matrices /applications/raytrace.html` – Pinholes and Beach-balls. That project visualizes spheres since infinite planes are boring and non-infinite planar surfaces are very fiddly.

The equations Ray and Plane above are four linear equations. A solution to them gives the d at which the ray intersects the plane, and we're done. Preview: we write down the equations, do a quick massage, write them in matrix form, and the answer is obvious.

Ray origin: (x_0, y_0, z_0).

Ray direction: $(\alpha_1, \alpha_2, \alpha_3)$

3 Ray Equations, 1 Plane Equation:

$x = x_0 + \alpha_1 d.$
$y = y_0 + \alpha_2 d$
$z = z_0 + \alpha_3 d$
$(A/C)x + (B/C)y + z = 0$

We need four equations for unknowns x, y, z, d, (where the ray pierces the plane and how long the ray must be produced from its origin to meet the plane). But clearly we only really need three numbers –f for instance from x, y, d we can easily solve for z using the Plane through Origin equation. Preemptively using that equation first, we get an expression for z to substitute into the third ray equation and we can rewrite the system as:
$x_0 = x - \alpha_1 d$
$y_0 = y - \alpha_2 d$
$z_0 = -(A/C)x - (B/C)y - \alpha_3 d.$

We could keep on substituting to solve the system, but let's express it as a matrix equation:

$$\begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} = M \begin{bmatrix} x \\ y \\ d \end{bmatrix}, \text{ where } M = \begin{bmatrix} 1 & 0 & -\alpha_1 \\ 0 & 1 & -\alpha_2 \\ -(A/C) & -(B/C) & \alpha_3 \end{bmatrix}$$

But hold it! For any vectors u, v, if $u = Mv$ then $v = M^{-1}u$. So we can solve for what we want (x, y, d) in terms of what we know or can easily compute: that is
x_0, y_0, z_0, and M^{-1}.

Then we ignore d, put x, y into the Plane through Origin equation, get z, and we've got our intersection-point $[x, y, z]^T$. Using matrix inversion to solve systems of linear equations is not the accepted way to do it (look up "Gaussian elimination"), but it's not wrong and it is easy to explain.

4 References

1. Tom Brown; *Course Notes, Optics 211* 2008, Part I.
2. F.L. Pedrotti and L.S. Pedrotti; *Introduction to Optics*, Prentice-Hall 1987.
3. Wikipedia.
4. R. W. Ditchburn; *Light*, Wiley Interscience, N.Y. 1963.
5. C. M. Brown; "Tutorial: Raycasting and Raytracing Programming Exercise", `http://hdl.handle.net/1802/29150`, URResearch Archive, 2011.

5 Acknowledgments

Randal Nelson gave a critical reading. Tom Brown generously shared his course notes, but any mistakes or infelicities are mine — especially since he did not respond to my invitation to comment on this tutorial. Thanks to Tom as well for sharing his amusing and admirably comprehensive examples of how not to code in Matlab.