4. Lecture, 14 September 1999

4.1 Introduction to analytical ray tracing

Light travels in a straight line until it encounters a surface, at which point its direction – and in general its amplitude – is changed and it proceeds further. The prescription of an optical system consists of the specification of the positions and shapes of all of its surfaces, and the nature of the medium into which the rays pass after they leave each surface. As complicated as this specification process is, repeated application of three principles, plus coordinate transformations and other aspects of algebraic organization, are sufficient to follow a given ray, or bundle of rays, all the way through an optical system to a final focus. These principles are simple results of the physics of reflection and refraction from planar dielectric and conducting surfaces; the first two are of course Snell's law and the mirror reflection rule:

$$n_i \sin\theta_i = n_t \sin\theta_t \tag{4.1}$$

$$\theta_i = -\theta_r \tag{4.2}$$

The third result is that the wavevector of incident light, the surface normal at the point of incidence, and the wavevectors of reflected or refracted light, all lie in the same plane, a fact that we have used implicitly until now. Note that as an algebraic definition, plane mirror reflection is the same as Snell's law if we use $n_r = -n_i$ and a sign convention such that angles measured clockwise (counterclockwise) from the surface normal are negative (positive).

If one knows the shapes and positions of all of the optical surfaces, and specifies a bundle of rays coming from some object, one can use these rules to find the location and size of an image exactly, without resort to the small angle or thin lens approximations. The procedure is to take each ray, find where it intersects the first surface, apply the three rules to get the direction of the reflected or refracted ray, follow this new ray to the next surface, find the new intersection point, apply the rules again, *et cetera*, until the rays reach the final image. This would be a large number of simple calculations, which a human might perhaps find tedious; far better to do it on a computer. In the following, we will show how to set the ray-tracing problem up to facilitate computer calculation. The first set of equations we obtain will be used in the homework for some ray tracing "by hand," to get a feel for what the computer is doing.

Why go through all of this? Because we will discover in exact ray traces that *all* optical systems actually blur images to some degree. To produce the sharpest astronomical pictures possible, we need to understand the blurring process; exact ray tracing is used in this optimization process.

We start by composing a computer-friendly description of rays. Now, rays simply consist of a direction, once they start off from some given point, and the directions are conventionally kept track of by using the angles, and the cosines thereof, that the ray makes with a Cartesian coordinate system. A ray, of course, keeps the same set of angles and direction cosines until it encounters the next surface. The usual setup is shown in Figure 4.1. If one considers the ray to be represented by a unit vector **ν**, it will have the following components in this coordinate system:

$$\boldsymbol{\nu} = \begin{bmatrix} \gamma \\ \delta \\ \varepsilon \end{bmatrix} = \begin{bmatrix} \cos\theta \\ \cos\phi \\ \cos\psi \end{bmatrix} \quad . \tag{5.3}$$

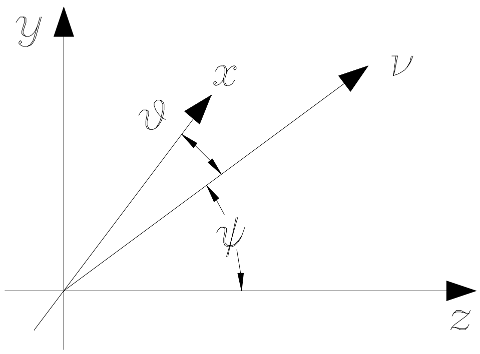

Figure 4.1: coordinate system for ray tracing.

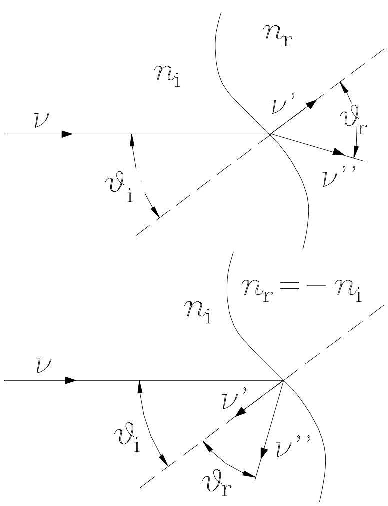

Figure 4.2: geometry of the input, surface normal and output unit vectors, for refraction (top) and reflection (bottom).

The components γ, δ, and ε are called the direction cosines of the ray. Note that since **ν** is a unit vector, $\sqrt{\gamma^2 + \delta^2 + \varepsilon^2} = 1$.

To find out what happens to this ray at the next surface, one needs to know where it intersects that surface, and at what angle, so that Snell's law or the mirror-reflection rule can be applied. The intersection can be determined easily if one has an analytical expression for the surface and the line that the input ray follows, in the same coordinate system. To find the incidence angle, one needs the direction of the unit vector normal to the surface at the intersection, $\boldsymbol{\nu}' = (\gamma' \quad \delta' \quad \varepsilon')$; this is perpendicular to the gradient of the surface at the intersection point. By convention, $\boldsymbol{\nu}'$ points toward the medium in which the output light will propagate, as shown in Figure 4.2. The incidence angle, and thus the angle of reflection or refraction, can be had from

$$\begin{aligned} &\cos\theta_i = \boldsymbol{\nu}\cdot\boldsymbol{\nu}' = \gamma\gamma' + \delta\delta' + \varepsilon\varepsilon' \\ \text{or} \quad &\theta_i = \cos^{-1}(\gamma\gamma' + \delta\delta' + \varepsilon\varepsilon') \quad , \\ \text{and} \quad &\theta_r = \sin^{-1}\left[\frac{n_i}{n_r}\sin\left(\cos^{-1}[\gamma\gamma' + \delta\delta' + \varepsilon\varepsilon']\right)\right] \quad , \end{aligned} \tag{4.4}$$

where Snell's law has been used in the last step. Note that the two possibilities for orientation of $\boldsymbol{\nu}'$ involve equal values for each direction cosine, but opposite signs. The proper orientation can be obtained by choosing the combination that gives $\cos\theta_i$ in Equation 4.4 the same sign as the ratio of n_r / n_i, if one uses $n_r = -n_i$ to describe the medium "behind" a reflecting surface. (As we've mentioned, this latter step also turns Snell's law into the mirror reflection rule.)

The resulting ray goes off in the direction $\boldsymbol{\nu}'' = (\gamma'' \quad \delta'' \quad \varepsilon'')$, and we therefore need three equations to solve for the three direction cosines. Since we know θ_i and θ_r, we have three independent equations in γ'', δ'' and ε'' from the inner products involving $\boldsymbol{\nu}''$ in Figure 4.2, and from the condition that all three unit vectors lie in a plane; these are, respectively,

$$\begin{aligned} \boldsymbol{\nu}\cdot\boldsymbol{\nu}'' &= \cos(\theta_i - \theta_r) \quad \text{(refraction)} \\ &= -\cos(2\theta_r) \quad \text{(reflection)} \\ \boldsymbol{\nu}'\cdot\boldsymbol{\nu}'' &= \cos\theta_r \\ (\boldsymbol{\nu}\times\boldsymbol{\nu}')\cdot\boldsymbol{\nu}'' &= 0 \end{aligned} \quad , \tag{4.5}$$

with θ_i and θ_r given by Equations 4.4. Including the direction cosines explicitly, and referring to Figure 4.2, we obtain

$$
\begin{aligned}
\gamma\gamma'' + \delta\delta'' + \varepsilon\varepsilon'' &= \cos(\theta_i - \theta_r) \text{ or } -\cos(2\theta_r) \\
\gamma'\gamma'' + \delta'\delta'' + \varepsilon'\varepsilon'' &= \cos\theta_r \\
(\varepsilon\delta' - \varepsilon'\delta)\gamma'' + (\gamma\varepsilon' - \gamma'\varepsilon)\delta'' + (\delta\gamma' - \delta'\gamma)\varepsilon'' &= 0
\end{aligned}
\tag{4.6}
$$

which can be expressed in matrix form as follows:

$$
\begin{bmatrix} \gamma & \delta & \varepsilon \\ \gamma' & \delta' & \varepsilon' \\ \varepsilon\delta' - \varepsilon'\delta & \gamma\varepsilon' - \gamma'\varepsilon & \delta\gamma' - \delta'\gamma \end{bmatrix}
\begin{bmatrix} \gamma'' \\ \delta'' \\ \varepsilon'' \end{bmatrix}
=
\begin{bmatrix} \cos(\theta_i - \theta_r) \text{ or } -\cos(2\theta_r) \\ \cos\theta_r \\ 0 \end{bmatrix}
\tag{4.7}
$$

or

$$
\ddot{\boldsymbol{\alpha}} \cdot \mathbf{v}'' = \boldsymbol{\beta}
$$

where the matrices $\ddot{\boldsymbol{\alpha}}$ and $\boldsymbol{\beta}$ are known. This can easily be solved for the components of $\mathbf{v}''$ in closed form, but it's usually easier simply to tell the computer how to solve the matrix equation; there are many fast numerical methods for doing so. Since only a 3×3 matrix in involved, even Cramer's rule can be used efficiently, with which

$$
\gamma'' = \frac{|\ddot{\boldsymbol{\alpha}}_1|}{|\ddot{\boldsymbol{\alpha}}|} \qquad \delta'' = \frac{|\ddot{\boldsymbol{\alpha}}_2|}{|\ddot{\boldsymbol{\alpha}}|} \qquad \varepsilon'' = \frac{|\ddot{\boldsymbol{\alpha}}_3|}{|\ddot{\boldsymbol{\alpha}}|} \quad , \tag{4.8}
$$

where $|\ldots|$ means a determinant is taken, $\ddot{\boldsymbol{\alpha}}_n$ is the matrix generated by replacing the *n*th column of $\ddot{\boldsymbol{\alpha}}$ with the corresponding components of the vector $\boldsymbol{\beta}$, and where it is assumed that $\ddot{\boldsymbol{\alpha}}$ has a nonvanishing determinant.

One way or another, it is simple for the computer to solve for $\mathbf{v}''$. It can then use this vector as the new $\mathbf{v}$, and the process repeats at the next surface.

Example 4.1
Consider two parallel rays incident on a concave paraboloidal mirror with focal length f; with one ray traveling along the paraboloid's symmetry axis and the other a distance x_0 away. The two rays reflect and intersect. Show by tracing the rays that they intersect on the symmetry axis, f from the apex. (We have already proven this to be true, in section 3.2.)

An appropriate coordinate system is shown in Figure 4.3. The unit vectors for the incident rays are the same, since they travel the same direction, and in this system are given by

$$
\mathbf{v} = \begin{bmatrix} \gamma \\ \delta \\ \varepsilon \end{bmatrix} = \begin{bmatrix} \cos\theta \\ \cos\phi \\ \cos\psi \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad . \tag{4.9}
$$

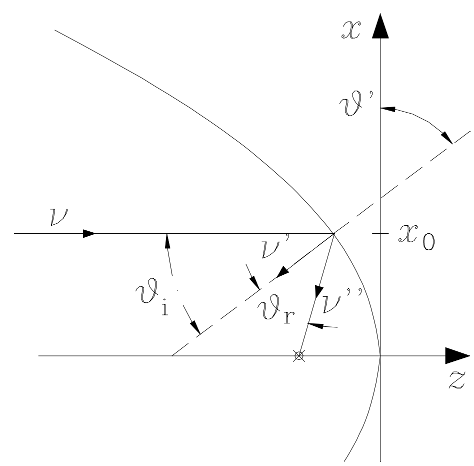

Figure 4.3: geometry of Example 4.1.

The paraboloid is described by

$$
z(x) = -\frac{x^2}{4f} \quad , \tag{4.10}
$$

so the slope is

$$\frac{dz}{dx}(x) = -\frac{x}{2f} \quad , \tag{4.11}$$

and the slope of the paraboloid's normal is

$$m = -\left(\frac{dz}{dx}(x)\right)^{-1} = \frac{2f}{x} \quad . \tag{4.12}$$

Since we have calculated slopes in such a way that $m = \tan\theta$, where θ as usual is the angle with respect to the x-axis, and since by convention the surface normal should point toward the output medium, the direction cosines of the normal unit vector $\mathbf{v}'$ are given from

$$\begin{aligned} \tan\theta' &= \frac{2f}{x} = \frac{\sin\theta'}{\cos\theta'} \\ \cos\theta' &= \gamma' = -\frac{x}{\sqrt{4f^2 + x^2}} \\ \sin\theta' &= \cos\psi' = \varepsilon' = -\frac{2f}{\sqrt{4f^2 + x^2}} \quad . \end{aligned} \tag{4.13}$$

Also, $\cos\varphi' = \delta' = 0$, since the whole figure lies in the x-z plane:

$$\mathbf{v}' = \begin{bmatrix} -x/\sqrt{4f^2 + x^2} \\ 0 \\ -2f/\sqrt{4f^2 + x^2} \end{bmatrix} \quad . \tag{4.14}$$

The incidence angle is given from

$$\cos\theta_i = \mathbf{v}\cdot\mathbf{v}' = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -x/\sqrt{4f^2 + x^2} \\ 0 \\ -2f/\sqrt{4f^2 + x^2} \end{bmatrix} = -\frac{2f}{\sqrt{4f^2 + x^2}} \quad . \tag{4.15}$$

Now, since $\theta_r = -\theta_i$, we know that the magnitudes of $\cos\theta_r$ and $\cos\theta_i$ must be the same. However, $\cos\theta_r$ must have the same sign as $\mathbf{v}\cdot\mathbf{v}''$, which is opposite the sign of $\mathbf{v}\cdot\mathbf{v}'$, so we must have

$$\begin{aligned} \cos\theta_r &= -\cos\theta_i = \frac{2f}{\sqrt{4f^2 + x^2}} \\ -\cos 2\theta_r &= -\left(\cos^2\theta_i - \sin^2\theta_i\right) = 1 - 2\cos^2\theta_i \\ &= 1 - \frac{8f^2}{4f^2 + x^2} \end{aligned} \tag{4.16}$$

Thus the matrix equation becomes

$$\begin{bmatrix} 0 & 0 & 1 \\ -x/\sqrt{4f^2+x^2} & 0 & -2f/\sqrt{4f^2+x^2} \\ 0 & x\big/\sqrt{4f^2+x^2} & 0 \end{bmatrix}\begin{bmatrix} \gamma'' \\ \delta'' \\ \varepsilon'' \end{bmatrix} = \begin{bmatrix} 1-8f^2/\left(4f^2+x^2\right) \\ 2f/\sqrt{4f^2+x^2} \\ 0 \end{bmatrix} \quad . \tag{4.17}$$

This can be solved with Cramer's Rule, but it's easier in this case to multiply out the component equations, though, for which we get

$$\varepsilon'' = 1 - \frac{8f^2}{4f^2+x^2} \quad , \tag{4.18}$$

$$\frac{\gamma'' x}{\sqrt{4f^2+x^2}} + \frac{\varepsilon'' 2f}{\sqrt{4f^2+x^2}} = \frac{-2f}{\sqrt{4f^2+x^2}} \quad , \tag{4.19}$$

and

$$\frac{x}{\sqrt{4f^2+x^2}}\delta'' = 0 \quad . \tag{4.20}$$

The last of these equations gives $\delta'' = 0$, unsurprisingly. Equation 4.19 can be combined with 4.18 to give

$$\gamma'' = \frac{4f}{x}\left(\frac{4f^2}{4f^2+x^2} - 1\right) \quad . \tag{4.21}$$

As you'll note, this satisfies $\varepsilon''^2 + \gamma''^2 = 1$, as it must. So, for the ray incident upon the mirror's apex,

$$\varepsilon'' = -1 \quad , \qquad \gamma'' = 0 \quad , \tag{4.22}$$

where l'Hôpital's rule, or the summing to unity of the squares of the direction cosines, is used because of the indeterminacy of the latter expression at $x = 0$. As expected, this ray reflects back along the mirror's symmetry axis. At $x = x_0$,

$$m'' = \frac{dz}{dx} = \frac{\varepsilon''}{\gamma''} = \frac{1 - \dfrac{8f^2}{4f^2+x_0^2}}{\dfrac{4f}{x_0}\left(\dfrac{4f^2}{4f^2+x_0^2} - 1\right)} = \frac{1}{x_0}\left(f - \frac{x_0^2}{4f}\right) \quad , \tag{4.23}$$

so the line along which the output ray propagates is

$$\begin{aligned} z - z_0 &= m''(x - x_0) \\ \text{or} \qquad z + \frac{x_0^2}{4f} &= \frac{1}{x_0}\left(f - \frac{x_0^2}{4f}\right)(x - x_0) \quad , \end{aligned} \tag{4.24}$$

and it duly intersects the symmetry axis ($x = 0$) at $z = -f$.