Chapter 1

GEOMETRICAL OPTICS

Electromagnetics problems are analytically intractable for all but the simplest situations. We know that any solution must satisfy the Helmholtz wave equation in a simple medium, but this offers marginal analytical insight into complicated propagation problems. This is particularly true if we start looking at inhomogeneous media – propagation through regions with varied permittivity and permeability. However, when the wavelength of radiation becomes very small relative to the objects, it is possible to invoke *geometrical optics* approximations and greatly enhance our ability to solve complicated wave propagation problems.

Geometrical optics allows engineers to construct wave propagation solutions with simple geometrical constructs called *rays*. The rays of geometrical optics trace out the paths of waves as they travel through an environment. In complicated wavescapes, this tracing of rays is conducted by computer programs. In fact, ray-tracing for radio wave propagation prediction has become a staple in certain areas of wireless communications. Cellular phone providers will often employ ray-tracing tools, for example, to predict signal strength in heavily urbanized areas to maximize coverage and service quality.

This chapter begins with a basic exposition of geometrical optics theory. After the geometrical concepts of rays have been formalized, the presentation shifts to computer algorithms and practical applications of ray theory.

1.1 Ray Representation of Waves

Geometrical optics requires a high-frequency approximation to the wave equation. This section describes the approximation and illustrates how to solve the simplified wave equation.

1.1.1 Approximating the Wave Equation

Let us start by making a slight modification to the Helmholtz wave equation to account for permittivities and permeabilities that vary with space. The modified Helmholtz wave equation is

$$(\nabla^2 + k_o^2\, n^2(\vec{\mathrm{r}})) \begin{pmatrix} \tilde{\vec{\mathsf{H}}} \\ \tilde{\vec{\mathsf{E}}} \end{pmatrix} = \mathbf{0} \tag{1.1.1}$$

where k_o is the wavelength in free space and the function $n(\vec{\mathrm{r}})$ is the *index of refraction* of the medium, defined to be

$$n(\vec{\mathrm{r}}) = \frac{\text{Speed of Light in Medium at } \vec{\mathrm{r}}}{\text{Speed of Light in Vacuum}} = \sqrt{\frac{\epsilon(\vec{\mathrm{r}})\,\mu(\vec{\mathrm{r}})}{\epsilon_0\,\mu_0}} \tag{1.1.2}$$

which is perfectly valid for source-free, linear, isotropic, non-conducting media. In subsequent equations, the positional dependence of n on $\vec{\mathrm{r}}$ will be implied but not written. Note that many physics and optics books will write $n = \sqrt{\epsilon_r(\vec{\mathrm{r}})}$, leaving unspoken the assumption that the medium is non-magnetic ($\mu(\vec{\mathrm{r}}) = \mu_0$, as it nearly always is for optical frequencies).

We will now solve the Helmholtz wave equation for a single field component, $\tilde{E}_\diamond(\vec{\mathrm{r}})$, recognizing that the solution could be generalized to all electric and magnetic field components. Furthermore, let us force our solution to be expressed in the following form:

$$\tilde{E}_\diamond(\vec{\mathrm{r}}) = E_\diamond(\vec{\mathrm{r}}) \exp\left(-jk_o S(\vec{\mathrm{r}})\right) \tag{1.1.3}$$

$E_\diamond$ is the amplitude of the wave and S is the phase of the wave, both of which are real functions of position $\vec{\mathrm{r}}$; S is also called the *eikonal function* and will track the direction of propagation. In fact, the vector field formed by the gradient of the eikonal, $\nabla S(\vec{\mathrm{r}})$, will represent the direction of propagation. As we will see, the streamlines of this field are actually the *rays* of geometrical optics.

Now insert Equation (1.1.3) into the wave equation. After some basic vector calculus operations, simplification, and rearrangement, the following results:

$$\underbrace{\left[\nabla^2 E_\diamond + k_o^2 n^2 E_\diamond - E_\diamond k_o^2 \|\nabla S\|^2\right]}_{\text{Term 1}} - j\, \underbrace{\left[k_o E_\diamond \nabla^2 S + 2k_o \nabla E_\diamond \cdot \nabla S\right]}_{\text{Term 2}} = 0 \tag{1.1.4}$$

This result has been grouped into real and imaginary portions, both of which must evaluate to zero if $E_\diamond$ and S are to form a Maxwellian solution. The equation resulting from the first term of this solution is written below:

$$\|\nabla S\|^2 = \frac{\nabla^2 E_\diamond}{k_o^2 E_\diamond} + n^2$$

Note that, for very high frequencies (and, thus, small wavelengths) this equation can be approximate as

$$|\nabla S(\vec{r})| = n(\vec{r}) \qquad \text{for } \left|\frac{\nabla^2 E_\diamond}{k_o^2 E_\diamond}\right| \ll n^2 \tag{1.1.5}$$

Equation (1.1.5) is often called the *eikonal equation* or the *ray-tracing equation*. Many of the waves that we have already study satisfy this equation, such as the plane wave in Example 1.1.

The term $\left|\frac{\nabla^2 E_\diamond}{k_o^2 E_\diamond}\right|$ in Equation (1.1.5) can be viewed as a diffraction factor. It indicates how much a wave solution will deviate outside of the geometrical optics prediction. Note that high frequency or high k_o is only part of minimizing this term; the relative change in the amplitude $E_\diamond$ over space is another factor. This is why rough features and sharp boundaries always pose problems in geometrical optics predictions.

Example 1.1: Plane Wave Eikonal

Problem: Show that a plane wave in a simple, homogeneous medium solves the eikonal equation.

Solution: Starting with the electric field component of a plane wave:

$$\tilde{\vec{\mathsf{E}}}(\vec{r}) = E_0 \hat{e} \exp(j[\phi - k\hat{k}\cdot\vec{r}])$$

We will show compliance with the x component of $\tilde{\vec{\mathsf{E}}}$, since the other components are identical in form:

$$\hat{x}\cdot\tilde{\vec{\mathsf{E}}}(\vec{r}) = E_x \exp(j\underbrace{[\phi - k\hat{k}\cdot\vec{r}]}_{-k_o S(\vec{r})}) \quad \longrightarrow \quad S(\vec{r}) = -\frac{\phi}{k_o} + \frac{k}{k_o}\hat{k}\cdot\vec{r}$$

Now take the gradient of the eikonal:

$$|\nabla S(\vec{r})| = \left|\nabla\left[-\frac{\phi}{k_o} + \frac{k}{k_o}\hat{k}\cdot\vec{r}\right]\right| = \frac{k}{k_o}\underbrace{|\nabla(\hat{k}\cdot\vec{r})|}_{=1} = \frac{\sqrt{\epsilon_r\mu_r}k_o}{k_o} = \underbrace{\sqrt{\epsilon_r\mu_r}}_{n}$$

which clearly satisfies the eikonal equation.

1.1.2 Solving the Eikonal Equation with Rays

Although the geometrical concept of describing a *ray* is fairly intuitive, it is much more difficult to attach a formal mathematical description. A ray is an intrinsically one-dimensional construct, but may bend in an arbitrary manner through three-dimensional space. Straight rays are fairly easy to describe with a vector function $\vec{\varrho}(s)$, where the components of $\vec{\varrho}$ are the Cartesian coordinates of points along the ray while s is a unit of length measured from the starting point of the ray. If a straight ray starts at the point $\vec{\varrho}(0)$, then its position as a function of distance may be written as

$$\vec{\varrho}(s) = \vec{\varrho}(0) + s\hat{\mathrm{t}} \tag{1.1.6}$$

where $\hat{\mathrm{t}}$ is a unit vector that points in the direction of ray travel.

For rays that curve, we must extend our formulation to account for a $\hat{\mathrm{t}}$ that varies with position s along the ray. We will define a general ray with the collection of positions $\vec{\varrho}$ that follow the general system of equations:

$$\frac{d\vec{\varrho}(s)}{ds} = \hat{\mathrm{t}} \qquad \frac{d^2\vec{\varrho}(s)}{ds^2} = \kappa\hat{\mathrm{u}} \qquad \hat{\mathrm{t}}\cdot\hat{\mathrm{u}} = 0 \text{ for all } s \tag{1.1.7}$$

According to Equation (1.1.7), there is a unit vector $\hat{\mathrm{t}}$ that is always tangent to the path of the ray. The first derivative of $\vec{\varrho}$ with respect to ray length must, by definition, point along this unit vector and have length of 1. Therefore $\frac{d\vec{\varrho}}{ds}$ is allowed to change directions as s increases, but its magnitude must be a constant 1 in order for s to be a natural measure of length in space. The second derivative of $\vec{\varrho}$ with respect to s, which is $\frac{d^2\vec{\varrho}}{ds^2}$ or $\frac{d\hat{\mathrm{t}}}{ds}$, is therefore allowed to be non-zero, but must point in a direction $\hat{\mathrm{u}}$ that is perpendicular to the direction of ray travel $\hat{\mathrm{t}}$. The magnitude of this second derivative is related to the *curvature* κ of the ray at a given point. The relationships of Equation (1.1.7) are illustrated in Figure 1.1.

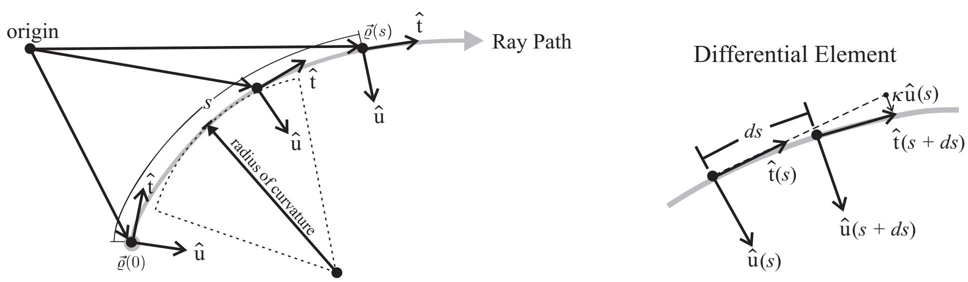

Figure 1.1. Geometry of a traced ray. The path of the ray is mapped by the vector function $\varrho(s)$, where s is distance from the start of the ray. $\hat{\mathrm{t}}$ and $\hat{\mathrm{u}}$ are unit vectors that point tangential and normal to the ray path, respectively.

After establishing the geometrical properties of the ray, it is now time to apply wave-related conditions. If the rays are to trace waves, then the direction of ray travel $\hat{t}$ must point along the gradient of the eikonal function S:

$$\hat{t} = \frac{\nabla S}{\|\nabla S\|} \quad \longrightarrow \quad n\frac{d\vec{\varrho}(s)}{ds} = \nabla S \tag{1.1.8}$$

which is modified by noting that $\|\nabla S\| = n$, according to Equation (1.1.5). Now take the derivative of both sides of Equation (1.1.8) with respect to s:

$$\underbrace{\frac{dn}{ds}}_{\nabla n\cdot\hat{t}}\frac{d\vec{\varrho}(s)}{ds} + n\underbrace{\frac{d^2\vec{\varrho}(s)}{ds^2}}_{\kappa\hat{u}} = \underbrace{\frac{d\nabla S}{ds}}_{\nabla\|\nabla S\|} \tag{1.1.9}$$

which reduces to

$$(\nabla n \cdot \hat{t})\hat{t} + (n\kappa)\hat{u} = \nabla n \tag{1.1.10}$$

As seen from Equation (1.1.10), the curvature κ of the ray is proportional to the portion of ∇n that projects *transverse* to the direction of ray travel $\hat{t}$. A ray will always curve into the direction of the index of refraction gradient. If the spatial change of n is perfectly aligned with the direction of ray travel $\hat{t}$, then the curvature is 0 and the ray travels in a straight line. Maximum curvature of the ray is achieved when the gradient of n is perpendicular to the direction of travel $\hat{t}$.

We may calculate the ray curvature κ and the bending direction $\hat{u}$ with the following formulas:

$$\kappa = \frac{1}{n}\left\|\nabla n - (\nabla n \cdot \hat{t})\hat{t}\right\| \qquad \hat{u} = \hat{t} \times (\nabla n \times \hat{t}) \tag{1.1.11}$$

One of the most important principles of Equation (1.1.11) is that *waves are traced by straight rays in homogeneous media.* In such media, the gradient of n is always the zero-vector. This greatly simplifies many of our calculations for radio wave propagation and free-space optics.

1.1.3 Calculating Amplitudes of Rays

Tracing the path of a ray is only part of constructing a geometrical optics solution. After the path has been traced, we must calculate the wave's amplitude at points of interest along the path. To accomplish this, start with the second, imaginary term of Equation (1.1.4). This term must also be equal to zero, resulting in the following differential equation:

$$\underbrace{\nabla E_\diamond \cdot \nabla S}_{\|\nabla S\|\frac{dE_\diamond}{ds}} = -\frac{1}{2}E_\diamond \nabla^2 S \tag{1.1.12}$$

Grouping similar terms and integrating produces

$$\int_{E_\diamond(s_o)}^{E_\diamond(s+s_o)} \frac{dE_\diamond}{E_\diamond} = -\int_{s_o}^{s_o+s} \frac{\nabla^2 S}{2\|\nabla S\|}\, ds' \tag{1.1.13}$$

which simplifies to

$$E_\diamond^2(s+s_o) = E_\diamond^2(s_o)\exp\left(-\int_{s_o}^{s_o+s} \frac{\nabla^2 S}{\|\nabla S\|}\, ds'\right) \tag{1.1.14}$$

This equation relates amplitudes of $E_\diamond$ to one another at different points along the ray. Thus, if the amplitude at s_o is known, Equation (1.1.14) may be used to project the amplitude to any new location s along the ray. The difference in amplitude at those two points is a function of how the eikonal S has changed. In fact, there is a direct relationship between the *shape* of S and the *power* of the wave, since it is the square of the wave amplitude (which is proportional to power) that appears in Equation (1.1.14). The right-hand integral in Equation (1.1.14) measures the *Gaussian curvature* of a surface [Cor94]. Gaussian curvature, κ_G, is defined as the inverse product of two radii of curvature, ρ_1 and ρ_2, as measured in orthogonal directions at a specific point on the surface. This concept is illustrated in Figure 1.2.

In a simple medium, where the rays are straight lines, the formula for mapping the curvature of equi-phase surfaces is fairly straight forward. As seen in Figure 1.2, the Gaussian curvatures at points $\mathcal{A}$ and $\mathcal{B}$ in the diagram are

$$\kappa_{G\mathcal{A}} = \frac{1}{\rho_1\rho_2} \qquad \kappa_{G\mathcal{B}} = \frac{1}{(\rho_1+s)(\rho_2+s)} \tag{1.1.15}$$

When moving s units from point $\mathcal{A}$ to point $\mathcal{B}$, the eikonal surface increases its radii of curvature each by a length of R as well. Thus, as rays propagate, their overall Gaussian curvature decreases (see Example 1.2). Note that, within this framework, it is perfectly acceptable to have a *negative* initial radius of curvature in one or both principle directions along the eikonal surface; this would correspond to a wavefront that was concave with respect to the direction of propagation – a case that arises often in optical propagation when lenses are involved.

Example 1.2: Gaussian Curvature of an Eikonal Surface

Problem: Demonstrate that the integral in Equation (1.1.14) represents the ratio of Gaussian curvatures of an eikonal surface at points $\mathcal{A}$ and $\mathcal{B}$ in a simple medium.

Solution: Let us define the unit vector ray field $\vec{\varrho}(s,u,v)$ where s corresponds to movement along the ray and u and v represent movement along an equi-phase surface and, consequently, jumping to different rays. In a homogeneous medium, we recognize the identity $\hat{\mathrm{t}} = \nabla S/n$. The divergence of $\hat{\mathrm{t}}$ may then be broken into constitutive operations:

$$\nabla\cdot\hat{\mathrm{t}} = \underbrace{\nabla_t\cdot\hat{\mathrm{t}}}_{=0} + \underbrace{\nabla_u\cdot\hat{\mathrm{t}}}_{(\rho_1+s)^{-1}} + \underbrace{\nabla_v\cdot\hat{\mathrm{t}}}_{(\rho_2+s)^{-1}} \qquad \text{where } \nabla_x\cdot\hat{\mathrm{t}} = \left(\hat{\mathrm{x}}\cdot\frac{\partial\hat{\mathrm{t}}}{\partial x}\right)$$

where the first term is identically equal to 0, since a unit vector in a ray field cannot change size. The second and third substitutions are a result of straight rays in a homogeneous medium. If we say that point $\mathcal{A}$ lies at $s = 0$, and that the principle radii of curvature in the directions $\hat{\mathrm{u}}$ and $\hat{\mathrm{v}}$ are known to be ρ_1 and ρ_2, respectively, then their directional divergences are inversely proportional to these radii plus any accumulated ray distance, s. This basic property stems directly from the fact that straight rays continue to diverge in space inversely proportional to the distance they travel. The rest of the proof-sketch follows mechanically from these facts:

$$\begin{aligned}
\exp\left(-\int_0^s \underbrace{\nabla\cdot\hat{\mathrm{t}}}_{\nabla^2 S/n}\, ds'\right) &= \exp\left(-\int_0^s \frac{ds'}{\rho_1+s'}\right)\exp\left(-\int_0^s \frac{ds'}{\rho_2+s'}\right)\\
&= \exp\left(-\ln(\rho_1+s')\Big|_0^s\right)\exp\left(-\ln(\rho_2+s')\Big|_0^s\right)\\
&= \left(\frac{\rho_1}{\rho_1+s}\right)\left(\frac{\rho_2}{\rho_2+s}\right) = \frac{\frac{1}{(\rho_1+s)(\rho_2+s)}}{\frac{1}{\rho_1\rho_2}} = \frac{\kappa_{G\mathcal{B}}}{\kappa_{G\mathcal{A}}}
\end{aligned}$$

where point $\mathcal{B}$ is s units down the ray from $\mathcal{A}$. For a one-dimensional *path* in two-dimensional space, the curvature κ at a point is inversely proportional to the radius of curvature. For a two-dimensional *surface* in three-dimensional space, the Gaussian curvature κ_G is defined as the product of two radii of curvature in principle directions, which appears in our final result.

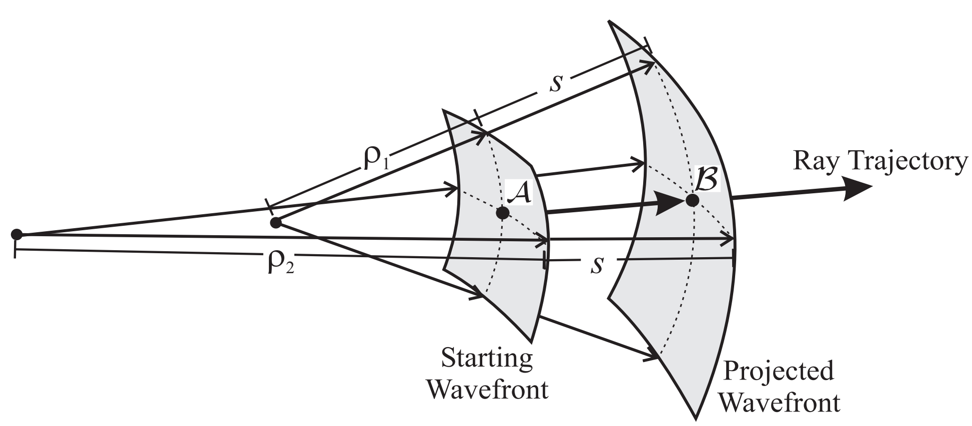

Figure 1.2. Geometrical interpretation of Gaussian curvature for a wavefront. The distances ρ_1 and ρ_2 are the principle radii of curvature for the surface at point $\mathcal{A}$. The radii of curvature change as the wavefront is translated to point $\mathcal{B}$.

Therefore, the full electric and magnetic vector field solution may be synthesized from a trajectory and a set of initial amplitude conditions. In a simple medium, we may write the solution as

$$\begin{aligned} \tilde{\vec{\mathsf{E}}}(s) &= \tilde{E}_o \hat{\mathrm{e}} \exp(-jks) \sqrt{\tfrac{\rho_1 \rho_2}{(\rho_1+s)(\rho_2+s)}} \\ \tilde{\vec{\mathsf{H}}}(s) &= \tfrac{\tilde{E}_o}{\eta} (\hat{\mathrm{t}} \times \hat{\mathrm{e}})^* \exp(-jks) \sqrt{\tfrac{\rho_1 \rho_2}{(\rho_1+s)(\rho_2+s)}} \end{aligned} \tag{1.1.16}$$

where $\tilde{E}_o$ is the amplitude and phase at $s = 0$ and ê is the polatization of the electric field at the initial wavefront.

Although the full calculation for Gaussian curvature looks daunting, it is particularly simple for the common forms of canonical waves. Consider three basic canonical waves in three-dimensional space: plane, spherical, and cylindrical waves. These cases, illustrated in Figure 1.3 with traced rays, have simple expressions for their Gaussian curvatures, summarized in Table 1.1. The power loss relative to an initial phase condition is summarized in this table.

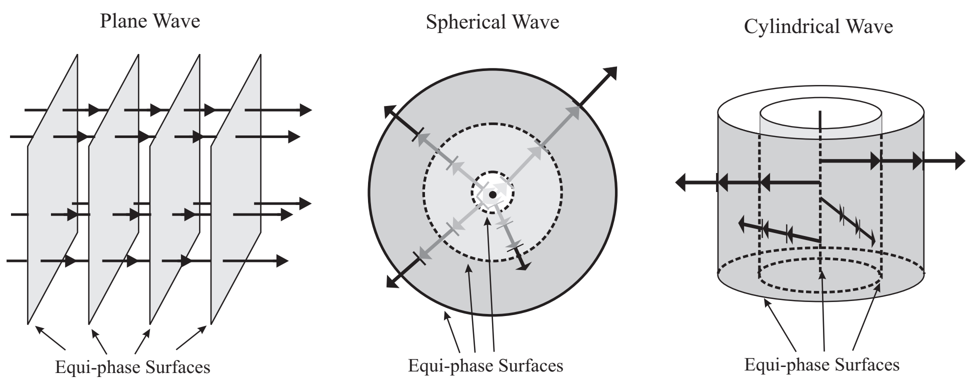

Figure 1.3. Examples of three types of canonical waves traced with rays.

The waves described in Figure 1.3 and Table 1.1 represent special cases of wave propagation where either the principle radii of wavefront curvature are equal (spherical wave) or at least one principle radius is infinite (cylindrical or plane wave). If both radii of curvature are finite, but dissimilar, the wave is said to be *astigmatic*. Astigmatic waves are particularly common in radio wave propagation when geometrical optics is extended to the *geometrical theory of diffraction*.

Note: Astigmatisms

Many people wear glasses or contact lenses to correct an *astigmatism*, a condition that arises when the eye's lens has dissimilar principle radii of curvature. Such a lens cannot focus an image cleanly. If you were to ray trace a plane wave's path through the lens, the rays that would pass through would become *astigmatic* ($\rho_1 \neq \rho_2$).

Table 1.1. Table for Gaussian curvature and path loss of 3 canonical waves relative to position s_o along a ray.

Canonical Wave Shape	Coordinate System	Example Equation	Laplacian of Eikonal, $\frac{\nabla^2 S}{n}$	Path Loss w.r.t. s_o
Planar	(x, y, z)	$E_\diamond \exp(-jnk_o(x-s_o))$	0	1
Cylindrical	(ρ, θ, z)	$\frac{E_\diamond}{\sqrt{\rho}} \exp(-jnk_o(\rho-s_o))$	$\frac{1}{\rho}$	$\frac{s_o}{s}$
Spherical	(r, ϕ, θ)	$\frac{E_\diamond}{r} \exp(-jnk_o(r-s_o))$	$\frac{2}{r}$	$\frac{s_o^2}{s^2}$