1

Scalar wave equations and diffraction of laser radiation

1.1 Introduction

Radiation from lasers is different from conventional optical light because, like microwave radiation, it is approximately monochromatic. Although each laser has its own fine spectral distribution and noise properties, the electric and magnetic fields from lasers are considered to have precise phase and amplitude variations in the first-order approximation. Like microwaves, electromagnetic radiation with a precise phase and amplitude is described most accurately by Maxwell's wave equations. For analysis of optical fields in structures such as optical waveguides and single-mode fibers, Maxwell's vector wave equations with appropriate boundary conditions are used. Such analyses are important and necessary for applications in which we need to know the detailed characteristics of the vector fields known as the modes of these structures. They will be discussed in Chapters 3 and 4.

For devices with structures that have dimensions very much larger than the wavelength, e.g. in a multimode fiber or in an optical system consisting of lenses, prisms or mirrors, the rigorous analysis of Maxwell's vector wave equations becomes very complex and tedious: there are too many modes in such a large space. It is difficult to solve Maxwell's vector wave equations for such cases, even with large computers. Even if we find the solution, it would contain fine features (such as the fringe fields near the lens) which are often of little or no significance to practical applications. In these cases we look for a simple analysis which can give us just the main features (i.e. the amplitude and phase) of the dominant component of the electromagnetic field in directions close to the direction of propagation and at distances reasonably far away from the aperture.

When one deals with laser radiation fields which have slow transverse variations and which interact with devices that have overall dimensions much larger than the optical wavelength λ, the fields can often be approximated as transverse electric and magnetic (TEM) waves. In TEM waves both the dominant electric field and the dominant magnetic field polarization lie approximately in the plane perpendicular to the direction of propagation. The polarization direction does not change substantially within a propagation distance comparable to wavelength. For such waves, we usually need only to solve the scalar wave equations to obtain the amplitude and the phase of the dominant electric field along its local polarization direction. The dominant magnetic field can be calculated directly from the dominant electric field. Alternatively, we can first solve the scalar equation of the dominant magnetic field, and the electric field can be calculated from the magnetic field. We have encountered TEM waves in undergraduate electromagnetic field courses usually as plane waves that have no transverse amplitude and phase variations. For TEM waves in general, we need a more sophisticated analysis than plane wave analysis to account for the transverse variations. Phase information for TEM waves is especially important for laser radiation because many applications, such as spatial filtering, holography and wavelength selection by grating, depend critically on the phase information.

The details with which we normally describe the TEM waves can be divided into two categories, depending on application. (1) When we analyze how laser radiation is diffracted, deflected or reflected by gratings, holograms or optical components with finite apertures, we calculate the phase and amplitude variations of the dominant transverse electric field. Examples include the diffraction of laser radiation in optical instruments, signal processing using laser light, or modes of solid state or gas lasers. (2) When we are only interested in the propagation velocity and the path of the TEM waves, we describe and analyze the optical beams only by reference to the path of such optical rays. Examples include modal dispersion in multimode fibers and lidars. The analyses of ray optics are fairly simple; they are discussed in many optics books and articles [1, 2]. They are also known as geometrical optics. They will not be presented in this book.

We will first learn what is meant by a scalar wave equation in Section 1.2. In Section 1.3, we will learn mathematically how the solution of the scalar wave equation by Green's function leads to the well known Kirchhoff diffraction integral solution. The mathematical derivations in these sections are important not only in order to present rigorously the theoretical optical analyses but also to allow us to appreciate the approximations and limitations implied in various results. Further approximations of Kirchhoff's integral then lead to the classical Fresnel and Fraunhofer diffraction integrals. Applications of Kirchhoff's integral are illustrated in Section 1.4.

Fraunhofer diffraction from an aperture at the far field demonstrates the classical analysis of diffraction. Although the intensity of the diffracted field is the primary concern of many conventional optics applications, we will emphasize both the amplitude and the phase of the diffracted field that are important for many laser applications. For example, Fraunhofer diffraction and Fourier transform relations at the focal plane of a lens provide the theoretical basis of spatial filtering. Spatial filtering techniques are employed frequently in optical instruments, in optical computing and in signal processing.

Understanding the origin of the integral equations for laser resonators is crucial in allowing us to comprehend the origin and the limitation of the Gaussian mode description of lasers. In Section 1.5, we will illustrate several applications of transformation techniques of Gaussian beams based on Kirchhoff's diffraction integral, which is valid for TEM laser radiation.

Please note that the information given in Sections 1.2, 1.3 and 1.4 is also presented extensively in classical optics books [3, 4, 5]. Readers are referred to those books for many other applications.

1.2 The scalar wave equation

The simplest way to understand why we can use a scalar wave equation is to consider Maxwell's vector wave equation in a sourceless homogeneous medium. It can be written in terms of the rectangular coordinates as

$$\nabla^2 \underline{E} - \frac{1}{c^2}\frac{\partial^2 \underline{E}}{\partial t^2} = 0,$$
$$\underline{E} = E_x \underline{i_x} + E_y \underline{i_y} + E_z \underline{i_z},$$

where c is the velocity of light in the homogeneous medium. If $\underline{E}$ has only one dominant component $E_x \underline{i_x}$, then E_y, E_z, and the unit vector $\underline{i_x}$ can be dropped from the above equation. The resultant equation is a scalar wave equation for E_x.

In short, for TEM waves, we usually describe the dominant electromagnetic (EM) field by a scalar function U. In a homogeneous medium, U satisfies the scalar wave equation

$$\nabla^2 U - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} U = 0. \tag{1.1}$$

In an elementary view, U is the instantaneous amplitude of the transverse electric field in its direction of polarization when the polarization is approximately constant (i.e. $|U|$ varies slowly within a distance comparable to the wavelength). From a different point of view, when we use the scalar wave equation, we have implicitly assumed that the curl equations in Maxwell's equations do not yield a sufficient magnitude of electric field components in other directions that will affect significantly the TEM characteristics of the field. The magnetic field is calculated directly from the dominant electric field. In books such as that by Born and Wolf [3], it is shown that U can also be considered as a scalar potential for the optical field. In that case, electric and magnetic fields can be derived from the scalar potential.

Both the scalar wave equation in Eq. (1.1) and the boundary conditions are derived from Maxwell's equations. The boundary conditions (i.e. the continuity of tangential electric and magnetic fields across the boundary) are replaced by boundary conditions of U (i.e. the continuity of U and normal derivative of U across the boundary). Notice that the only limitation imposed so far by this approach is that we can find the solution for the EM fields by just one electric field component (i.e. the scalar U). We will present further simplifications on how to solve Eq. (1.1) in Section 1.3.

For wave propagation in a complex environment, Eq. (1.1) can be considered as the equation for propagation of TEM waves in the local region when TEM approximation is acceptable. In order to obtain a global analysis of wave propagation in a complex environment, solutions obtained for adjacent local regions are then matched in both spatial and time variations at the boundary between adjacent local regions.

For monochromatic radiation with a harmonic time variation, we usually write

$$U(x, y, z; t) = U(x, y, z)e^{j\omega t}. \tag{1.2}$$

Here, $U(x, y, z)$ is complex, i.e. U has both amplitude and phase. Then U satisfies the Helmholtz equation,

$$\nabla^2 U + k^2 U = 0, \tag{1.3}$$

where $k = \omega/c = 2\pi/\lambda$ and $c =$ free space velocity of light $= 1/\sqrt{\varepsilon_0 \mu_0}$. The boundary conditions are the continuity of U and the normal derivative of U across the dielectric discontinuity boundary.

In this section, we have defined the equation governing U and discussed the approximations involved when we use it. In the first two chapters of this book, we will accept the scalar wave equation and learn how to solve for U in various applications of laser radiation.

We could always solve for U for each individual case as a boundary value problem. This would be the case when we solve the equation by numerical methods. However, we would also like to have an analytical expression for U in a homogeneous medium when its value is known at some boundary surface. The well known method used to obtain U in terms of its known value on some boundary is the Green's function method, which is derived and discussed in Section 1.3.

1.3 The solution of the scalar wave equation by Green's function – Kirchhoff's diffraction formula

Green's function is nothing more than a mathematical technique which facilitates the calculation of U at a given position in terms of the fields known at some remote boundary without explicitly solving the differential Eq. (1.4) for each individual case [3, 6]. In this section, we will learn how to do this mathematically. In the process we will learn the limitations and the approximations involved in such a method.

Let there be a Green's function G such that G is the solution of the equation

$$\begin{aligned}\nabla^2 G(x, y, z; x_0, y_0, z_0) + k^2 G &= -\,\delta(x - x_0, y - y_0, z - z_0)\\ &= -\,\delta(\underline{r} - \underline{r_0}).\end{aligned} \tag{1.4}$$

Equation (1.4) is identical to Eq. (1.3) except for the δ function. The boundary conditions for G are the same as those for U; δ is a unit impulse function which is zero when $x \neq x_0$, $y \neq y_0$ and $z \neq z_0$. It goes to infinity when (x, y, z) approaches the discontinuity point (x_0, y_0, z_0), and δ satisfies the normalization condition

$$\begin{aligned}&\iiint_V \delta(x - x_0, y - y_0, z - z_0)\, dx\, dy\, dz = 1\\ &\quad = \iiint_V \delta(\underline{r} - \underline{r_0})\, dv,\end{aligned} \tag{1.5}$$

where $\underline{r} = x\underline{i_x} + y\underline{i_y} + z\underline{i_z}$, $\underline{r_0} = x_0\underline{i_x} + y_0\underline{i_y} + z_0\underline{i_z}$ and $dv = dx\,dy\,dz = r^2 \sin\theta\, dr\, d\theta\, d\phi$. V is any volume including the point $(x_0,\ y_0,\ z_0)$. First we will show how a solution for G of Eq. (1.4) will let us find U at any given observer position $(x_0,\ y_0,\ z_0)$ from the U known at some distant boundary.

From advanced calculus [7],

$$\nabla \cdot (G\nabla U - U\nabla G) = G\nabla^2 U - U\nabla^2 G.$$

Applying a volume integral to both sides of the above equation and utilizing Eqs. (1.4) and (1.5), we obtain

$$\begin{aligned}&\iiint_V \nabla\cdot (G\nabla U - U\nabla G)\, dv\\ &\quad = \iint_S (G\underline{n}\cdot \nabla U - U\underline{n}\cdot \nabla G)\, ds\\ &\quad = \iiint_V \left[-k^2 GU + k^2 UG + U\delta(\underline{r} - \underline{r_0})\right] dv = U(\underline{r_0}).\end{aligned} \tag{1.6}$$

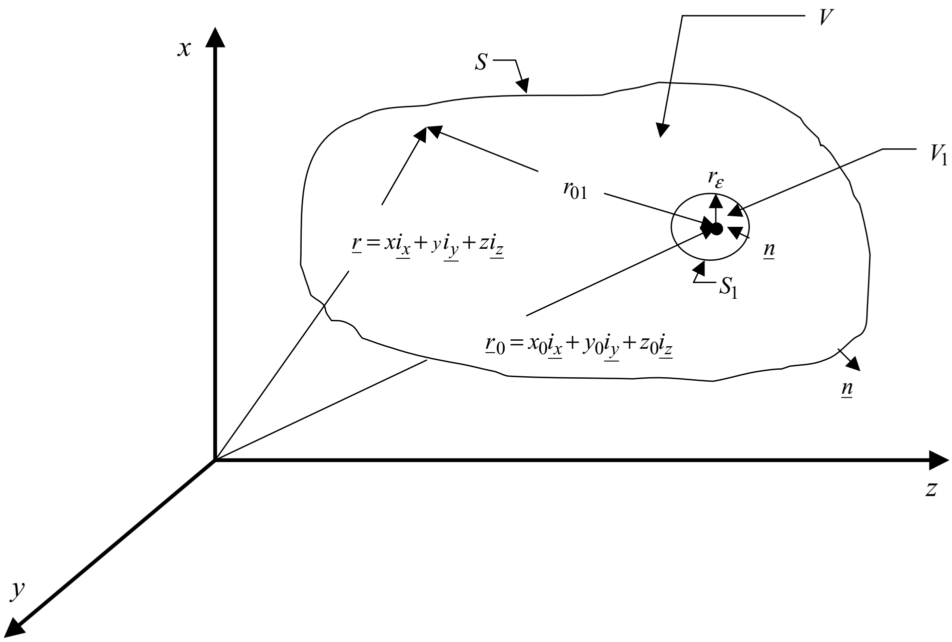

Figure 1.1. Illustration of volumes and surfaces to which Green's theory applies. The volume to which Green's function applies is V', which has a surface S. The outward unit vector of S is $\underline{n}$; $\underline{r}$ is any point in the x, y, z space. The observation point within V is $\underline{r_0}$. For the volume V', V_1 around $\underline{r_0}$ is subtracted from V. V_1 has surface S_1, and the unit vector $\underline{n}$ is pointed outward from V'.

V is any closed volume (within a boundary S) enclosing the observation point $\underline{r_0}$ and n is the unit vector perpendicular to the boundary in the outward direction, as illustrated in Fig. 1.1.

Equation (1.6) is an important mathematical result. It shows that, when G is known, the U at position (x_0, y_0, z_0) can be expressed directly in terms of the values of U and ∇U on the boundary S, without solving explicitly the Helmholtz equation, Eq. (1.3). Equation (1.6) is known mathematically as Green's identity. The key problem is how to find G.

Fortunately, G is well known in some special cases that are important in many applications. We will present three cases of G in the following.

1.3.1 The general Green's function G

The general Green's function G has been derived in many classical optics textbooks; see, for example, [3]:

$$G = \frac{1}{4\pi} \frac{\exp(-jkr_{01})}{r_{01}}, \tag{1.7}$$

where

$$r_{01} = |\underline{r_0} - \underline{r}| = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}.$$

As shown in Fig. 1.1, r_{01} is the distance between $\underline{r_0}$ and $\underline{r}$.

This G can be shown to satisfy Eq. (1.4) in two steps.

(1) By direct differentiation, $\nabla^2 G + k^2 G$ is clearly zero everywhere in any homogeneous medium except at $\underline{r} \approx \underline{r_0}$. Therefore, Eq. (1.4) is satisfied within the volume V', which is V minus V_1 (with boundary S_1) of a small sphere with radius r_ε enclosing $\underline{r_0}$ in the limit as r_ε approaches zero. V_1 and S_1 are also illustrated in Fig. 1.1.

(2) In order to find out the behavior of G near $\underline{r_0}$, we note that $|G| \to \infty$ as $r_{01} \to 0$. If we perform the volume integration of the left hand side of Eq. (1.4) over the volume V_1, we obtain:

$$\begin{aligned}\operatorname*{Lim}_{r_\varepsilon \to 0} \iiint_{V_1} [\nabla \cdot \nabla G + k^2 G]\, dv &= \iint_{S_1} \nabla G \cdot \underline{n}\, ds \\ &= \operatorname*{Lim}_{r_\varepsilon \to 0} \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} \left[-\frac{e^{-jkr_\varepsilon}}{4\pi\, r_\varepsilon^2} \right] r_\varepsilon^2 \sin\theta\, d\theta\, d\phi = -1.\end{aligned}$$

Thus, using this Green's function, the volume integration of the left hand side of Eq. (1.4) yields the same result as the volume integration of the δ function. In short, the G given in Eq. (1.7) satisfies Eq. (1.4) for any homogeneous medium.

From Eq. (1.6) and G, we obtain the well known Kirchhoff diffraction formula,

$$U(\underline{r_0}) = \iint_S (G \nabla U - U \nabla G) \cdot \underline{n}\, ds. \tag{1.8}$$

Note that we need only to know both U and ∇U on the boundary in order to calculate its value at $\underline{r_0}$ inside the boundary.

1.3.2 Green's function, G_1, for U known on a planar aperture

For many practical applications, U is known on a planar aperture, followed by a homogeneous medium with no additional radiation source. Let the planar aperture be the surface $z = 0$; a known radiation U is incident on the aperture Ω from $z < 0$, and the observation point is located at $z > 0$. As a mathematical approximation to this geometry, we define V to be the semi-infinite space at $z \geq 0$, bounded by the surface S. S consists of the plane $z = 0$ on the left and a large spherical surface with radius R on the right, as $R \to \infty$. Figure 1.2 illustrates the semi-sphere.

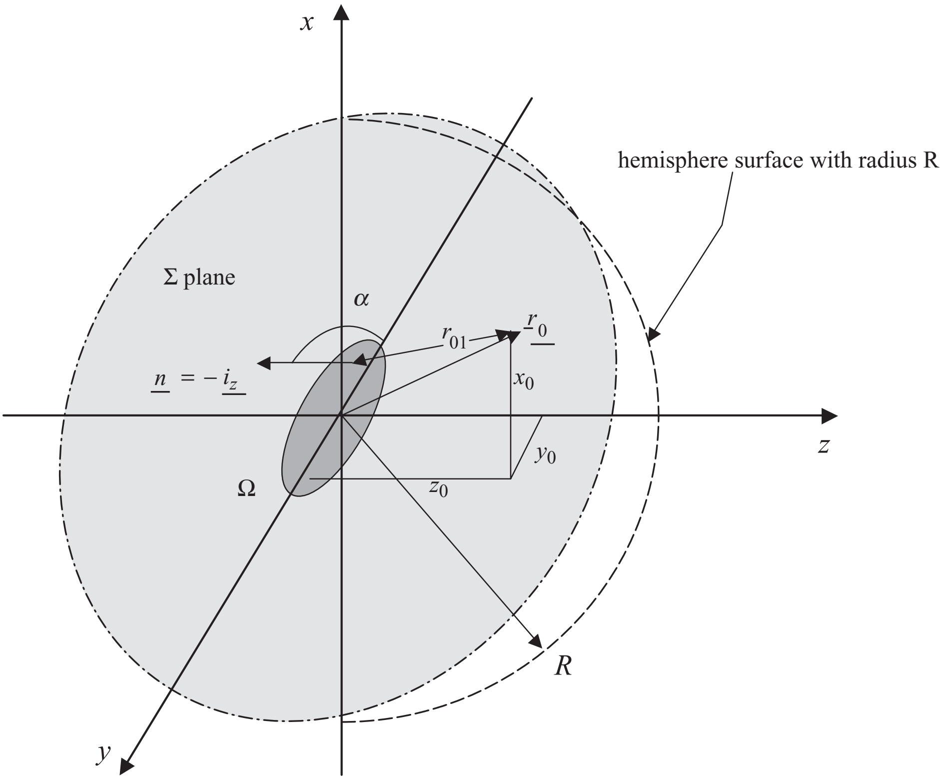

Figure 1.2. Geometrical configuration of the semi-spherical volume for the Green's function G_1. The surface to which the Green's function applies consists of Σ, which is part of the *xy* plane, and a very large hemisphere that has a radius *R*, connected with Σ. The incident radiation is incident on Ω, which is an open aperture within Σ. The outward normal of the surfaces Σ and Ω is $-i_z$. The coordinates for the observation point r_0 are x_0, y_0 and z_0.

The boundary condition for a sourceless U at $z > 0$ is given by the radiation condition at very large R; as $R \to \infty$ [8],

$$\lim_{R\to\infty} R\left(\frac{\partial U}{\partial n} + jkU\right) = 0. \tag{1.9}$$

The radiation condition is essentially a mathematical statement that there is no incoming wave at very large R. Any U which represents an outgoing wave in the $z > 0$ space will satisfy Eq. (1.9).

If we do not want to use the ∇U term in Eq. (1.8), we like to have a Green's function which is zero on the plane boundary (i.e. $z = 0$). Since we want to apply Eq. (1.8) to the semi-sphere boundary S, Eq. (1.4) needs to be satisfied only for $z > 0$. In order to find such a Green's function, we note first that any function F in the form $\exp(-jkr)/r$, expressed in Eq. (1.7), will satisfy $\lfloor \nabla F + k^2 F = 0 \rfloor$ as

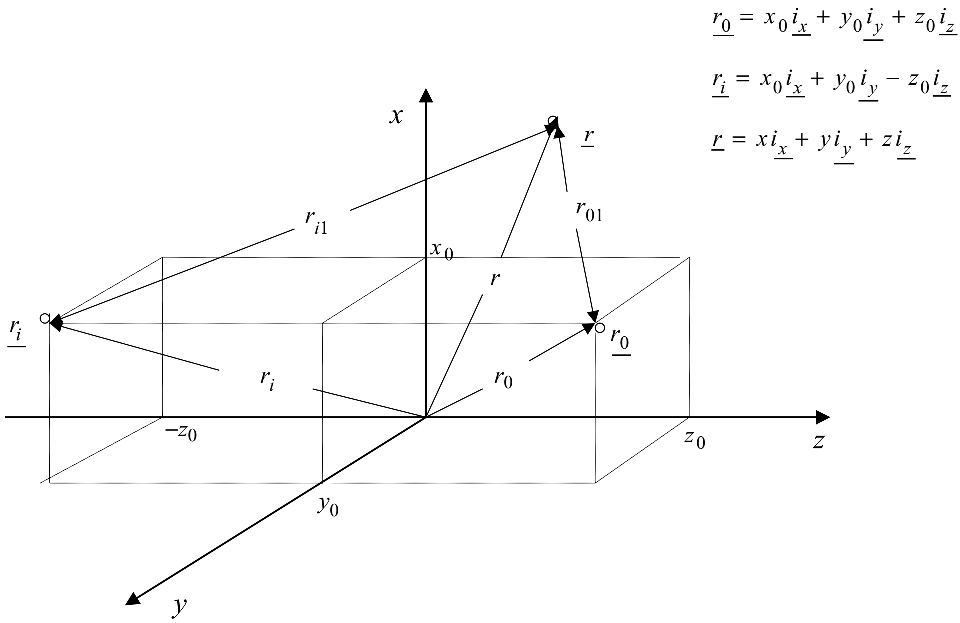

Figure 1.3. Illustration of $\underline{r}$, the point of observation $\underline{r_0}$ and its image $\underline{r_j}$, in the method of images. For G, the image plane Σ is the xy plane, and $\underline{r_i}$ is the image of the observation point $\underline{r_0}$ in Σ. The coordinates of $\underline{r_0}$ and $\underline{r_i}$ are given.

long as r is not allowed to approach zero. We can add such a second term to the G given in Eq. (1.7) and still satisfy Eq. (1.4) for $z > 0$ as long as r never approaches zero for $z > 0$. To be more specific, let $\underline{r_i}$ be a mirror image of (x_0, y_0, z_0) across the $z = 0$ plane at $z < 0$. Let the second term be $e^{-jkr_{i1}}/r_{i1}$, where r_{i1} is the distance between (x, y, z) and $\underline{r_i}$. Since our Green's function will only be used for $z_0 > 0$, the r_{i1} for this second term will never approach zero for $z \geq 0$. Thus, as long as we seek the solution of U in the space $z > 0$, Eq. (1.4) is satisfied for $z > 0$. However, the difference of the two terms is zero when (x, y, z) is on the $z = 0$ plane. This is known as the "method of images" in electromagnetic theory. Such a Green's function is constructed mathematically in the following.

Let the Green's function for this configuration be designated as G_1, where

$$G_1 = \frac{1}{4\pi}\left[\frac{e^{-jkr_{01}}}{r_{01}} - \frac{e^{-jkr_{i1}}}{r_{i1}}\right], \tag{1.10}$$

where $\underline{r_i}$ is the image of $\underline{r_0}$ in the $z = 0$ plane. It is located at $z < 0$, as shown in Fig. 1.3. G_1 is zero on the xy plane at $z = 0$. When G_1 is used in the Green's identity, Eq. (1.8), we obtain

$$U(\underline{r_0}) = \iint_{\Omega} U(x, y, z = 0)\,\frac{\partial G_1}{\partial z}\,dx\,dy. \tag{1.11}$$

Here, Ω refers to the xy plane at $z = 0$. Because of the radiation condition expressed in Eq. (1.9), the value of the surface integral over the very large semi-sphere enclosing the $z > 0$ volume (with $R \to \infty$) is zero.

For most applications, $U \neq 0$ only in a small sub-area of Σ, e.g. the radiation U is incident on an opaque screen that has a limited open aperture Ω. In that case, $-\partial G_1/\partial z$ at $z_0 \gg \lambda$ can be simplified to obtain

$$-\nabla G_1 \cdot \underline{i_z} = 2\cos\alpha\,\frac{e^{-jkr_{01}}}{4\pi\, r_{01}}\,(-jk),$$

where α is as illustrated in Fig. 1.2. Therefore, the simplified expression for U is

$$U(\underline{r_0}) = \frac{j}{\lambda}\iint_{\Omega} U\,\frac{e^{-jkr_{01}}}{r_{01}}\cos\alpha\,dx\,dy. \tag{1.12}$$

This result has also been derived from the Huygens principle in classical optics.

Let us now define the paraxial approximation for the observer at position (x_0, y_0, z_0) in a direction close to the direction of propagation and at a distance reasonably far from the aperture, i.e. $\alpha \approx 180^\circ$ and $|r_{01}| \approx |z| \approx \rho$. Then, for observers in the paraxial approximation, α is now approximately a constant in the integrand of Eq. (1.12) over the entire aperture Ω, while the change of ρ in the denominator of the integrand also varies very slowly over the entire Ω. Thus, U can now be simplified further to yield

$$U(z \approx \rho) = \frac{-j}{\lambda\,\rho}\iint_{\Omega} U e^{-jkr_{01}}\,dx\,dy. \tag{1.13}$$

Note that $k = 2\pi/\lambda$ and ρ/λ is a very large quantity. A small change of r_{01} in the exponential can affect significantly the value of the integral, while the ρ factor in the denominator of the integrand can be considered as a constant in the paraxial approximation.

Both Eqs. (1.8) and (1.13) are known as Kirchhoff's diffraction formula [3]. In the case of paraxial approximation, limited aperture and $z \gg \lambda$, Eq. (1.8) yields