國立臺灣大學理學院物理學系

碩士論文

Department of Physics

College of Science

National Taiwan University

Master Thesis

繞射之線積分理論

Line-Integral Representation of Diffraction Theory

盧奕銓

Yi-Chuan Lu

指導教授：陳義裕 博士

Advisor: Yih-Yuh Chen, Ph.D.

中華民國 100 年 6 月

June, 2011

Contents

Acknowledgement

感謝我的指導教授陳義裕博士，在每一次的討論中，都能夠給予我許多獨到的見解與清晰的物理詮釋。在我遇到瓶頸的時候，他也能夠適時的指引我新的方向。感謝我的指導教授在這段過程中悉心的指導，讓我最後能夠順利完成本篇論文。

中文摘要

本篇論文探討以沿著孔洞邊緣的線積分來表示一個純量波經過孔洞之後的繞射行為，並探討不同線積分的形式所能給出的不同物理詮釋。

Abstract

Traditionally, the diffraction of a scalar wave satisfying Helmholtz equation through an aperture on an otherwise black screen can be solved approximately by Kirchhoff's integral over the aperture. Rubinowicz, on the other hand, was able to split the solution into two parts: one is the geometrical optics wave that appears only in the geometrical illuminated region, and the other representing the reflected wave is a line integral along the edge of the aperture. Though providing us with an alternative perspective on the diffraction phenomena, this decomposition theory is not entirely satisfactory in the sense that the two separated fields are discontinuous at the boundary of the illuminated region. Also, the functional form of the line integral is not what one would expect an ordinary reflection wave should be due to some confusing factors in the integrand. Finally, the boundary conditions on the screen imposed by Kirchhoff's approximation are mathematically inconsistent, and therefore to be more rigorous this decomposition formulation must be slightly modified by taking into account the correct boundary conditions.

In this thesis, we derive a slightly different decomposition formulation that avoids the discontinuity, and also we deform the functional form of the line integral into another one that mimics the ordinary reflection behavior of waves, and finally, all

these works are done based on the mathematically consistent boundary conditions.

In the appendix, we digress a little to see how to solve diffraction problems subject to "physical" boundary conditions, which best describe the diffraction phenomena in the real world.

Chapter 1

Introduction

The diffraction phenomenon is quite universal. A propagating wave, once encountered with an obstacle object with its size comparable to the wave length, will be distorted from its original pattern predicted by geometrical optics. This phenomenon could happen to any kind of waves, including scalar waves (such as acoustic waves or Schrödinger's wave functions) or vectorial waves (such as electromagnetic waves); although these waves may possess totally different characters in nature, they all "suffer" from diffraction. This suggests that the diffraction phenomenon is a general property of waves, and therefore we may analyze the diffraction problem by considering the simplest example–the diffraction of scalar waves.

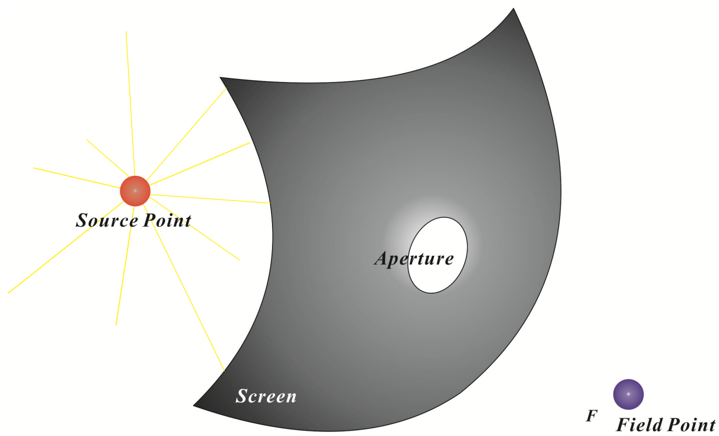

Figure (1.1): The typical diffraction problem

So the question we focus on is quite simple: Given a source distribution and an opaque screen with a hole on it (see Fig.(1.1)), we are asked to find the diffraction wave at the observation point $\vec{r}_F$ (or called the field point) behind the screen. We also assume the wave ψ itself satisfies the Helmholtz equation

$$\left(\nabla^2 + k^2\right)\psi = 0$$

behind the screen, where the constant k represents the wave number. So ψ actually describes a monotonic wave with definite frequency, and thus there is no need to wonder if the media is dispersive. All we have to care is the behavior of ψ due to the presence of the screen and the shape of the hole.

Traditionally this problem is solved by using Kirchhoff's integral formula, which is based on the well-known Huygens' principle that predicts the future shape of a given wave is the "superposition" of the spherical waves constructed at each point on the original wave front. (If the readers are not familiar with the Kirchhoff's integral formula, Appendix A provides a quick review of this powerful theory.) But actually even before Fresnel, Thomas Young had another totally different interpretation of the diffraction phenomenon when he observed the behavior of water wave. Young proposed that part of the incident wave undergoes a kind of *reflection* at the edges of the hole (aperture), and the rest just *goes through* the aperture without any perturbation, and the final diffraction wave is the sum (interference) of the two waves. However, Fresnel had pointed out that Young's interpretation had failed to describe the diffraction phenomena quantitatively and as result Young's theory has been forgotten for a long time [1].

About 100 years later, however, Rubinowicz proved mathematically that Kirchhoff's diffraction formula could be exactly decomposed into two parts: one is the direct wave through the aperture and appears only in the ordinary geometrical illuminated region; the other one is a line integral along the edge of the aperture [2, 3]. This big surprise not only represented the triumph of Young's interpretation, but also provided us a new perspective on the diffraction phenomena.

However, this theory is not entirely satisfactory in the sense that the two separated fields are discontinuous at the boundary of the illuminated region. Also, the functional form of the line integral is not what one would expect an ordinary reflection wave should be due to some confusing factors in the integrand. Finally, the boundary conditions on the screen imposed by Kirchhoff's formula are mathematically inconsistent, and therefore to be more rigorous this decomposition formulation must be slightly modified by taking into account the correct boundary conditions.

In this thesis, we solve the problem of inconsistent boundary condition (B.C.) by imposing a new B.C., which guarantees the uniqueness and self-consistency of the solution, and then exploit this modified formulation to derive a neater representation of the line integral along the edge of the aperture, correlating the functional form of the integrand with the ordinary reflection in geometric optics. We also derive a slightly different decomposition formulation which avoids the discontinuity across the geometrical illuminated region, and give a line integral representation in terms of the solid angle subtended by the edge of the aperture, which is also derived by Asvestas by a different approach.

In Appendix B we digress a little discussing on how to solve the diffraction problem subject to a more rigorous boundary condition (called physical boundary condition) for an arbitrary shape of aperture.

Chapter 2

Maggi-Rubinowicz' Formulation

Kirchhoff's diffraction formula has been used widely both theoretically and experimentally with triumphant success. For example, the far-field diffraction pattern predicted by Kirchhoff's formula is accurate, at least for the ordinary experimental equipment. Based on the Kirchhoff's formula, Rubinowicz could also decompose the representation of diffraction wave into geometrical and reflected parts. However, as shown in Appendix A, the mathematically inconsistent boundary condition, which assumes both the value of ψ and the gradient $\partial\psi/\partial n$ to vanish on the screen, makes the whole theory a little unsatisfactory–the solution to the Kirchhoff's formula does not take the boundary values imposed at the very beginning. However, as suggested by Arnold Sommerfeld [4], this problem can be settled down by using proper Green's function for specific geometric shape of screen, and Sommerfeld also derived the modified Kirchhoff's formula for plane screen by using the Green's function for an infinite plane. However, in his book the Rubinowicz' decomposition trick is still treated based on the original mathematically inconsistent B.C.s.

In this chapter we first present Rubinowicz' original formulation based on the original boundary conditions, and then in the following section we reformulate it by using the Green's function for an infinite plane with the mathematically *consistent* B.C.s.

2.1 Rubinowicz' Original Work

The time-independent scalar wave ψ behind the plane screen $z > 0$ (in the opposite side of the source) satisfies the Helmholtz equation

$$\left(\nabla^2 + k^2\right)\psi = 0.$$

Assume that on the aperture, the wave ψ and $\partial\psi/\partial n$ have exactly the same values as the *unperturbed* source wave ψ_s :

$$\begin{cases} \psi = \psi_s \\ \dfrac{\partial \psi}{\partial n} = \dfrac{\partial \psi_s}{\partial n} \end{cases} \quad \text{on the aperture,} \tag{2.1}$$

where $\hat{n}$ is the outward normal of the plane screen pointing to the $-z$ direction. We also require the screen to be "opaque" in the sense that it does not permit any variations of ψ :

$$\begin{cases} \psi = 0 \\ \dfrac{\partial \psi}{\partial n} = 0 \end{cases} \quad \text{on the screen.} \tag{2.2}$$

From Green's second identity, the wave ψ at the field point $\vec{r}_F$ behind the screen can be expressed by the boundary values

$$\begin{aligned}\psi\left(\vec{r}_F\right) &= -\frac{1}{4\pi}\int_{\text{plane}}\left(\psi\frac{\partial G_K}{\partial n}-G_K\frac{\partial\psi}{\partial n}\right)da \\ &= -\frac{1}{4\pi}\int_{\text{aperture}}\left(\psi_s\frac{\partial G_K}{\partial n}-G_K\frac{\partial\psi_s}{\partial n}\right)da \qquad (2.3)\end{aligned}$$

where

$$G_K=\frac{e^{ik\|\vec{r}-\vec{r}_F\|}}{\|\vec{r}-\vec{r}_F\|}\equiv\frac{e^{ikr}}{r} \qquad (2.4)$$

is the Green's function for an infinite plane. The subscript K reminds us that it is the Kirchhoff's type. To simplify the notation, let

$$\vec{F}=-\psi_s\vec{\nabla}G_K+G_K\vec{\nabla}\psi_s$$

and Eq.(2.3) reduces to

$$\begin{aligned}\psi\left(\vec{r}_F\right) &= \frac{1}{4\pi}\int_{\text{aperture}}\vec{F}\cdot d\vec{a}, \\ d\vec{a} &\equiv \hat{n}da.\end{aligned}$$

Now we make an auxiliary surface that shares the boundary of the aperture and encloses the "illuminated region" of geometrical optics. For example, if ψ_s is a plane wave, then the auxiliary surface is a cylinder parallel to the direction of propagation, as shown in Fig.(2.1); if ψ_s is a point source, then the surface is a cone whose vertex coincides the source point, as shown in Fig.(2.2).

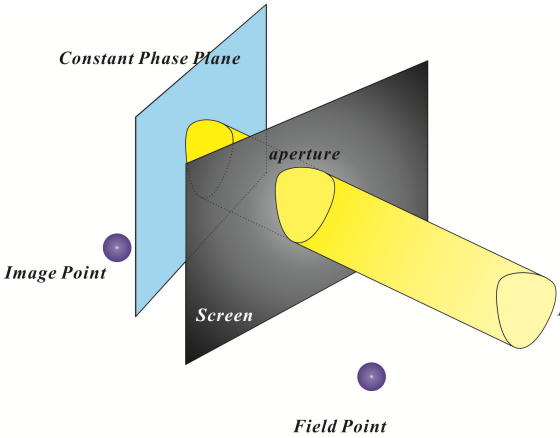

Figure (2.1): Plane Wave

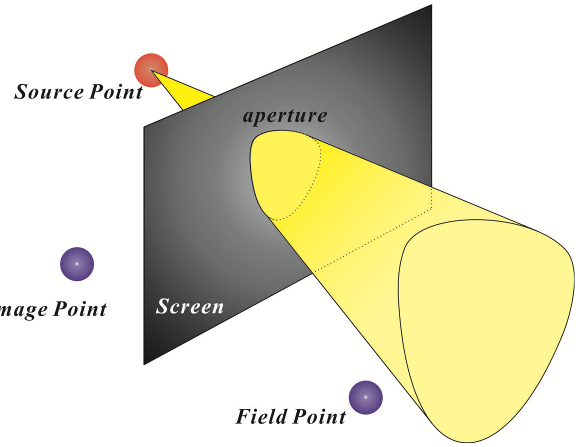

Figure (2.2): Point Source

Now, since

$$\begin{aligned}\vec{\nabla}\cdot\vec{F} &= G_K\nabla^2\psi_s - \psi_s\nabla^2 G_K \\ &= \frac{e^{ikr}}{r}\left(-k^2\psi_s\right) - \psi_s\left(-k^2\frac{e^{ikr}}{r} - 4\pi\delta^3\left(\vec{r}-\vec{r}_F\right)\right) \\ &= 4\pi\psi_s\delta^3\left(\vec{r}-\vec{r}_F\right),\end{aligned}$$

we apply the divergence theorem to the illuminated region $\mathcal{V}$ enclosed by the auxiliary surface and the aperture:

$$\int_{\text{aperture}}\vec{F}\cdot d\vec{a} + \int_{\text{surface}}\vec{F}\cdot d\vec{a} = \int_{\mathcal{V}}\vec{\nabla}\cdot\vec{F}d\tau,$$

then we have

$$\psi\left(\vec{r}_F\right) = -\frac{1}{4\pi}\int_{\text{surface}}\vec{F}\cdot d\vec{a} + \begin{cases}\psi_s\left(\vec{r}_F\right) & ,\vec{r}_F \text{ in the illuminated region} \\ 0 & ,\text{otherwise}\end{cases} .$$

The calculations of the integral over the auxiliary surface follows what Rubinowicz did in 1917: on the boundary $\vec{\nabla}\psi_s = \vec{0}$, and therefore

$$\begin{aligned} -\int_{\text{surface}} \vec{F} \cdot d\vec{a} &= \int_{\text{surface}} \psi_s \vec{\nabla} \frac{e^{ikr}}{r} \cdot d\vec{a} \\ &= \int_{\text{surface}} \psi_s \vec{\nabla} \left(\frac{e^{ikr}}{r} \right) \cdot d\vec{a} \\ &= \int_{\text{surface}} \psi_s \left(\frac{ik}{r} - \frac{1}{r^2} \right) e^{ikr} \hat{r} \cdot d\vec{a} \end{aligned} \tag{2.5}$$

Notice that $\hat{r}$ and $\vec{\rho}_f \equiv \vec{r} - \vec{r}_F$ (the vector from $\vec{r}_F$ to the edge) is related by

$$\hat{r} \cdot d\vec{a} = \frac{\vec{r}}{r} \cdot d\vec{a} = \frac{\vec{\rho}_f}{r} \cdot d\vec{a} \tag{2.6}$$

and actually the quantity $\vec{r} \cdot d\vec{a}$ is the shortest distance between the field point $\vec{r}_F$ and the tangent plane at the integrating point on the surface.

Up to the present we have not made any assumption to the functional form of ψ_s–it can be a point source or a plane wave, for example. To proceed, we must specify the type of the source, as discussed in the following two cases.

Case (I) *Plane Wave Diffraction*

In this case

$$\psi_s = Ae^{ik\rho} \tag{2.7}$$

where ρ is the distance the wave has traveled from an arbitrary constant phase plane, and A is the amplitude.

The illuminated region now is a cylinder parallel to the direction of propagation, and therefore the area element of the auxiliary surface is

$$d\vec{a} = d\vec{l} \times \hat{\rho}_s d\rho \tag{2.8}$$

where $d\vec{l}$ is the line element along the edge of the aperture (counterclockwise as observed from $z > 0$) and $\hat{\rho}_s$ is the unit vector parallel to the propagation[1]. From Eqs.(2.6)(2.7)(2.8), we get

$$\begin{aligned} -\int_{\text{surface}} \vec{F} \cdot d\vec{a} &= A \int_{\text{surface}} e^{ik(\rho+r)} \left(\frac{ik}{r^2} - \frac{1}{r^3} \right) \vec{\rho}_f \cdot \left(d\vec{l} \times \hat{\rho}_s \right) d\rho \\ &= A \oint_{\text{edge}} \vec{\rho}_f \cdot \left(d\vec{l} \times \hat{\rho}_s \right) \int_{\rho_s}^{\infty} e^{ik(\rho+r)} \left(\frac{ik}{r^2} - \frac{1}{r^3} \right) d\rho \end{aligned}$$

Use the law of cosine:

$$r^2 = \rho_f^2 + (\rho - \rho_s)^2 + 2\rho_f (\rho - \rho_s) \hat{\rho}_f \cdot \hat{\rho}_s$$

and then differentiate it with respect to ρ :

$$\begin{aligned} r\frac{dr}{d\rho} &= \rho - \rho_s + \rho_f \hat{\rho}_f \cdot \hat{\rho}_s, \\ r\left(1 + \frac{dr}{d\rho}\right) &= r + \rho - \rho_s + \rho_f \hat{\rho}_f \cdot \hat{\rho}_s. \end{aligned}$$

Use the identities above, and notice that

$$\frac{d}{d\rho} \left(\frac{e^{ik(\rho+r)}}{r\left(r + \rho - \rho_s + \rho_f \hat{\rho}_f \cdot \hat{\rho}_s\right)} \right) = e^{ik(\rho+r)} \left(\frac{ik}{r^2} - \frac{1}{r^3} \right)$$

we have

$$-\int_{\text{surface}} \vec{F} \cdot d\vec{a} = -\oint_{\text{edge}} A e^{ik\rho_s} \frac{e^{ik\rho_f}}{\rho_f} \left(\frac{\hat{\rho}_s \times \hat{\rho}_f}{1 + \hat{\rho}_s \cdot \hat{\rho}_f} \right) \cdot d\vec{l}. \tag{2.9}$$

[1] In this case $\hat{\rho}_s$ is just a constant vector, but as we'll see in the next case, $\hat{\rho}_s$ changes its direction along the edge.

Case (II) *Point Source Diffraction*

In this case

$$\psi_s = A\frac{e^{ik\rho}}{\rho}$$

where $\rho = \|\vec{r} - \vec{r}_S\|$ and $\vec{r}_S$ is the position of the source. The illuminated region is now a cone with vertex at $\vec{r}_S$, with surface element

$$d\vec{a} = \frac{\rho}{\rho_s}\left(d\vec{l} \times \hat{\rho}_s\right) d\rho.$$

As before, the surface integral thus turns out to be

$$\begin{aligned}
-\int_{\text{surface}} \vec{F} \cdot d\vec{a} &= A\int_{\text{surface}} \frac{e^{ik(\rho+r)}}{\rho_s}\left(\frac{ik}{r^2} - \frac{1}{r^3}\right) e^{ikr}\vec{\rho}_f \cdot \left(d\vec{l} \times \hat{\rho}_s\right) d\rho \\
&= A\oint_{\text{edge}} \frac{\vec{\rho}_f}{\rho_s} \cdot \left(d\vec{l} \times \hat{\rho}_s\right) \int_{\rho_s}^{\infty} e^{ik(\rho+r)}\left(\frac{ik}{r^2} - \frac{1}{r^3}\right) d\rho \\
&= -\oint_{\text{edge}} A\frac{e^{ik\rho_s}}{\rho_s}\frac{e^{ik\rho_f}}{\rho_f}\left(\frac{\hat{\rho}_s \times \hat{\rho}_f}{1 + \hat{\rho}_s \cdot \hat{\rho}_f}\right) \cdot d\vec{l}
\end{aligned} \tag{2.10}$$

As we can see in both cases,

$$\psi(\vec{r}_F) = -\frac{1}{4\pi}\oint_{\text{edge}} \psi_s\frac{e^{ik\rho_f}}{\rho_f}\left(\frac{\hat{\rho}_s \times \hat{\rho}_f}{1 + \hat{\rho}_s \cdot \hat{\rho}_f}\right) \cdot d\vec{l} + \begin{cases} \psi_s(\vec{r}_F) & ,\vec{r}_F \in \text{illuminated region} \\ 0 & ,\text{otherwise} \end{cases} \tag{2.11}$$

and the functional form of the line integral is *independent* of the type of the source field.

2.2 Modified Version of Rubinowicz' Decomposition

As mentioned earlier, the derivation in the previous section is based on the inconsistent boundary conditions Eqs.(2.1)(2.2). The B.C.s are inconsistent because the requirements are more than enough; there is no need to impose both ψ and $\partial\psi/\partial n$ on the boundary to determine the unique solution $\psi\left(\vec{r}_F\right)$, and imposing excessive B.C.s generally leads to the self-inconsistency. It is straightforward to show that the solution Eq.(2.11) does not satisfy the B.C.s on the aperture and the screen [2].

Traditionally there are three major types of boundary conditions[6]: the Dirichlet type (the value ψ on the boundary is given), the Neumann type (the gradient of ψ on the boundary is given), and the mixed type (imposing ψ on part of the boundary and $\partial\psi/\partial n$ on the rest part), which are standard B.C.s that can uniquely[3] determine the solution $\psi\left(\vec{r}_F\right)$. In this section we attempt to reformulate Rubinowicz' Decomposition trick based on the Dirichlet boundary condition as suggested by Sommerfeld:

$$\begin{cases} \psi = \psi_s & \text{, on the aperture} \\ \psi = 0 & \text{, on the screen} \end{cases} , \tag{2.12}$$

(the Neumann's version can be easily accomplished by the same procedure.) Introduce

[2] Although Eq.(2.11) is consequently not the mathematical solution to the problem at hand, it IS a solution for another situation called "saltus problem" provided by Kottler, see [5], for example.

[3] Or up to a constant for the Neumann's type.

the Green's function

$$G_D = \frac{e^{ik\|\vec{r}-\vec{r}_F\|}}{\|\vec{r}-\vec{r}_F\|} - \frac{e^{ik\left\|\vec{r}-\vec{r}_F^*\right\|}}{\|\vec{r}-\vec{r}_F^*\|} \equiv \frac{e^{ikr}}{r} - \frac{e^{ikr^*}}{r^*} \tag{2.13}$$

which vanishes at the whole screen yet still satisfies

$$\left(\nabla^2 + k^2\right) G_D = -4\pi\delta^3\left(\vec{r}-\vec{r}_F\right), \text{ for } z > 0$$

where $\vec{r}_F^*$ is the *image* of the field point with respect to the screen[4], as also shown in Figs.(2.1)(2.2). The subscript K reminds us that G_K is the Green's function for the Dirichlet-type B.C.. So again we apply Green's second identity

$$\begin{aligned}
\psi\left(\vec{r}_F\right) &= -\frac{1}{4\pi}\int_{\text{whole screen}} \left(\psi\frac{\partial G_D}{\partial n} - G_D\frac{\partial \psi}{\partial n}\right) da \\
&= -\frac{1}{4\pi}\int_{\text{aperture}} \psi_s\frac{\partial G_D}{\partial n} da \\
&= -\frac{1}{4\pi}\int_{\text{aperture}} \left(\psi_s\frac{\partial G_D}{\partial n} - G_D\frac{\partial \psi_s}{\partial n}\right) da. \qquad (2.14)
\end{aligned}$$

In the third step we add a zero term $G_D\left(\partial\psi_s/\partial n\right)$ to make the integrand "divergence-free", as shown in the following calculations: Let

$$\vec{F} = -\psi_s\vec{\nabla}G_D + G_D\vec{\nabla}\psi_s$$

and Eq.(2.14) reduces to

$$\psi\left(\vec{r}_F\right) = \frac{1}{4\pi}\int_{\text{aperture}} \vec{F}\cdot d\vec{a}.$$

[4]Or, the *inversion* point w.r.t the screen.

As before, we make the same auxiliary surface and apply the divergence theorem to the enclosed region $\mathcal{V}$:

$$\int_{\text{aperture}} \vec{F} \cdot d\vec{a} + \int_{\text{surface}} \vec{F} \cdot d\vec{a} = \int_{\mathcal{V}} \vec{\nabla} \cdot \vec{F} d\tau$$

Although the Green's function now takes a *new* form Eq.(2.13), the divergence of $\vec{F}$ is still *the same* due to the zero term $G_D\left(\partial\psi_s/\partial n\right)$

$$\begin{aligned}
\vec{\nabla} \cdot \vec{F} &= G_D \nabla^2 \psi_s - \psi_s \nabla^2 G_D \\
&= \left(\frac{e^{ikr}}{r} - \frac{e^{ikr^*}}{r^*}\right)\left(-k^2\psi_s\right) - \psi_s \left(-k^2 \frac{e^{ikr}}{r} - 4\pi\delta^3\left(\vec{r} - \vec{r}_F\right) + k^2 \frac{e^{ikr^*}}{r^*}\right) \\
&= 4\pi\psi_s \delta^3\left(\vec{r} - \vec{r}_F\right),
\end{aligned}$$

so again we have

$$\psi\left(\vec{r}_F\right) = -\frac{1}{4\pi}\int_{\text{surface}} \vec{F} \cdot d\vec{a} + \begin{cases} \psi_s\left(\vec{r}_F\right) & ,\vec{r}_F \in \text{illuminated region} \\ 0 & ,\text{otherwise} \end{cases}.$$

What's more, the zero term $G_D\left(\partial\psi_s/\partial n\right)$ still vanishes on the auxiliary surface since $\vec{\nabla}\psi_s = \vec{0}$. Therefore,

$$\begin{aligned}
-\int_{\text{surface}} \vec{F} \cdot d\vec{a} &= \int_{\text{surface}} \left(\psi_s \vec{\nabla} G_D - G_D \vec{\nabla}\psi_s\right) \cdot d\vec{a} \\
&= \int_{\text{surface}} \psi_s \vec{\nabla} G_D \cdot d\vec{a} \\
&= \int_{\text{surface}} \psi_s \vec{\nabla}\left(\frac{e^{ikr}}{r} - \frac{e^{ikr^*}}{r^*}\right) \cdot d\vec{a} \\
&= \int_{\text{surface}} \psi_s \vec{\nabla}\left(\frac{e^{ikr}}{r}\right) \cdot d\vec{a} - \int_{\text{surface}} \psi_s \vec{\nabla}\left(\frac{e^{ikr^*}}{r^*}\right) \cdot d\vec{a}.
\end{aligned}$$

Here we identify that the first term is just Eq.(2.5):

$$\int_{\text{surface}} \psi_s \vec{\nabla}\left(\frac{e^{ikr}}{r}\right) \cdot d\vec{a} = -\frac{1}{4\pi}\oint_{\text{edge}} \psi_s \frac{e^{ik\rho_f}}{\rho_f}\left(\frac{\hat{\rho}_s \times \hat{\rho}_f}{1 + \hat{\rho}_s \cdot \hat{\rho}_f}\right) \cdot d\vec{l}. \tag{2.15}$$

The second part can be evaluated as the same way in the previous section, with replacing $\vec{r}_F$ with $\vec{r}_F^*$, and thus we have a similar line integral

$$-\int_{\text{surface}} \psi_s \vec{\nabla}\left(\frac{e^{ikr^*}}{r^*}\right)\cdot d\vec{a} = \frac{1}{4\pi}\oint_{\text{edge}} \psi_s \frac{e^{ik\rho_f}}{\rho_f}\left(\frac{\hat{\rho}_s \times \hat{\rho}_f^*}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}\right)\cdot d\vec{l}, \tag{2.16}$$

where $\vec{\rho}_f^* \equiv \vec{r} - \vec{r}_F^* =$ (the vector from $\vec{r}_F^*$ to the edge), and $\hat{\rho}_f^*$ is its unit vector. Notice that $\rho_f^* \equiv \left\|\vec{\rho}_f^*\right\| = \rho_f$ since $\vec{r}_F^*$ is the inversion of $\vec{r}_F$ w.r.t the plane. Together with the two line integral representations,

$$\begin{aligned}\psi\left(\vec{r}_F\right) \ &= \begin{cases}\psi_s\left(\vec{r}_F\right) & , \vec{r}_F \in \text{illuminated region} \\ 0 & , \text{otherwise}\end{cases} \\ &\quad -\frac{1}{4\pi}\oint_{\text{edge}} \psi_s \frac{e^{ik\rho_f}}{\rho_f}\left(\frac{\hat{\rho}_s \times \hat{\rho}_f}{1+\hat{\rho}_s\cdot\hat{\rho}_f} - \frac{\hat{\rho}_s \times \hat{\rho}_f^*}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}\right)\cdot d\vec{l}\end{aligned} \tag{2.17}$$

Historically, in his original paper Rubinowicz has ever pointed out that if one used a different Green's function (like what we've done in this section) then there would be an image term associated with the line integral. For some reason most people (including Rubinowicz himself) seem to ignore the image term in the following related papers. Thus traditionally the formula of the reflected part of ψ reads

$$-\frac{1}{4\pi}\oint_{\text{edge}} \psi_s \frac{e^{ik\rho_f}}{\rho_f}\left(\frac{\hat{\rho}_s \times \hat{\rho}_f}{1+\hat{\rho}_s\cdot\hat{\rho}_f}\right)\cdot d\vec{l}. \tag{2.18}$$

However, as mentioned earlier, this formulation Eq.(2.18) has the defect of self-inconsistency: the total wave ψ does not vanish at the screen nor does it take on the value ψ_s as imposed as an assumption in the derivation. Furthermore, as we shall see in Chapter 5, the complete representation Eq.(2.17) exhibits some merits, which in some sense suggests that the complete representation could describe a more physical situation.

Chapter 3

"Reflective" Representation

3.1 Motivations from Geometrical Optics

As derived in the previous chapter, by using the proper Green's function Eq.(2.13) we have a more satisfactory result that exhibits no mathematical inconsistency. However, the so called "reflected wave" part

$$-\frac{1}{4\pi}\oint_{\mathrm{edge}}\psi_s\frac{e^{ik\rho_f}}{\rho_f}\left(\frac{\hat{\rho}_s\times\hat{\rho}_f}{1+\hat{\rho}_s\cdot\hat{\rho}_f}-\frac{\hat{\rho}_s\times\hat{\rho}_f^*}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}\right)\cdot d\vec{l}. \tag{3.1}$$

has become more cumbersome and fails to convey clearly what really happens on the boundary of the aperture due to the confusing factor

$$\frac{\hat{\rho}_s\times\hat{\rho}_f}{1+\hat{\rho}_s\cdot\hat{\rho}_f}-\frac{\hat{\rho}_s\times\hat{\rho}_f^*}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}. \tag{3.2}$$

Traditionally Eq.(3.1)[1] is interpreted as follows[7]: the reflected wave is obtained

[1] Actually, the explanation is given to the incomplete representation Eq.(2.18).

by constructing a spherical wave

$$\psi_{\text{spherical}} \equiv \frac{e^{ik\rho_f}}{\rho_f}$$

at each point on the edge with amplitude ψ_s (evaluated at the boundary), multiplying the angular factor Eq.(3.2), and finally summing over all the spherical waves. But this is actually nothing to do with the "reflection"[2]: In ordinary geometrical optics, the reflection can be comprehended by drawing the "image point" behind the "mirror", as shown in Fig.(3.1), and the reflected wave is equal to incident wave from this image source. Therefore, if what Young really meant (in the early day when he saw the diffraction phenomena) by "reflection" was the reflection in geometrical optics, then we *expect* the line integral should take the form

$$\psi_{\text{reflection}} \sim \oint_{\text{edge}} e^{ik\left(\rho_s+\rho_f\right)} \text{ for plane wave,} \tag{3.3}$$

$$\psi_{\text{reflection}} \sim \oint_{\text{edge}} \frac{e^{ik\left(\rho_s+\rho_f\right)}}{\rho_s+\rho_f} \text{ for point source.} \tag{3.4}$$

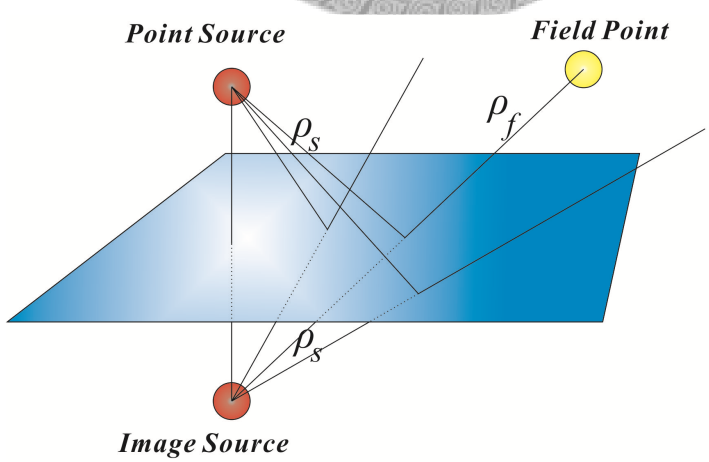

Figure (3.1): Ordinary Reflection

[2]Instead, it resembles Huygens' Principle.

That is, the source wave ψ_s travels an optical path $\rho_s + \rho_f$, where $\rho_s = \|\vec{r} - \vec{r}_S\|$ =the distance from the source to the edge, $\rho_f = \|\vec{r} - \vec{r}_F\|$ =the distance form the edge to the field point, and the edge plays the role of the mirror. This expectation actually can be accomplished by some deformations of Eq.(3.1). But let's do it another way: to derive the "reflected wave" from the beginning Eq.(2.14). This will make the idea more tangible and will enlighten the spirit of Young's primordial interpretation, as shown in the following two sections.

3.2 Reflection at the Boundary

By Eq.(2.14), we have

$$\begin{aligned}
\psi\left(\vec{r}_F\right) &= -\frac{1}{4\pi}\int_{\text{aperture}} \psi_s \frac{\partial}{\partial n}\left(\frac{e^{ikr}}{r} - \frac{e^{ikr^*}}{r^*}\right) da \\
&= -\frac{1}{2\pi}\int_{\text{aperture}} \psi_s \frac{\partial}{\partial n}\left(\frac{e^{ikr}}{r}\right) da \\
&= \frac{1}{2\pi}\int_{\text{aperture}} \psi_s \frac{\partial}{\partial r}\left(\frac{e^{ikr}}{r}\right)\frac{\partial r}{\partial z} da \qquad (3.5)
\end{aligned}$$

Case (I) *Plane Wave Diffraction*

To simplify the calculation, let's assume the wave propagates in the direction perpendicular to the screen–and it is reasonable to make this assumption since experimentally it is the most common configuration. As shown in Fig.(3.2), we make a projection of F on the screen, and call it the origin O. Notice that O does not necessarily lie inside the aperture. Next, let $\vec{l}_f \equiv \overrightarrow{OF}$, and $\vec{l}$ be the vector from O

to a specific point on the edge of the aperture, so that every points along $\vec{l}$ can be described by $s\vec{l}$, where $0 \leq s \leq 1$. Therefore

$$r \equiv \|\vec{r} - \vec{r}_F\| = \left\| s\vec{l} - \vec{l}_f \right\| = \sqrt{s^2 l^2 + l_f^2}$$

and the area element on the aperture is

$$d\vec{a} = \vec{l} ds \times s d\vec{l}.$$

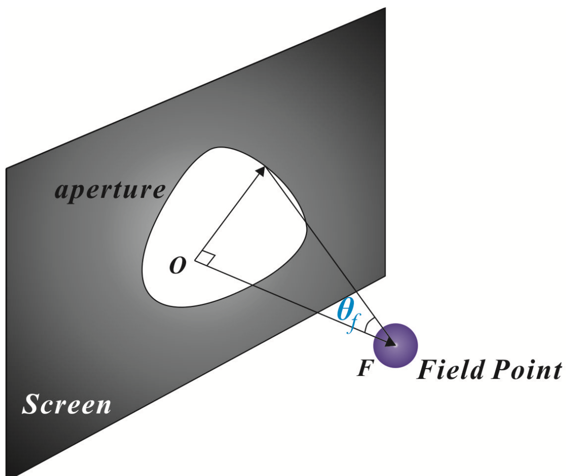

Figure (3.2): Integration over the Aperture

So Eq.(3.5) can be evaluated as (for simplicity, the amplitude A of the source wave is taken to be 1 from now on)

$$\begin{aligned}
\psi(\vec{r}_F) &= -\frac{1}{2\pi} \int_{\text{aperture}} e^{ik\rho} \left(\frac{ik}{r} - \frac{1}{r^2} \right) e^{ikr} \frac{z_f}{r} da \\
&= -\frac{l_f e^{ik\rho_s}}{2\pi} \int_{\text{aperture}} \left(\frac{ik}{r^2} - \frac{1}{r^3} \right) e^{ikr} da \\
&= -\frac{l_f e^{ik\rho_s}}{2\pi} \oint_{\text{edge}} \int_{s=0}^{s=1} \left(\frac{ik}{s^2 l^2 + l_f^2} - \frac{1}{\left(s^2 l^2 + l_f^2\right)^{3/2}} \right) e^{ik\sqrt{s^2 l^2 + l_f^2}} \left(\vec{l} ds \times s d\vec{l} \right) \cdot \hat{e}_z. \\
&= -\frac{l_f e^{ik\rho_s}}{2\pi} \oint_{\text{edge}} \frac{\left(\vec{l} \times d\vec{l} \right) \cdot \hat{e}_z}{l^2} \left(\frac{e^{ik\sqrt{s^2 l^2 + l_f^2}}}{\sqrt{s^2 l^2 + l_f^2}} \right) \Bigg|_{s=0}^{s=1}
\end{aligned}$$

But

$$\left(\vec{l}\times d\vec{l}\right)\cdot\hat{e}_z = l^2 d\phi,$$

where ϕ is the angle subtended by the arc of the boundary of the aperture as measured from O. So

$$\begin{aligned}\psi\left(\vec{r}_F\right) &= \frac{l_f e^{ik\rho_s}}{2\pi}\oint_{\text{edge}} d\phi\left(\frac{e^{ikl_f}}{l_f}-\frac{e^{ik\sqrt{l^2+l_f^2}}}{\sqrt{l^2+l_f^2}}\right)\\ &= \frac{1}{2\pi}\oint_{\text{edge}} d\phi\left(\psi_s\left(\vec{r}_F\right)-e^{ik\left(\rho_s+\rho_f\right)}\cos\theta_f\right)\end{aligned}$$

where θ_f is indicated in Fig.(3.2).

This formula can be transformed into Rubinowicz' type easily: Note that

$$\frac{1}{2\pi}\oint_{\text{edge}}\psi_s\left(\vec{r}_f\right)d\phi=\begin{cases}\psi_s\left(\vec{r}_F\right) & \text{, if } O \text{ lies inside the aperture}\\ 0 & \text{, otherwise}\end{cases}$$

so

$$\boxed{\psi\left(\vec{r}_F\right)=\begin{cases}\psi_s\left(\vec{r}_F\right) & \text{, if } O \text{ lies inside the aperture}\\ 0 & \text{, otherwise}\end{cases}\quad -\frac{1}{2\pi}\oint_{\text{edge}} e^{ik\left(\rho_s+\rho_f\right)}\cos\theta_f d\phi} \tag{3.6}$$

Notice that the "reflected wave"

$$\psi_{\text{reflection}} = -\frac{1}{2\pi}\oint_{\text{edge}} e^{ik\left(\rho_s+\rho_f\right)}\cos\theta_f d\phi$$

now takes the form of Eq.(3.3) accompanied by a more friendly geometric factor $\cos\theta_f$.

Eq.(3.6) also simplifies the calculation of $\psi\left(\vec{r}_F\right)$. For example, when $l_f \to 0$, then $\cos\theta_f \to 0$ and we have

$$\psi\left(\vec{r}_F\right) \to \begin{cases} \psi_s\left(\vec{r}_F\right) & \text{, if } O \text{ lies inside the aperture} \\ 0 & \text{, otherwise} \end{cases},$$

as expected. Furthermore, if the aperture is circular with radius a, then $\psi\left(\vec{r}_f\right)$ right above the center of the circle can be evaluated explicitly:

$$\psi\left(\vec{r}_F\right) = e^{ikl_f} - \frac{l_f e^{ik\sqrt{a^2+l_f^2}}}{\sqrt{a^2+l_f^2}}.$$

Case (II) *Point Source Diffraction*

Similar to the previous case: we assume the $\overline{SF}$ is perpendicular to the screen. Then define O, $\vec{l}_f$, $\vec{l}$ as before, and define a new vector $\vec{l}_s$ to be the vector from O to S. Therefore,

$$\rho \equiv \left\|\vec{r} - \vec{r}_S\right\| = \left\|s\vec{l} - \vec{l}_s\right\| = \sqrt{s^2l^2 + l_s^2}$$

So Eq.(3.5) can be evaluated as

$$\begin{aligned}
\psi\left(\vec{r}_F\right) &= -\frac{1}{2\pi}\int_{\text{aperture}} \frac{e^{ik\rho}}{\rho}\left(\frac{ik}{r} - \frac{1}{r^2}\right) e^{ikr}\frac{l_f}{r}da \\
&= -\frac{l_f}{2\pi}\int_{\text{aperture}} \frac{e^{ik(\rho+r)}}{\rho}\left(\frac{ik}{r^2} - \frac{1}{r^3}\right) da \\
&= -\frac{l_f}{2\pi}\oint_{\text{edge}}\int_{s=0}^{s=1} \frac{e^{ik\left(\sqrt{s^2l^2+l_f^2}+\sqrt{s^2l^2+l_s^2}\right)}}{\sqrt{s^2l^2+l_s^2}}\left(\frac{ik}{s^2l^2+l_f^2} - \frac{1}{\left(s^2l^2+l_f^2\right)^{3/2}}\right)\left(\vec{l}ds \times sd\vec{l}\right)\cdot\hat{e}_z \\
&= -\frac{l_f}{2\pi}\oint_{\text{edge}} \frac{\left(\vec{l}\times d\vec{l}\right)\cdot\hat{e}_z}{l^2}\left(\frac{e^{ik\sqrt{s^2l^2+l_f^2}}}{\sqrt{s^2l^2+l_f^2}}\frac{e^{ik\sqrt{s^2l^2+l_s^2}}}{\sqrt{s^2l^2+l_f^2}+\sqrt{s^2l^2+l_s^2}}\right)\Bigg|_{s=0}^{s=1}
\end{aligned}$$

So

$$
\begin{aligned}
\psi\left(\vec{r}_F\right) &= \frac{l_f}{2\pi}\oint_{\text{edge}}\left(\frac{e^{ikl_f}}{l_f}\frac{e^{ikl_s}}{l_f+l_s}-\frac{e^{ik\sqrt{l^2+l_f^2}}}{\sqrt{l^2+l_f^2}}\frac{e^{ik\sqrt{l^2+l_s^2}}}{\sqrt{l^2+l_f^2}+\sqrt{l^2+l_s^2}}\right)d\phi \\
&= \frac{1}{2\pi}\oint_{\text{edge}}\left(\frac{e^{ik\left(l_f+l_s\right)}}{l_f+l_s}-\frac{e^{ik\left(\rho_s+\rho_f\right)}}{\rho_s+\rho_f}\frac{l_f}{\rho_f}\right)d\phi \\
&= \frac{1}{2\pi}\oint_{\text{edge}}\left(\psi_s\left(\vec{r}_f\right)-\frac{e^{ik\left(\rho_s+\rho_f\right)}}{\rho_s+\rho_f}\cos\theta_f\right)d\phi
\end{aligned}
$$

where θ_f has the same definition as before.

This formula can as well be transformed to Rubinowicz' type:

$$
\boxed{\psi\left(\vec{r}_F\right)=\begin{cases}\psi_s\left(\vec{r}_F\right) & \text{, if } O \text{ lies inside the aperture} \\ 0 & \text{, otherwise}\end{cases}-\frac{1}{2\pi}\oint_{edge}\frac{e^{ik\left(\rho_s+\rho_f\right)}}{\rho_s+\rho_f}\cos\theta_f d\phi}
\tag{3.7}
$$

Notice that the "reflected wave"

$$
\psi_{\text{reflection}}=-\frac{1}{2\pi}\oint_{\text{edge}}\frac{e^{ik\left(\rho_s+\rho_f\right)}}{\rho_s+\rho_f}\cos\theta_f d\phi
$$

takes the form of Eq.(3.4) accompanied by the same factor $\cos\theta_f$.

As before, when $l_f\to 0$, then $\cos\theta_f\to 0$ and

$$
\psi\left(\vec{r}_F\right)\to\begin{cases}\psi_s\left(\vec{r}_F\right) & \text{, if } O \text{ lies inside the aperture} \\ 0 & \text{, otherwise}\end{cases}.
$$

Also, if the aperture is circular with radius a, then $\psi\left(\vec{r}_f\right)$ just right above the center of the circle is

$$
\psi\left(\vec{r}_F\right)=\frac{e^{ik\left(l_s+l_f\right)}}{l_s+l_f}-\frac{e^{ik\left(\sqrt{a^2+l_s^2}+\sqrt{a^2+l_f^2}\right)}}{\sqrt{a^2+l_s^2}+\sqrt{a^2+l_f^2}}\frac{l_f}{\sqrt{a^2+l_f^2}}.
$$

3.3 A More Elegant Derivation

There is a more elegant and easier derivation that utilizes the property of "normal incidence"–the line through S and F is perpendicular to the screen. But before going into the detail, we first mention a lemma from calculus.

Lemma *Assume f is differentiable, g is continuous, and G is the antiderivative of g, then*

$$\begin{aligned}\frac{d}{dx}\int_{f_1(x)}^{f_2(x)} g(t)\,dt &= \frac{d}{dx}\left[G\left(f_1(x)\right) - G\left(f_2(x)\right)\right] \\ &= G'\left(f_1(x)\right) f_1'(x) - G'\left(f_2(x)\right) f_2'(x) \\ &= g\left(f_1(x)\right) f_1'(x) - g\left(f_2(x)\right) f_2'(x)\,.\end{aligned}$$

With this lemma, we can derive the results of Eqs.(3.6)(3.7) easily, as shown in the following two cases:

Case (I) *Plane Wave Diffraction*

Start from Eq.(2.14), and note that ψ_s =constant on the aperture, we have

$$\begin{aligned}\psi\left(\vec{r}_F\right) &= -\frac{1}{4\pi}\int_{\text{aperture}} \psi_s \frac{\partial G_D}{\partial n} da = \frac{\psi_s}{2\pi}\int_{\text{aperture}} \frac{\partial}{\partial z}\left(\frac{e^{ikr}}{r}\right) da \\ &= -\frac{\psi_s}{2\pi}\oint_{\text{edge}}\int_{s=0}^{s=1} \frac{\partial}{\partial z_f}\left(\frac{e^{ikr}}{r}\right) l^2 s ds d\phi = -\frac{\psi_s}{2\pi}\oint_{\text{edge}} \frac{\partial}{\partial z_f}\int_{s=0}^{s=1}\left(\frac{e^{ikr}}{r}\right) l^2 s ds d\phi\end{aligned}$$

where $\partial/\partial z_f$ is the differentiation with respect to the z coordinate of the *field point* $\vec{r}_F$, and thus it can go outside the integral sign.

Since $r=\sqrt{s^2l^2+l_f^2}$, we have $rdr=l^2sds$, and

$$da=l^2sdsd\phi=rdrd\phi.$$

So

$$\begin{aligned}\psi\left(\vec{r}_F\right) &= -\frac{\psi_s}{2\pi}\oint_{\text{edge}}\frac{\partial}{\partial z_f}\int_{s=0}^{s=1}\left(\frac{e^{ikr}}{r}\right)rdrd\phi=-\frac{\psi_s}{2\pi}\oint_{\text{edge}}\frac{\partial}{\partial z_f}\int_{r=l_f}^{r=\rho_f}e^{ikr}drd\phi\\ &= -\frac{\psi_s}{2\pi}\oint_{\text{edge}}\left(e^{ik\rho_f}\frac{\partial\rho_f}{\partial z_f}-e^{ikl_f}\frac{\partial l_f}{\partial z_f}\right)d\phi=\frac{1}{2\pi}\oint_{\text{edge}}\left(e^{ik\left(l_s+l_f\right)}-e^{ik\left(\rho_s+\rho_f\right)}\cos\theta_f\right)d\phi.\end{aligned}$$

In this case the derivation seems to be trivial–we can derive the same result easily without the aid of the lemma. But as shown in the next case, when the integration cannot be performed explicitly, the lemma gives us a way to *bypass* the integration.

Case (II) *Point Source Diffraction*

Also start from Eq.(2.14), and note that $\psi_s=\frac{e^{ik\rho}}{\rho}$ on the aperture, we have

$$\begin{aligned}\psi\left(\vec{r}_F\right) &= -\frac{1}{4\pi}\int_{\text{aperture}}\psi_s\frac{\partial G_D}{\partial n}da=\frac{1}{2\pi}\int_{\text{aperture}}\frac{e^{ik\rho}}{\rho}\frac{\partial}{\partial z}\left(\frac{e^{ikr}}{r}\right)da\\ &= -\frac{1}{2\pi}\oint_{\text{edge}}\frac{\partial}{\partial z_f}\int_{s=0}^{s=1}\frac{e^{ik\rho}}{\rho}\left(\frac{e^{ikr}}{r}\right)l^2sdsd\phi.\end{aligned}$$

Now let

$$u=r+\rho=\sqrt{l^2s^2+l_s^2}+\sqrt{l^2s^2+l_f^2},$$

then

$$\begin{aligned}
du &= \left(\frac{1}{\sqrt{l^2s^2+l_s^2}}+\frac{1}{\sqrt{l^2s^2+l_f^2}}\right)l^2sds \\
&= \left(\frac{\sqrt{l^2s^2+l_s^2}+\sqrt{l^2s^2+l_f^2}}{\sqrt{l^2s^2+l_s^2}\sqrt{l^2s^2+l_f^2}}\right)l^2sds \\
&= \frac{u}{\rho r}l^2sds,
\end{aligned}$$

and therefore

$$\psi\left(\vec{r}_F\right)=-\frac{1}{2\pi}\oint_{\text{edge}}\frac{\partial}{\partial z_f}\left(\int_{u=l_s+l_f}^{u=\rho_s+\rho_f}\frac{e^{iku}}{u}du\right)d\phi.$$

Actually the integral

$$\int\frac{e^{iku}}{u}du$$

cannot be expressed in terms of elementary functions. However, with the aid of the previous lemma, we can skip this problem: remember what we need is the derivatives (with respect to z_f):

$$\begin{aligned}
\psi\left(\vec{r}_F\right) &= \frac{1}{2\pi}\oint_{\text{edge}}\left(\frac{e^{ik\left(l_s+l_f\right)}}{l_s+l_f}\frac{dl_f}{dz_f}-\frac{e^{ik\left(\rho_s+\rho_f\right)}}{\rho_s+\rho_f}\frac{d\rho_f}{dz_f}\right)d\phi \\
&= \frac{1}{2\pi}\oint_{\text{edge}}\left(\frac{e^{ik\left(l_s+l_f\right)}}{l_s+l_f}-\frac{e^{ik\left(\rho_s+\rho_f\right)}}{\rho_s+\rho_f}\cos\theta_f\right)d\phi.
\end{aligned}$$

Note that the geometric and the reflected part of the waves come from the lower and upper limits of the integration, respectively. Upon setting up the coordinate on the aperture (as shown in Fig.(3.3)) the $s-$integral represents integration over different loops from the origin to the edge.

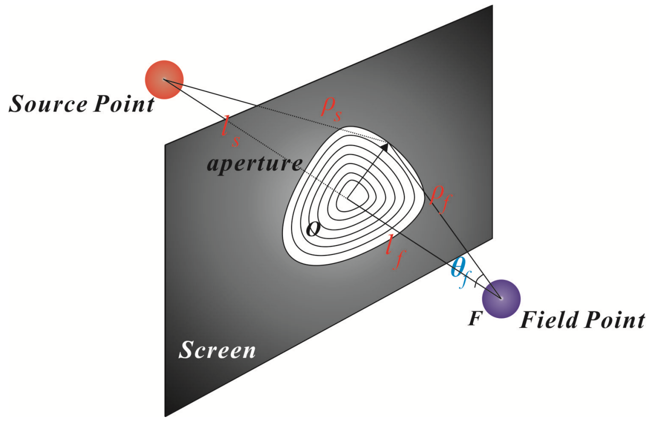

Figure (3.3)

The integral at the innermost loop corresponds to the geometric light, which vanishes if the origin is outside the aperture; the integral at the outermost loop corresponds to the reflection wave at the edge. The integral for the loops in-between somehow eliminate (or interfere) with each other. Moreover, the integrand is always of the form

$$\frac{e^{ik(\rho+r)}}{\rho+r}, \tag{3.8}$$

which means that the light travels an optical path from the source to the edge the aperture, and then to the field point. For example, when $s=0$ Eq.(3.8) becomes

$$\frac{e^{ik\left(l_s+l_f\right)}}{l_s+l_f},$$

which represents a light ray propagating form the source $\vec{r}_S$ directly to $\vec{r}_F$ (geometric light), and when $s=1$, Eq.(3.8) becomes

$$\frac{e^{ik\left(\rho_s+\rho_f\right)}}{\rho_s+\rho_f},$$

representing a light ray propagating form the source to the edge, and then being "reflected" to $\vec{r}_F$ (reflected wave).

Chapter 4

Solid-Angle Representation

4.1 Motivation from Electrostatics

Although $\psi\left(\vec{r}_F\right)$ in Eq.(2.17) is mathematically self-consistent, it still exhibits the same problem as what Rubinowicz encountered–the geometric part

$$\psi_{\text{geometric}}\left(\vec{r}_F\right) \equiv \begin{cases} \psi_s\left(\vec{r}_F\right) & , \vec{r}_F \in \text{illuminated region} \\ 0 & , \text{otherwise} \end{cases}$$

is discontinuous across the surface enclosing the illuminated region, so is the reflected part

$$\psi_{\text{reflected}}\left(\vec{r}_F\right) \equiv -\frac{1}{4\pi}\oint_{\text{edge}} \psi_s \frac{e^{ik\rho_f}}{\rho_f}\left(\frac{\hat{\rho}_s \times \hat{\rho}_f}{1+\hat{\rho}_s \cdot \hat{\rho}_f} - \frac{\hat{\rho}_s \times \hat{\rho}_f^*}{1+\hat{\rho}_s \cdot \hat{\rho}_f^*}\right) \cdot d\vec{l},$$

since, as $\vec{r}_F$ crosses the surface, $\hat{\rho}_s \cdot \hat{\rho}_f$ becomes -1 and the integrand diverges[1].

To overcome this problem, we seek for the analogy in electrostatics: Consider a grounded infinite plane with a finite region σ at which the potential is held at a

[1] The image's term does not exhibit the discontinuity, since $\hat{\rho}_s \cdot \hat{\rho}_f^*$ never takes the value -1.

constant value V_0 :

$$V = \begin{cases} V_0 & \text{, on the region } \sigma \\ 0 & \text{, otherwise} \end{cases} .$$

To solve the potential at the field point $\vec{r}_F$, we define the Green's function:

$$\begin{aligned} G &\equiv \frac{1}{\|\vec{r} - \vec{r}_F\|} - \frac{1}{\|\vec{r} - \vec{r}_F^{*}\|} \equiv \frac{1}{r} - \frac{1}{r^*} \\ \vec{r}_F^{*} &\equiv \text{the image of } \vec{r}_F \text{ w.r.t the plane} \end{aligned}$$

and apply Green's second identity

$$\begin{aligned} V(\vec{r}_F) &= -\frac{1}{4\pi} \int_{\text{whole plane}} \left(V\vec{\nabla} G - G\vec{\nabla} V \right) \cdot d\vec{a} \\ &= -\frac{1}{4\pi} \int_{\sigma} V_0 \vec{\nabla} G \cdot d\vec{a} \\ &= -\frac{1}{2\pi} \int_{\sigma} V_0 \vec{\nabla} \left(\frac{1}{r} \right) \cdot d\vec{a} \\ &= \frac{1}{2\pi} \int_{\sigma} V_0 \left(\frac{\hat{r}}{r^2} \right) \cdot d\vec{a} = \frac{1}{2\pi} \int_{\sigma} V_0 d\Omega. \end{aligned}$$

So

$$V(\vec{r}_F) = \frac{\Omega_f}{2\pi} V_0 \tag{4.1}$$

where Ω_f is the solid angle subtended by the aperture as observed at $\vec{r}_F$.

Inspired by the electrostatic result, we attempt a solution for diffraction problem of the form

$$\psi(\vec{r}_F) = \frac{\Omega_f}{2\pi} \psi_s(\vec{r}_F) + (\text{a line-integral}),$$

that is, we demand the geometric part to possess a solid angle term $\Omega_f/2\pi$. As the field point $\vec{r}_F$ approaches to the black screen, $\Omega_f \to 0$ and the geometric part vanishes;

yet as $\vec{r}_F$ approaches to the aperture, $\Omega_f \to 2\pi$ and the geometric part dominates, and, finally, we regain ψ_s when $\vec{r}_F$ is exactly on the aperture (the reflected part now vanishes as before.) The advantages of this formulation is that both the geometric and reflected parts now vary continuously, without any jump discontinuity across the boundary of the illuminated region.

4.2 Derivations

Let's begin with Eq.(2.14): $\psi\left(\vec{r}_F\right)$ can be expressed as

$$\begin{aligned}
\psi\left(\vec{r}_F\right) &= -\frac{1}{4\pi}\int_{\text{aperture}}\left[\psi_s\frac{\partial G_D}{\partial n} - G_D\frac{\partial \psi_s}{\partial n}\right] da \\
&= -\frac{1}{4\pi}\int_{\text{aperture}}\left[\psi_s\frac{\partial}{\partial n}\left(\frac{e^{ikr}}{r}\right) - \frac{e^{ikr}}{r}\frac{\partial \psi_s}{\partial n}\right] da \\
&\quad + \frac{1}{4\pi}\int_{\text{aperture}}\left[\psi_s\frac{\partial}{\partial n}\left(\frac{e^{ikr^*}}{r^*}\right) - \frac{e^{ikr^*}}{r^*}\frac{\partial \psi_s}{\partial n}\right] da. \\
&\equiv J + J^*
\end{aligned}$$

To evaluate J, we define

$$\vec{F} = -\psi_s\vec{\nabla}\frac{e^{ikr}}{r} + \frac{e^{ikr}}{r}\vec{\nabla}\psi_s$$

as before, and

$$J = \frac{1}{4\pi}\int_{\text{aperture}}\vec{F}\cdot d\vec{a}.$$

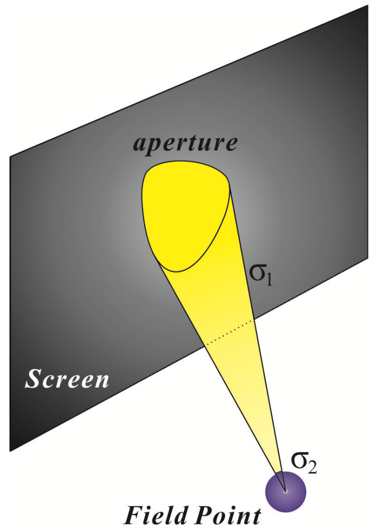

Figure (4.1): Deforming the Integration Region

Next we do a trick slightly different from what Rubinowicz did. We make an auxiliary surface with the vertex at the *field point*, as shown in Fig.(4.1). Also make a small ball centered at the field point, and define σ_2 to be the surface of the small ball inside the cone, while define σ_1 to be the surface of the cone outside the small ball. Apply the divergence theorem to the region enclosed by σ_1, σ_2 and the aperture:

$$\int_{\text{aperture}} \vec{F} \cdot d\vec{a} + \int_{\sigma_1} \vec{F} \cdot d\vec{a} + \int_{\sigma_2} \vec{F} \cdot d\vec{a} = 0$$

or

$$4\pi J + \int_{\sigma_1} \vec{F} \cdot d\vec{a} - \frac{\Omega_f}{4\pi} \left(4\pi\psi_s\left(\vec{r}_F\right)\right) = 0.$$

The third term comes from the fact that σ_2 encloses $\Omega_f/4\pi$ part of the singularity of $\vec{\nabla} \cdot \vec{F}$, which is $4\pi\psi_s\left(\vec{r}_F\right)$. The next task is to evaluate $\int_{\sigma_1} \vec{F} \cdot d\vec{a}$. This can be done by the same trick presented by Rubinowicz, as discuss in the following cases.

Case (I) *Plane Wave Diffraction*

$$
\begin{aligned}
\int_{\sigma_1} \vec{F}\cdot d\vec{a} &= \int_{\sigma_1}\left(-e^{ik\rho}\vec{\nabla}\frac{e^{ikr}}{r}+\frac{e^{ikr}}{r}\vec{\nabla}e^{ik\rho}\right)\cdot d\vec{a} \\
&= \int_{\sigma_1}\frac{e^{ikr}}{r}\vec{\nabla}e^{ik\rho}\cdot d\vec{a} = \int_{\sigma_1}\frac{e^{ikr}}{r}ike^{ik\rho}\hat{\rho}\cdot\left(\frac{\hat{\rho}_f\times d\vec{l}}{\rho_f}rdr\right) \\
&= \int_{\sigma_1} ike^{ik(r+\rho)}\hat{\rho}_s\cdot\left(\frac{\hat{\rho}_f\times d\vec{l}}{\rho_f}dr\right) \\
&= \oint_{\text{edge}} ik\hat{\rho}_s\cdot\left(\frac{\hat{\rho}_f\times d\vec{l}}{\rho_f}\right)\int_{r=0}^{r=\rho_f} e^{ik(r+\rho)}dr.
\end{aligned}
$$

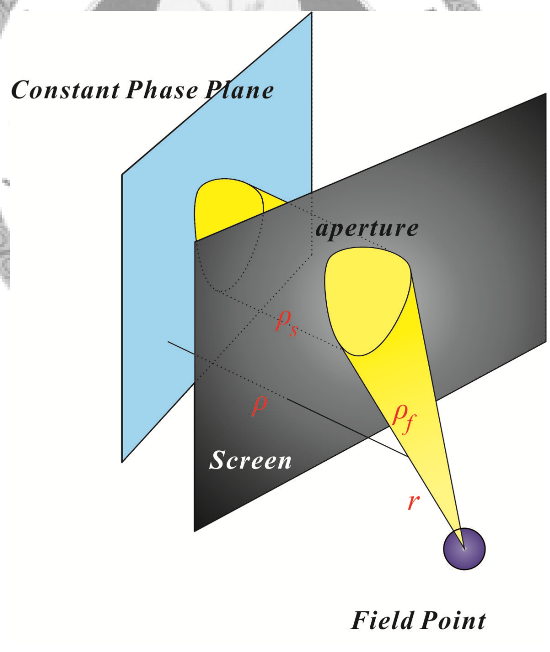

Figure (4.2): Solid Angle for Plane Waves

But from the geometry structure of Fig.(4.2), we have

$$\rho = -\left(\rho_f - r\right)\hat{\rho}_s\cdot\hat{\rho}_f + \rho_s$$

so

$$\begin{aligned}\int_{r=0}^{r=\rho_f} e^{ik(r+\rho)}dr &= e^{ik\left(-\rho_f\hat{\rho}_s\cdot\hat{\rho}_f+\rho_s\right)}\int_{r=0}^{r=\rho_f} e^{ikr\left(1+\hat{\rho}_s\cdot\hat{\rho}_f\right)}dr \\ &= \frac{1}{ik}\frac{e^{ik\left(\rho_s+\rho_f\right)}-e^{ik\left(-\rho_f\hat{\rho}_s\cdot\hat{\rho}_f+\rho_s\right)}}{1+\hat{\rho}_s\cdot\hat{\rho}_f}\end{aligned}$$

and

$$\int_{\sigma_1}\vec{F}\cdot d\vec{a}=\oint_{\text{edge}} e^{ik\rho_s}\frac{e^{ik\rho_f}}{\rho_f}\left(1-e^{-ik\rho_f\left(1+\hat{\rho}_s\cdot\hat{\rho}_f\right)}\right)\left(\frac{\hat{\rho}_s\times\hat{\rho}_f}{1+\hat{\rho}_s\cdot\hat{\rho}_f}\right)\cdot d\vec{l}.$$

Therefore

$$\begin{aligned}J &= \frac{\Omega_f}{4\pi}\psi_s\left(\vec{r}_F\right) \\ &\quad -\frac{1}{4\pi}\oint_{\text{edge}} e^{ik\rho_s}\frac{e^{ik\rho_f}}{\rho_f}\left(1-e^{-ik\rho_f\left(1+\hat{\rho}_s\cdot\hat{\rho}_f\right)}\right)\left(\frac{\hat{\rho}_s\times\hat{\rho}_f}{1+\hat{\rho}_s\cdot\hat{\rho}_f}\right)\cdot d\vec{l},\end{aligned}$$

and it is straightforward to check that both parts of J are now continuous across the surface of illuminated region.

To evaluate J^*, we construct the same surface as used in Chapter 2–the boundary of illuminated region, and apply divergence theorem again:

$$\int_{\text{aperture}}\left[\psi_s\vec{\nabla}\frac{e^{ikr^*}}{r^*}-\frac{e^{ikr^*}}{r^*}\vec{\nabla}\psi_s\right]\cdot d\vec{a}+\int_{\text{surface}}\left[\psi_s\vec{\nabla}\frac{e^{ikr^*}}{r^*}-\frac{e^{ikr^*}}{r^*}\vec{\nabla}\psi_s\right]\cdot d\vec{a}=0$$

or

$$4\pi J^*+\int_{\text{surface}}\psi_s\vec{\nabla}\left(\frac{e^{ikr^*}}{r^*}\right)\cdot d\vec{a}=0.$$

The second part is given by Eq.(2.16) for plane wave. Thus

$$J^*=\frac{1}{4\pi}\oint_{\text{edge}} e^{ik\rho_s}\frac{e^{ik\rho_f}}{\rho_f}\left(\frac{\hat{\rho}_s\times\hat{\rho}_f^*}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}\right)\cdot d\vec{l}$$

Combining J and J^* :

$$\begin{aligned}\psi\left(\vec{r}_F\right) &= \frac{\Omega_f}{4\pi}\psi_s\left(\vec{r}_F\right)\\ &\quad -\frac{1}{4\pi}\oint_{\text{edge}} e^{ik\rho_s}\frac{e^{ik\rho_f}}{\rho_f}\left[\left(1-e^{-ik\rho_f\left(1+\hat{\rho}_s\cdot\hat{\rho}_f\right)}\right)\left(\frac{\hat{\rho}_s\times\hat{\rho}_f}{1+\hat{\rho}_s\cdot\hat{\rho}_f}\right)-\left(\frac{\hat{\rho}_s\times\hat{\rho}_f^*}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}\right)\right]\cdot d\vec{l}.\end{aligned} \tag{4.2}$$

Although this result fits our demand–$\psi\left(\vec{r}_F\right)$ is now expressed in terms of the solid angle Ω_f, it is still unsatisfactory since the factor accompanied is $1/4\pi$ instead of $1/2\pi$. Accordingly, if $\vec{r}_F$ approaches to the aperture, the geometrical part only gives one half of the total source wave $\psi_s\left(\vec{r}_F\right)$, and thus the reflected part must contribute the rest half part of $\psi_s\left(\vec{r}_F\right)$:

$$\begin{cases}\text{Geometric Part} \rightarrow \dfrac{1}{2}\psi_s\left(\vec{r}_F\right)\\ \text{Reflected Part} \rightarrow \dfrac{1}{2}\psi_s\left(\vec{r}_F\right)\end{cases}, \text{ as } \vec{r}_F \text{ approaches to the aperture.}$$

To fix the problem, we take the limit $k \rightarrow 0$, and thus

$$\psi_s = e^{i0\rho} \equiv 1.$$

So by Eq.(4.1), we must have

$$\psi\left(\vec{r}_F\right) \rightarrow \frac{\Omega_f}{4\pi} + \frac{1}{4\pi}\oint_{\text{edge}}\frac{1}{\rho_f}\left(\frac{\hat{\rho}_s\times\hat{\rho}_f^*}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}\right)\cdot d\vec{l} \overset{(4.1)}{=} \frac{\Omega_f}{2\pi},$$

or

$$\frac{\Omega_f}{4\pi} = \frac{1}{4\pi}\oint_{\text{edge}}\frac{1}{\rho_f}\left(\frac{\hat{\rho}_s\times\hat{\rho}_f^*}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}\right)\cdot d\vec{l}. \tag{4.3}$$

This is a line-integral representation of solid angle, which has been also derived by John S. Asvestas [8]. Note that $\hat{\rho}_s$ is an arbitrary constant unit vector which can point to any direction, so the representation is *not unique*.

Finally, we construct $\psi\left(\vec{r}_F\right)$ with proper geometric part by adding Eq.(4.3) into Eq.(4.2)

$$
\begin{aligned}
\psi\left(\vec{r}_F\right) &= \frac{\Omega_f}{2\pi}\psi_s\left(\vec{r}_F\right) - \frac{1}{4\pi}\oint_{\text{edge}} e^{ik\rho_s}\frac{1}{\rho_f}\left(\frac{\hat{\rho}_s\times\hat{\rho}_f^*}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}\right)\cdot d\vec{l} \\
&\quad -\frac{1}{4\pi}\oint_{\text{edge}} e^{ik\rho_s}\frac{e^{ik\rho_f}}{\rho_f}\left[\left(1-e^{-ik\rho_f\left(1+\hat{\rho}_s\cdot\hat{\rho}_f\right)}\right)\left(\frac{\hat{\rho}_s\times\hat{\rho}_f}{1+\hat{\rho}_s\cdot\hat{\rho}_f}\right)-\left(\frac{\hat{\rho}_s\times\hat{\rho}_f^*}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}\right)\right]\cdot d\vec{l} \\
&= \boxed{\frac{\Omega_f}{2\pi}\psi_s\left(\vec{r}_F\right) - \frac{1}{4\pi}\oint_{\text{edge}} \psi_s\frac{e^{ik\rho_f}}{\rho_f}\left[\begin{array}{l}\left(1-e^{-ik\rho_f\left(1+\hat{\rho}_s\cdot\hat{\rho}_f\right)}\right)\left(\dfrac{\hat{\rho}_s\times\hat{\rho}_f}{1+\hat{\rho}_s\cdot\hat{\rho}_f}\right) \\ -\left(1-e^{-ik\left(\rho_s+\rho_f\right)}\right)\left(\dfrac{\hat{\rho}_s\times\hat{\rho}_f^*}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}\right)\end{array}\right]\cdot d\vec{l}}
\end{aligned}
$$

Case (II) *Point Source Diffraction*

$$
\begin{aligned}
\int_{\sigma_1}\vec{F}\cdot d\vec{a} &= \int_{\sigma_1}\left(-\frac{e^{ik\rho}}{\rho}\vec{\nabla}\frac{e^{ikr}}{r}+\frac{e^{ikr}}{r}\vec{\nabla}\frac{e^{ik\rho}}{\rho}\right)\cdot d\vec{a} \\
&= \int_{\sigma_1}\frac{e^{ikr}}{r}\vec{\nabla}\frac{e^{ik\rho}}{\rho}\cdot d\vec{a} = \int_{\sigma_1}\frac{e^{ik(r+\rho)}}{r}\left(\frac{ik}{\rho}-\frac{1}{\rho^2}\right)\hat{\rho}\cdot\left(\frac{\hat{\rho}_f\times d\vec{l}}{\rho_f}r\,dr\right) \\
&= \int_{\sigma_1} e^{ik(r+\rho)}\left(\frac{ik}{\rho}-\frac{1}{\rho^2}\right)\frac{\vec{\rho}_s}{\rho}\cdot\left(\frac{\hat{\rho}_f\times d\vec{l}}{\rho_f}dr\right) \\
&= \oint_{\text{edge}}\vec{\rho}_s\cdot\left(\frac{\hat{\rho}_f\times d\vec{l}}{\rho_f}\right)\int_{r=0}^{r=\rho_f} e^{ik(r+\rho)}\left(\frac{ik}{\rho^2}-\frac{1}{\rho^3}\right)dr
\end{aligned}
$$

Follow the Rubinowicz' trick:

$$
\begin{aligned}
\rho^2 &= \rho_s^2+\left(\rho_f-r\right)^2-2\rho_s\left(\rho_f-r\right)\hat{\rho}_s\cdot\hat{\rho}_f \\
\rho\left(1+\frac{d\rho}{dr}\right) &= r+\rho-\rho_f+\rho_s\hat{\rho}_s\cdot\hat{\rho}_f
\end{aligned}
$$

and

$$
\begin{aligned}
\int_{r=0}^{r=\rho_f} e^{ik(r+\rho)}\left(\frac{ik}{\rho^2}-\frac{1}{\rho^3}\right)dr &= \left.\frac{e^{ik(r+\rho)}}{\rho\left(r+\rho-\rho_f+\rho_s\hat{\rho}_s\cdot\hat{\rho}_f\right)}\right|_{r=0}^{r=\rho_f} \\
&= \frac{e^{ik\left(\rho_f+\rho_s\right)}}{\rho_s^2\left(1+\hat{\rho}_s\cdot\hat{\rho}_f\right)}-\frac{e^{ik\rho_0}}{\rho_0\left(\rho_0-\rho_f+\rho_s\hat{\rho}_s\cdot\hat{\rho}_f\right)} \\
&= \frac{e^{ik\left(\rho_f+\rho_s\right)}}{\rho_s^2\left(1+\hat{\rho}_s\cdot\hat{\rho}_f\right)}-\frac{e^{ik\rho_0}}{\rho_0^2\left(1+\hat{\rho}_0\cdot\hat{\rho}_f\right)}
\end{aligned}
$$

where

$$
\vec{\rho}_0\equiv\vec{\rho}_s-\vec{\rho}_f
$$

is the vector from S to F. So

$$
\begin{aligned}
\int_{\sigma_1}\vec{F}\cdot d\vec{a} &= \oint_{\text{edge}}\left(\frac{e^{ik\left(\rho_f+\rho_s\right)}}{\rho_s\left(1+\hat{\rho}_s\cdot\hat{\rho}_f\right)}-\frac{\rho_s}{\rho_0^2}\frac{e^{ik\rho_0}}{\left(1+\hat{\rho}_0\cdot\hat{\rho}_f\right)}\right)\frac{\hat{\rho}_s\times\hat{\rho}_f}{\rho_f}\cdot d\vec{l} \\
&= \oint_{\text{edge}}\frac{e^{ik\rho_s}}{\rho_s}\frac{e^{ik\rho_f}}{\rho_f}\left(\frac{1}{1+\hat{\rho}_s\cdot\hat{\rho}_f}-\frac{\rho_s^2}{\rho_0^2}\frac{e^{-ik\left(\rho_s+\rho_f-\rho_0\right)}}{1+\hat{\rho}_0\cdot\hat{\rho}_f}\right)\left(\hat{\rho}_s\times\hat{\rho}_f\right)\cdot d\vec{l}.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
J &= \frac{\Omega_f}{4\pi}\psi_s\left(\vec{r}_F\right) \\
&\quad -\frac{1}{4\pi}\oint_{\text{edge}}\frac{e^{ik\rho_s}}{\rho_s}\frac{e^{ik\rho_f}}{\rho_f}\left(\frac{1}{1+\hat{\rho}_s\cdot\hat{\rho}_f}-\frac{\rho_s^2}{\rho_0^2}\frac{e^{-ik\left(\rho_s+\rho_f-\rho_0\right)}}{1+\hat{\rho}_0\cdot\hat{\rho}_f}\right)\left(\hat{\rho}_s\times\hat{\rho}_f\right)\cdot d\vec{l},
\end{aligned}
$$

and it is straightforward to check that both parts of J are now continuous across the surface of illuminated region.

To evaluate J^*, we again construct the same surface as used in Chapter 2–the boundary of illuminated region, and apply the divergence theorem:

$$
\int_{\text{aperture}}\left[\psi_s\vec{\nabla}\frac{e^{ikr^*}}{r^*}-\frac{e^{ikr^*}}{r^*}\vec{\nabla}\psi_s\right]\cdot d\vec{a}+\int_{\text{surface}}\left[\psi_s\vec{\nabla}\frac{e^{ikr^*}}{r^*}-\frac{e^{ikr^*}}{r^*}\vec{\nabla}\psi_s\right]\cdot d\vec{a}=0
$$

or

$$4\pi J^* + \int_{\text{surface}} \psi_s \vec{\nabla}\left(\frac{e^{ikr^*}}{r^*}\right) \cdot d\vec{a} = 0.$$

The second part is again, given by Eq.(2.16) for point source, so

$$J^* = \frac{1}{4\pi}\oint_{\text{edge}} \frac{e^{ik\rho_s}}{\rho_s}\frac{e^{ik\rho_f}}{\rho_f}\left(\frac{\hat{\rho}_s \times \hat{\rho}_f^*}{1+\hat{\rho}_s \cdot \hat{\rho}_f^*}\right) \cdot d\vec{l}$$

and combine J and J^*, we have

$$\begin{aligned}\psi\left(\vec{r}_F\right) &= \frac{\Omega_f}{4\pi}\psi_s\left(\vec{r}_F\right) \\ &\quad -\frac{1}{4\pi}\oint_{\text{edge}} \frac{e^{ik\rho_s}}{\rho_s}\frac{e^{ik\rho_f}}{\rho_f}\left[\begin{array}{c}\left(\dfrac{1}{1+\hat{\rho}_s \cdot \hat{\rho}_f} - \dfrac{\rho_s^2}{\rho_0^2}\dfrac{e^{-ik\left(\rho_s+\rho_f-\rho_0\right)}}{1+\hat{\rho}_0 \cdot \hat{\rho}_f}\right)\left(\hat{\rho}_s \times \hat{\rho}_f\right) \\ -\dfrac{\hat{\rho}_s \times \hat{\rho}_f^*}{1+\hat{\rho}_s \cdot \hat{\rho}_f^*}\end{array}\right] \cdot d\vec{l}.\end{aligned} \tag{4.4}$$

To construct the correct factor $1/2\pi$, we use Eq.(4.3) to add another $\frac{\Omega_f}{4\pi}\psi_s\left(\vec{r}_F\right)$ to the geometric wave. But notice that in Eq.(4.3) $\hat{\rho}_s$ is a *constant* vector, and in the case of point source $\hat{\rho}_s$ changes its direction as we integrate along the edge, so we must specify one direction for $\hat{\rho}_s$ in Eq.(4.3) so that we can insert it into Eq.(4.4).

The result is most symmetric if we adopt $\hat{\rho}_s \equiv \hat{\rho}_0 \equiv$the unit vector from $\vec{r}_S$ to $\vec{r}_F$:

$$\begin{aligned}
\psi\left(\vec{r}_F\right) &= \frac{\Omega_f}{2\pi}\psi_s\left(\vec{r}_F\right) - \frac{1}{4\pi}\oint_{\text{edge}}\frac{e^{ik\rho_s}}{\rho_s}\frac{1}{\rho_f}\left(\frac{\hat{\rho}_0\times\hat{\rho}_f^*}{1+\hat{\rho}_0\cdot\hat{\rho}_f^*}\right)\cdot d\vec{l} \\
&\quad -\frac{1}{4\pi}\oint_{\text{edge}}\frac{e^{ik\rho_s}}{\rho_s}\frac{e^{ik\rho_f}}{\rho_f}\left[\begin{array}{c}\left(\frac{1}{1+\hat{\rho}_s\cdot\hat{\rho}_f}-\frac{\rho_s^2}{\rho_0^2}\frac{e^{-ik\left(\rho_s+\rho_f-\rho_0\right)}}{1+\hat{\rho}_0\cdot\hat{\rho}_f}\right)\left(\hat{\rho}_s\times\hat{\rho}_f\right) \\ -\frac{\hat{\rho}_s\times\hat{\rho}_f^*}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}\end{array}\right]\cdot d\vec{l} \\
&= \boxed{\frac{\Omega_f}{2\pi}\psi_s\left(\vec{r}_F\right) - \frac{1}{4\pi}\oint_{\text{edge}}\psi_s\frac{e^{ik\rho_f}}{\rho_f}\left[\begin{array}{c}\left(\frac{1}{1+\hat{\rho}_s\cdot\hat{\rho}_f}-\frac{\rho_s^2}{\rho_0^2}\frac{e^{-ik\left(\rho_s+\rho_f-\rho_0\right)}}{1+\hat{\rho}_0\cdot\hat{\rho}_f}\right)\left(\hat{\rho}_s\times\hat{\rho}_f\right) \\ -\frac{\hat{\rho}_s\times\hat{\rho}_f^*}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}+e^{-ik\rho_f}\left(\frac{\hat{\rho}_0\times\hat{\rho}_f^*}{1+\hat{\rho}_0\cdot\hat{\rho}_f^*}\right)\end{array}\right]\cdot d\vec{l}}
\end{aligned}$$

We have seen that in both cases the field $\psi\left(\vec{r}_F\right)$ has the form

$$\psi\left(\vec{r}_F\right) = \frac{\Omega_f}{2\pi}\psi_s\left(\vec{r}_F\right) - \frac{1}{4\pi}\oint_{\text{edge}}\psi_s\frac{e^{ik\rho_f}}{\rho_f}\left[\cdots\right]\cdot d\vec{l}.$$

However, the integrand in $[\cdots]$ now depends on the type of the source.

There is another point to be noticed: since $\Omega_f \propto 1/{\rho_f}^2$, in the far zone the geometric term is overwhelmed by the reflected field, which is proportional to $1/\rho_f$ of the source wave. The situation reverses in the near zone, of course.

Chapter 5

Approximations

As discussed in previous chapters, the boundary conditions based on Kirchhoff's theory is mathematically inconsistent, and by using the proper Green's function Eq.(2.13), the diffraction theory can be transformed into a boundary value problem of Dirichlet type which is mathematically admissible. However, this is not the whole story. As the source wave ψ_s propagates toward the aperture, the wave must be modified by the presence of the opaque screen, and thus ψ is not exactly equal to ψ_s, the unperturbed source, on the aperture. So the boundary value Eq.(2.12) imposed earlier is still, unsatisfactory in the physical sense.

Nevertheless, the formulation based on Eq.(2.12) can, to some extent, still describe the diffraction phenomena in the real world. For example, Sommerfeld has solved a 2-D straight edge diffraction problem rigorously without using the unperturbed source wave as boundary values [9], and his result can be derived from our formulation developed in Chapter 2 by letting the size of the aperture approach to infinity, while keeping the field point lie in the vicinity of the edge of the aperture. The result is, of course, a little different from Sommerfeld's solution due to the unphysical B.C. Eq.(2.12) we imposed. But if we use Kirchhoff's B.C., things get worse since the result has totally different functional form from Sommerfeld's solution. Consequently, we may "ranks" of boundary conditions as:

$$\text{Kirchhoff}_{(\text{inconsistent})} < \text{Dirichlet}_{(\text{consistent})} < \text{Sommerfeld}_{(\text{physical})}$$

In the following sections, we first solve the 2-D straight edge diffraction problem by using the formulation established in Section 2.2 (with a little approximations), and then compare the results with Sommerfeld's work. In the Appendix B, we provide a more rigorous method for treating the diffraction problem based on the physical boundary conditions.

5.1 Approximated Solution for a Point Source

Consider an infinite half plane lying on $z = 0$ and $x > 0$, with a point source lying in the region $z < 0$ as before. The solution of diffracted wave in the space $z > 0$ can be solved by Eq.(2.17)

$$\begin{aligned}
\psi\left(\vec{r}_F\right) &= \begin{cases} \psi_s\left(\vec{r}_F\right) & ,\vec{r}_F \in \text{illuminated region} \\ 0 & ,\text{otherwise} \end{cases} \\
&\quad -\frac{1}{4\pi}\int_{\text{edge}} \psi_s \frac{e^{ik\rho_f}}{\rho_f}\left(\frac{\hat{\rho}_s \times \hat{\rho}_f}{1+\hat{\rho}_s\cdot\hat{\rho}_f} - \frac{\hat{\rho}_s \times \hat{\rho}_f^*}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}\right)\cdot d\vec{l}.
\end{aligned}$$

where the line integral is performed along the infinite straight edge[1]. In the far field region $k\rho_f \gg 1$, we apply stationary-phase approximation to evaluate the reflected part

$$I = \frac{1}{4\pi}\int_{\text{edge}} A\frac{e^{ik\left(\rho_s+\rho_f\right)}}{\rho_s\rho_f}\left(\frac{\hat{\rho}_s\times\hat{\rho}_f}{1+\hat{\rho}_s\cdot\hat{\rho}_f}-\frac{\hat{\rho}_s\times\hat{\rho}_f^*}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}\right)\cdot d\vec{l}.$$

The the stationary point occurs when

$$\nabla\left(\rho_s+\rho_f\right)\cdot d\vec{l}=0,$$

or

$$\left(\hat{\rho}_s+\hat{\rho}_f\right)\cdot d\vec{l}=0.$$

Also, since $\rho_f=\rho_f^*$, at the stationary point, we also have

$$\left(\hat{\rho}_s+\hat{\rho}_f^*\right)\cdot d\vec{l}=0.$$

Now expand the phase at the stationary point:

$$\begin{aligned}
\rho_s+\rho_f &= \left(\rho_s+\rho_f\right)\Big|_0+\nabla\left(\rho_s+\rho_f\right)\Big|_0\cdot\delta\vec{l}+\frac{1}{2}\delta\vec{l}\cdot\left(\frac{\hat{\mathbf{1}}-\hat{\rho}_s\hat{\rho}_s}{\rho_s}+\frac{\hat{\mathbf{1}}-\hat{\rho}_f\hat{\rho}_f}{\rho_f}\right)\Bigg|_0\cdot\delta\vec{l}\\
&= \left(\rho_s+\rho_f\right)\Big|_0+\frac{1}{2}\left(\frac{1}{\rho_s}+\frac{1}{\rho_f}\right)\delta\vec{l}\cdot\left(\hat{\mathbf{1}}-\hat{\rho}_s\hat{\rho}_s\right)\Big|_0\cdot\delta\vec{l}\\
&= \left(\rho_s+\rho_f\right)\Big|_0+\frac{1}{2}\frac{\rho_s+\rho_f}{\rho_s\rho_f}\left(\delta\vec{l}^2-\delta\vec{l}^2\cos^2\left(d\vec{l},\hat{\rho}_s\right)\right)\Big|_0\\
&= \left(\rho_s+\rho_f\right)\Big|_0+\frac{1}{2}\frac{\rho_s+\rho_f}{\rho_s\rho_f}\delta\vec{l}^2\sin^2\left(d\vec{l},\hat{\rho}_s\right)\Big|_0,
\end{aligned}$$

where the subscript 0 denotes the stationary point, which in this special case is the point on the edge nearest to $\vec{r}_F$; also, $\hat{\mathbf{1}}$ is the identity operator in three dimensional

[1] And thus the integral sign $\int$ is used instead of $\oint$.

space. Thus

$$
\begin{aligned}
I &= \frac{1}{4\pi}\int_{\text{edge}} A\frac{e^{ik\left(\rho_s+\rho_f\right)}}{\rho_s\rho_f}\left(\frac{\hat{\rho}_s\times\hat{\rho}_f}{1+\hat{\rho}_s\cdot\hat{\rho}_f}-\frac{\hat{\rho}_s\times\hat{\rho}_f^*}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}\right)\cdot d\vec{l} \\
&\simeq \frac{1}{4\pi}A\frac{e^{ik\left(\rho_s+\rho_f\right)}}{\rho_s\rho_f}\left(\frac{\hat{\rho}_s\times\hat{\rho}_f}{1+\hat{\rho}_s\cdot\hat{\rho}_f}-\frac{\hat{\rho}_s\times\hat{\rho}_f^*}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}\right)\cdot\frac{d\vec{l}}{\left\|d\vec{l}\right\|}\Bigg|_0\int_{\text{edge}} e^{ik\left(\frac{1}{2}\frac{\rho_s+\rho_f}{\rho_s\rho_f}\sin^2\left(d\vec{l},\hat{\rho}_s\right)\right)\delta\vec{l}^2}d\left(\delta l\right) \\
&\simeq \frac{Ae^{ik\left(\rho_s+\rho_f\right)}}{\rho_s\rho_f}\left(\frac{\hat{\rho}_s\times\hat{\rho}_f}{1+\hat{\rho}_s\cdot\hat{\rho}_f}-\frac{\hat{\rho}_s\times\hat{\rho}_f^*}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}\right)\cdot\frac{d\vec{l}}{\left\|d\vec{l}\right\|}\Bigg|_0\sqrt{\frac{2\pi i}{k}\frac{\rho_s\rho_f}{\rho_s+\rho_f}}\frac{1}{\sin\left(d\vec{l},\hat{\rho}_s\right)}\Bigg|_0.
\end{aligned}
$$

But[2]

$$
\begin{aligned}
&\frac{1}{4\pi}A\frac{e^{ik\left(\rho_s+\rho_f\right)}}{\rho_s\rho_f}\left(\frac{\hat{\rho}_s\times\hat{\rho}_f}{1+\hat{\rho}_s\cdot\hat{\rho}_f}-\frac{\hat{\rho}_s\times\hat{\rho}_f^*}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}\right)\cdot\frac{d\vec{l}}{\left\|d\vec{l}\right\|} \\
={}& \frac{1}{4\pi}A\frac{e^{ik\left(\rho_s+\rho_f\right)}}{\rho_s\rho_f}\left(\frac{\hat{\rho}_f\cdot\left(d\vec{l}\times\hat{\rho}_s\right)}{1+\hat{\rho}_s\cdot\hat{\rho}_f}-\frac{\hat{\rho}_f^*\cdot\left(d\vec{l}\times\hat{\rho}_s\right)}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}\right)\frac{1}{\left\|d\vec{l}\right\|} \\
={}& \frac{1}{4\pi}A\frac{e^{ik\left(\rho_s+\rho_f\right)}}{\rho_s\rho_f}\left(\frac{\hat{\rho}_f\cdot\hat{n}}{1+\hat{\rho}_s\cdot\hat{\rho}_f}\sin\left(d\vec{l},\hat{\rho}_s\right)-\frac{\hat{\rho}_f^*\cdot\hat{n}}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}\sin\left(d\vec{l},\hat{\rho}_s\right)\right)
\end{aligned}
$$

where

$$
\hat{n}=\frac{d\vec{l}\times\hat{\rho}_s}{\left\|d\vec{l}\times\hat{\rho}_s\right\|}
$$

is the outward normal of the geometric light cone. Therefore,

$$
\begin{aligned}
I &\simeq \frac{1}{4\pi}A\frac{e^{ik\left(\rho_s+\rho_f\right)}}{\rho_s\rho_f}\left(\frac{\hat{\rho}_f\cdot\hat{n}}{1+\hat{\rho}_s\cdot\hat{\rho}_f}-\frac{\hat{\rho}_f^*\cdot\hat{n}}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}\right)\sin\left(d\vec{l},\hat{\rho}_s\right)\sqrt{\frac{2\pi i}{k}\frac{\rho_s\rho_f}{\rho_s+\rho_f}}\frac{1}{\sin\left(d\vec{l},\hat{\rho}_s\right)} \\
&= \frac{1}{4\pi}A\frac{e^{ik\left(\rho_s+\rho_f\right)}}{\rho_s\rho_f}\left(\frac{\hat{\rho}_f\cdot\hat{n}}{1+\hat{\rho}_s\cdot\hat{\rho}_f}-\frac{\hat{\rho}_f^*\cdot\hat{n}}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}\right)\sqrt{\frac{2\pi i}{k}\frac{\rho_s\rho_f}{\rho_s+\rho_f}}
\end{aligned}
$$

To simplify the factor in the parenthesis, refer to Fig.(5.1).

[2] We suppress the subscript 0 to makes the notation neat.

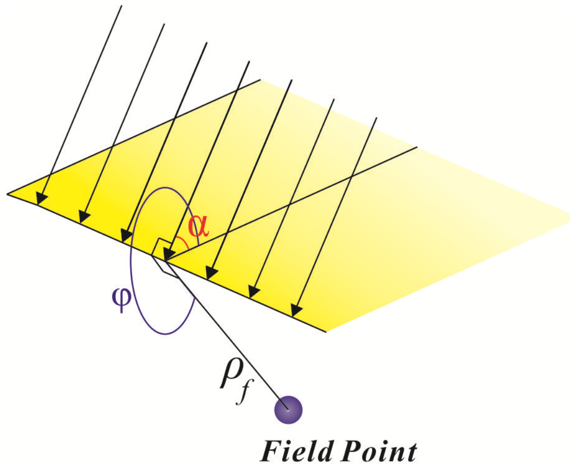

Figure (5.1): Sommerfeld's 2D diffraction configuration

and

$$
\begin{aligned}
1+\hat{\rho}_s\cdot\hat{\rho}_f &= 1+\cos(\phi-\alpha),\\
\hat{\rho}_f\cdot\hat{n} &= -\sin(\phi-\alpha),\\
1+\hat{\rho}_s\cdot\hat{\rho}_f^* &= 1+\cos(\phi+\alpha),\\
\hat{\rho}_f^*\cdot\hat{n} &= \sin(\phi+\alpha).
\end{aligned}
$$

So

$$
\begin{aligned}
&\frac{\hat{\rho}_f\cdot\hat{n}}{1+\hat{\rho}_s\cdot\hat{\rho}_f}-\frac{\hat{\rho}_f^*\cdot\hat{n}}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}\\
=\ &\frac{-\sin(\phi-\alpha)}{1+\cos(\phi-\alpha)}-\frac{\sin(\phi+\alpha)}{1+\cos(\phi+\alpha)}\\
=\ &-\frac{\sin(\phi-\alpha)+\sin(\phi+\alpha)+\sin(\phi-\alpha)\cos(\phi+\alpha)+\cos(\phi-\alpha)\sin(\phi+\alpha)}{1+\cos(\phi+\alpha)+\cos(\phi-\alpha)+\cos(\phi+\alpha)\cos(\phi-\alpha)}\\
=\ &-\frac{2\sin\phi}{\cos\alpha+\cos\phi}
\end{aligned}
$$

Finally,

$$\begin{aligned} I &\simeq \frac{1}{4\pi}A\frac{e^{ik\left(\rho_s+\rho_f\right)}}{\rho_s\rho_f}\left(\frac{\hat{\rho}_f\cdot\hat{n}}{1+\hat{\rho}_s\cdot\hat{\rho}_f}-\frac{\hat{\rho}_f^*\cdot\hat{n}}{1+\hat{\rho}_s\cdot\hat{\rho}_f^*}\right)\sqrt{\frac{2\pi i}{k}\frac{\rho_s\rho_f}{\rho_s+\rho_f}} \\ &= -A\frac{e^{ik\left(\rho_s+\rho_f\right)}}{\sqrt{2\pi k\rho_s\rho_f}}\frac{2\sin\phi}{\cos\alpha+\cos\phi}\frac{e^{i\frac{\pi}{4}}}{\sqrt{\rho_s+\rho_f}} \end{aligned} \tag{5.1}$$

The factor $e^{i\frac{\pi}{4}}$ can explain the reason why the diffraction pattern in the water has a phase delay compared to the incident wave.

5.2 Approximated Solution for Plane Waves

According to Sommerfeld, the diffraction wave based on the rigorous derivation has the form [9]:

$$\psi\left(\vec{r}_F\right)\simeq\begin{cases}\psi_s\left(\vec{r}_F\right)+\dfrac{1+i}{4\sqrt{\pi k\rho_f}}e^{ik\rho_f}\left(\dfrac{1}{\cos\frac{\phi-\alpha}{2}}-\dfrac{1}{\cos\frac{\phi+\alpha}{2}}\right) & ,\vec{r}_F\in\text{illuminated region} \\ \dfrac{1+i}{4\sqrt{\pi k\rho_f}}e^{ik\rho_f}\left(\dfrac{1}{\cos\frac{\phi-\alpha}{2}}-\dfrac{1}{\cos\frac{\phi+\alpha}{2}}\right) & ,\text{otherwise}\end{cases}.$$

for $k\rho_f\gg 1$. Note that the geometric part already has the same form as that of our formula. The rest of the work is to verify to what extent the reflected part derived from our formula can approximate Sommerfeld's. This can be done by considering Eq.(5.1) for plane wave source, i.e., by taking the limit

$$\rho_s\to\infty,\ A\to\infty,\ \text{while keeping }\frac{A}{\rho_s}\to\text{finite number taken to be 1}$$

The result is

$$
\begin{aligned}
I &\simeq -\frac{1}{2\pi}\frac{e^{ik\left(\rho_s+\rho_f\right)}}{\rho_f}\left(\frac{2\sin\phi}{\cos\alpha+\cos\phi}\right)\sqrt{\frac{2\pi i}{k}\rho_f} \\
&= -\frac{e^{ik\left(\rho_s+\rho_f\right)}}{\sqrt{2\pi k\rho_f}}\left(\frac{2\sin\phi}{\cos\alpha+\cos\phi}\right)e^{i\frac{\pi}{4}}
\end{aligned}
$$

The angle factor can be simplified as

$$
\begin{aligned}
\frac{2\sin\phi}{\cos\alpha+\cos\phi} &= -\frac{4\sin\frac{\phi}{2}\sin\frac{\alpha}{2}}{\cos\alpha+\cos\phi}\frac{\cos\frac{\phi}{2}}{2\sin\frac{\alpha}{2}} \\
&= \left(\frac{1}{\cos\frac{\phi-\alpha}{2}}-\frac{1}{\cos\frac{\phi+\alpha}{2}}\right)\frac{\cos\frac{\phi}{2}}{2\sin\frac{\alpha}{2}}
\end{aligned}
$$

and therefore

$$
I \simeq -\frac{1+i}{4\sqrt{\pi k\rho_f}}e^{ik\rho_f}\left(\frac{1}{\cos\frac{\phi-\alpha}{2}}-\frac{1}{\cos\frac{\phi+\alpha}{2}}\right)\frac{\cos\frac{\phi}{2}}{\sin\frac{\alpha}{2}}
$$

Here $e^{ik\rho_s}$ has been dropped since Sommerfeld assumed that the plane wave has phase 0 right at $\rho_f = 0$. So apart from a remaining factor $\frac{\cos\frac{\phi}{2}}{\sin\frac{\alpha}{2}}$, the representation mimic the form given by Sommerfeld. The discrepancy results from the different boundary conditions: in Sommerfeld's formulation, ψ represents the parallel component of the electric field (of the normal component of the magnetic field) w.r.t the screen, and thus $\psi \equiv 0$ when $\alpha = 0$–this is the reason why the factor $\left(\sin\frac{\alpha}{2}\right)$ appears in the denominator. Also, we've assumed the unperturbed ψ_s on the aperture while Sommerfeld used the perturbed ψ, so as the field point approaches the aperture, our formula predicts that the reflected wave I vanishes exactly–this is the reason for the presence of $\cos\frac{\phi}{2}$ in the numerator. Apart from these discrepancies, the angular form of the two solutions is exactly the same.

If, however, we use the mathematically inconsistent boundary conditions Eqs.(2.1)(2.2), then the reflected wave would be

$$I \simeq \frac{1+i}{4\sqrt{\pi k \rho_f}} e^{ik\rho_f} \frac{\sin(\phi-\alpha)}{1+\cos(\phi-\alpha)} = \frac{1+i}{4\sqrt{\pi k \rho_f}} e^{ik\rho_f} \tan\left(\frac{\phi-\alpha}{2}\right),$$

which is totally different from Sommerfeld's solution.

Chapter 6

Conclusion

By using mathematically consistent boundary conditions, we have seen that Rubinowicz' decomposition formulation can be more powerful: the functional form of the line integral becomes much neater and admits a simple interpretation of *reflection at edges*. The new boundary conditions also provide us a different approach that makes the diffraction be much similar to the electrostatic problem by using solid angle representation for the geometric part of the field. Finally, the field predicted by this formulation is much closer to the physical solution.

It is known that the Kirchhoff's surface integral formula would cause some troubles in the numerical simulations, since near the aperture the spherical waves in the Kirchhoff's formula make the integrand diverge. The line integral formulation can avoid such trouble since the edge is still far from the observation point, and remember it is a one-dimensional integration instead of two-dimensional surface integral, and this provides us an easier way for analyzing the diffraction problem with computers.

Appendices

Appendix A

The Mathematical Foundation of Huygens' Principle

In this appendix we present a short review to Kirchhoff's formulation of Huygens' principle, which plays a significant role in the motivation of Rubinowicz' decomposition trick. This is just a brief introduction, remember. If the readers want to have a more comprehensive and complete understanding of Kirchhoff's formula and its applications in optics, I particular recommend the book written by Born M. and Wolf E., *Principles of Optics*; or the book I mentioned throughout this thesis, Sommerfeld's *Optics*.

A.1 Kirchhoff's Formulation

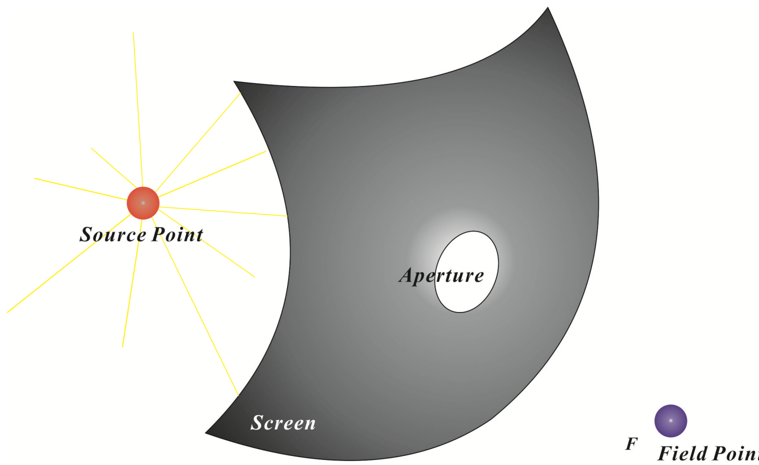

Figure (A.1): The typical diffraction porblem

The spirit of Kirchhoff's formulation is to regard the Huygens' principle as a boundary value problem. We've already know that the scalar wave ψ satisfies the Helmholtz equation

$$\left(\nabla^2 + k^2\right)\psi = 0$$

behind the screen. Using the Green's second identity,

$$\int_{\substack{\text{behind}\\ \text{the screen}}} \left(\psi\nabla^2 G - G\nabla^2\psi\right) d\tau = \int_{\text{whole screen}} \left(\psi\frac{\partial G}{\partial n} - G\frac{\partial \psi}{\partial n}\right) da \qquad \text{(A.1)}$$

where G is the Green's function defined by

$$G \equiv G_K = \frac{e^{ik\|\vec{r}-\vec{r}_F\|}}{\|\vec{r}-\vec{r}_F\|}$$

and n is the outward normal of the screen. Since $\left(\nabla^2 + k^2\right) G_K = -4\pi\delta\left(\vec{r} - \vec{r}_F\right)$, Eq.(A.1) can be reduced to

$$\psi\left(\vec{r}_F\right) = -\frac{1}{4\pi}\int_{\text{whole screen}}\left(\psi\frac{\partial G_K}{\partial n} - G_K\frac{\partial \psi}{\partial n}\right) da. \tag{A.2}$$

To simply the integral, Kirchhoff adopted the boundary conditions

$$\begin{cases} \psi = \psi_s \\ \dfrac{\partial \psi}{\partial n} = \dfrac{\partial \psi_s}{\partial n} \end{cases} \text{on the aperture} \tag{A.3}$$

where ψ_s is the the unperturbed source wave (ψ would be identically ψ_s if there is no screen), and

$$\begin{cases} \psi = 0 \\ \dfrac{\partial \psi}{\partial n} = 0 \end{cases} \text{on the rest of the screen.} \tag{A.4}$$

With these assumptions, Eq.(A.2) simplifies to

$$\psi\left(\vec{r}_F\right) = -\frac{1}{4\pi}\int_{\text{aperture}}\left(\psi_s\frac{\partial G_K}{\partial n} - G_K\frac{\partial \psi_s}{\partial n}\right) da. \tag{A.5}$$

This is the famous Kirchhoff's diffraction formula.

A.2 Sommerfeld's Modification

Kirchhoff's diffraction formula has been used widely since it can predict the diffraction pattern correctly and thus becomes the foundation of diffraction theory. However, there is a snake lurking in the derivation of Kirchhoff's formula. The boundary conditions imposed in the previous section is mathematically inconsistent, for, if an analytic function (satisfies the Helmholtz equation) has zero value $\psi = 0$ and zero gradient $\partial\psi/\partial n = 0$ on a finite region of a surface, then it vanishes *everywhere*. But from the expression of Eq.(A.5), it is clearly not a trivial function, so the only possibility is that wave function ψ given by Eq.(A.5) must not satisfy Eqs.(A.3)(A.4) on the boundary.

To solve this problem, Sommerfeld, on the other hand, proposed another boundary condition given by

$$\begin{cases} \psi = \psi_s & \text{, on the aperture} \\ \psi = 0 & \text{, on the rest of the screen} \end{cases} . \tag{A.6}$$

This boundary condition, on the contrary, constitutes a self-consistent boundary value problem of Dirichlet type, which admit the existence of the solution (and, as the expected, the solution is guaranteed to be unique.) But since we have only assumed the value of ψ on the screen and lacked of the information of $\partial\psi/\partial n$, we have to modify the Green's function such that the $G\,(\partial\psi/\partial n)$ term vanishes completely on the screen. Theoretically this can be accomplished by choosing appropriate Green's function satisfying $G \equiv 0$ on the screen. If we choose the screen to be an infinite plane, then this G can be solved explicitly:

$$G \equiv G_D = \frac{e^{ik\|\vec{r}-\vec{r}_F\|}}{\|\vec{r}-\vec{r}_F\|} - \frac{e^{ik\left\|\vec{r}-\vec{r}_F^{\,*}\right\|}}{\|\vec{r}-\vec{r}_F^{\,*}\|}$$

where $\vec{r}_F^{\,*}$ is image of the field point $\vec{r}_F$ with respective to the plane screen. Substitute

G into Eq.(A.1), together with the B.C. Eq.(A.6), and note that

$$\frac{\partial}{\partial n}\left(\frac{e^{ik\|\vec{r}-\vec{r}_F\|}}{\|\vec{r}-\vec{r}_F\|}\right)=-\frac{\partial}{\partial n}\left(\frac{e^{ik\|\vec{r}-\vec{r}_F^*\|}}{\|\vec{r}-\vec{r}_F^*\|}\right)\text{ on the screen,}$$

we have

$$\psi\left(\vec{r}_F\right)=-\frac{1}{2\pi}\int_{\text{aperture}}\psi_s\frac{\partial}{\partial n}\left(\frac{e^{ik\|\vec{r}-\vec{r}_F\|}}{\|\vec{r}-\vec{r}_F\|}\right)da. \tag{A.7}$$

In the short wavelength limit $k\to\infty$, Eq.(A.7) can be written as

$$\begin{aligned}\psi\left(\vec{r}_F\right) &= \frac{1}{2\pi}\int_{\text{aperture}}\psi_s\left(\frac{ik}{\|\vec{r}-\vec{r}_F\|}-\frac{1}{\|\vec{r}-\vec{r}_F\|^2}\right)\cos\theta_F e^{ik\|\vec{r}-\vec{r}_F\|}da\\ &\simeq \frac{ik}{2\pi}\int_{\text{aperture}}\psi_s\left(\frac{e^{ik\|\vec{r}-\vec{r}_F\|}}{\|\vec{r}-\vec{r}_F\|}\cos\theta_F\right)da\\ &\equiv \frac{ik}{2\pi}\int_{\text{aperture}}\psi_s\left(\frac{e^{ikr}}{r}\cos\theta_F\right)da. \end{aligned}\tag{A.8}$$

where θ_F is indicated in Fig.(A.2).

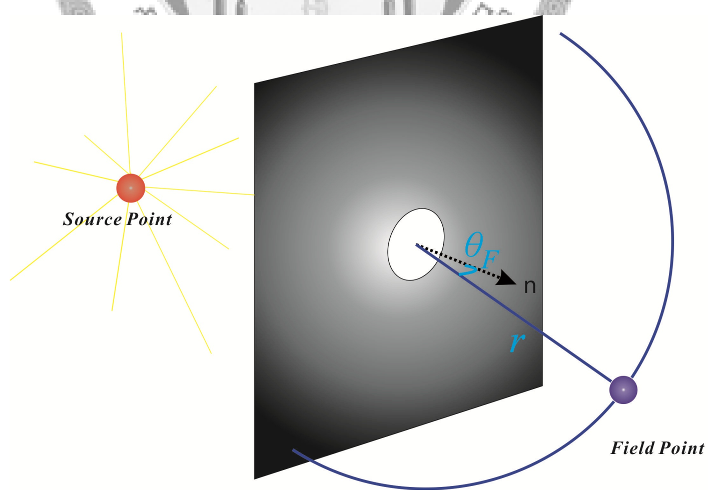

Figure (A.2): Huygens' principle of a plane screen

Eq.(A.8) not only is a self-consistent solution compatible with the B.C. Eq.(A.6), but it has a vivid physical interpretations as well. It tells us that the diffracted wave behind the screen is the summation of spherical waves e^{ikr}/r with amplitudes ψ_s emitted at each point inside the aperture. It also predict that the spherical waves are not isotropic, but has a modulating factor $\cos\theta_F$ that reduces the field to zero when $\vec{r}_F$ approaches to the screen.

Appendix B

Diffraction Problems with Physical Boundary Conditions

Up to the present we've been using the mathematically consistent boundary condition to obtain a series of consequences. However, as mentioned in Chapter 5, this boundary condition is still not entirely satisfactory since, by its nature, it assumes the unperturbed ψ_s on the aperture and thus somehow fails to describe the real phenomena near the aperture especially in the vicinity of the edge, since, as we know, the screen that "absorbs" the incident wave will produce an induced current (or induced charges, due to the model we use) on the surface, and the current then produces another wave in return such as to cancel the wave inside the bulk of the screen. As a result, the only boundary condition that we can imposed is

$$\psi = 0 \text{ on the screen}$$

and we are actually lack of information of the boundary values on the aperture.

This complicates the situation. Historically, Sommerfeld has solved the problem for a half-plane screen with a plane wave propagating perpendicular to the straight edge. Born and Wolf [10] have generalized this problem by demanding the incident wave to propagate in an arbitrary direction (still diffracted by a half-plane screen). In this chapter, we provide general approach that in principle deals with a general situation when the aperture is finite with arbitrary shape (that is, an arbitrary shape hole on an infinite opaque screen), with arbitrary source distribution. However, even the simplest case of normal incidence requires some numerical analysis to estimate the wave function ψ, as shown in the following sections.

B.1 The General Approach

Assume there is an arbitrary source distribution behind the screen. Question: what is the wave function (denoted by $\psi^{(0)}$) if there is no aperture on the screen? The problem is easy: behind the screen (the opposite side of the source) $\psi^{(0)} \equiv 0$; , in front of the screen we pretend to put some image source (same distribution with the "real" one, but with opposite sign) behind the screen, and the wave $\psi^{(0)}$ in front of the screen is the interference of the real and the image source waves. For example, assume that there is a point source at a distance d in front of the screen, then the wave function $\psi^{(0)}$ at the field point $\vec{r}_F = (x, y, z)$ is

$$\psi^{(0)} = \begin{cases} 0 & \text{, behind the screen } (z > 0) \\ \dfrac{e^{ik\sqrt{x^2+y^2+(z+d)^2}}}{\sqrt{x^2+y^2+(z+d)^2}} - \dfrac{e^{ik\sqrt{x^2+y^2+(z-d)^2}}}{\sqrt{x^2+y^2+(z-d)^2}} & \text{, in front of the screen } (z < 0) \end{cases} . \tag{B.1}$$

Now if we carve a hole on the screen, then the total wave can be regarded as a "perturbation" of the original wave $\psi^{(0)}$:

$$\psi = \psi^{(0)} + \psi^{(1)}$$

where $\psi^{(1)}$ is due to the presence of the aperture. But notice that $\psi^{(1)}$ does not need to be small–it is just a change between ψ and $\psi^{(0)}$, and the following derivations are all exact. The task is now reduced to finding the new function $\psi^{(1)}$. First we claim that $\psi^{(1)}$ is an even function of z–this can be comprehend easily from the physical perspective: we know that the contribution from the image source is actually produced by the current (or charge) on the screen, and thus Eq.(B.1)[1] can be expressed in terms of the current (denoted by σ):

$$\psi^{(0)} = \frac{e^{ik\sqrt{x^2+y^2+(z+d)^2}}}{\sqrt{x^2+y^2+(z+d)^2}} + \int_{\text{plane}} \sigma\left(x', y'\right) \frac{e^{ik\sqrt{(x-x')^2+(y-y')^2+z^2}}}{\sqrt{(x-x')^2+(y-y')^2+z^2}} dx'dy',$$

that is, σ acts as the surface source density (like the surface charge density in electrostatics). Now a hole on the screen will in general change the distribution σ to $\sigma + \Delta\sigma$,

[1] We take the point source case for illustration, but actually the consequence is quite general, for every source distribution can be regarded as the superposition of points of source.

so

$$\begin{aligned}
\psi &= \psi^{(0)} + \psi^{(1)} \\
&= \frac{e^{ik\sqrt{x^2+y^2+(z+d)^2}}}{\sqrt{x^2+y^2+(z+d)^2}} + \int_{\text{plane}} \frac{\left(\sigma\left(x',y'\right)+\Delta\sigma\left(x',y'\right)\right) e^{ik\sqrt{(x-x')^2+(y-y')^2+z^2}}}{\sqrt{\left(x-x'\right)^2+\left(y-y'\right)^2+z^2}} dx'dy' \\
&= \left(\frac{e^{ik\sqrt{x^2+y^2+(z+d)^2}}}{\sqrt{x^2+y^2+(z+d)^2}} + \int_{\text{plane}} \sigma\left(x',y'\right) \frac{e^{ik\sqrt{(x-x')^2+(y-y')^2+z^2}}}{\sqrt{\left(x-x'\right)^2+\left(y-y'\right)^2+z^2}} dx'dy' \right) \\
&\quad + \int_{\text{plane}} \Delta\sigma\left(x',y'\right) \frac{e^{ik\sqrt{(x-x')^2+(y-y')^2+z^2}}}{\sqrt{\left(x-x'\right)^2+\left(y-y'\right)^2+z^2}} dx'dy'
\end{aligned}$$

and hence we identify

$$\psi^{(1)} = \int_{\text{plane}} \Delta\sigma\left(x',y'\right) \frac{e^{ik\sqrt{(x-x')^2+(y-y')^2+z^2}}}{\sqrt{\left(x-x'\right)^2+\left(y-y'\right)^2+z^2}} dx'dy',$$

which is an even function of z.

Next we seek the boundary conditions that $\psi^{(1)}$ has to satisfy. Since $\psi \equiv 0$ on the screen, and $\psi^{(0)}$ is already equal to zero on the whole plane $z = 0$ (including the aperture and the screen), we have

$$\psi^{(1)} \equiv 0 \text{ on the screen.} \tag{B.2}$$

Furthermore, since there is no source on the aperture, $\partial\psi/\partial z$ must be continuous:

$$\left.\frac{\partial\psi}{\partial z}\right|_{z=0^+} = \left.\frac{\partial\psi}{\partial z}\right|_{z=0^-} \text{ on the aperture.}$$

Since $\psi^{(1)}$ is even in z, we have

$$\left.\frac{\partial\psi^{(1)}}{\partial z}\right|_{z=0^+} = -\left.\frac{\partial\psi^{(1)}}{\partial z}\right|_{z=0^+},$$

and thus we have[2]

$$\left.\frac{\partial\psi^{(1)}}{\partial z}\right|_{z=0}=\frac{\partial}{\partial z}\left(\frac{e^{ik\sqrt{x^2+y^2+(z+d)^2}}}{\sqrt{x^2+y^2+(z+d)^2}}\right)\Bigg|_{z=0}=\left.\frac{\partial\psi_s}{\partial z}\right|_{z=0}. \tag{B.3}$$

Since what we're interested is ψ in the the region $z>0$ (behind the screen), where $\psi=\psi^{(0)}+\psi^{(1)}=\psi^{(1)}$. Accordingly, the physically admitted boundary condition for ψ is

$$\begin{array}{ll}\dfrac{\partial\psi}{\partial z}=\dfrac{\partial\psi_s}{\partial z} & \text{, on the aperture}\\ \psi=0 & \text{, on the screen}\end{array}. \tag{B.4}$$

It is not the Dirichlet nor the Neumann's type, but a "mixed" boundary condition.

The next strategy is to turn the mixed-boundary-condition problem into the more familiar Neumann's type problem. Define the Green's function

$$G_N=\frac{e^{ik\|\vec{r}-\vec{r}_F\|}}{\|\vec{r}-\vec{r}_F\|}+\frac{e^{ik\|\vec{r}-\vec{r}_F^*\|}}{\|\vec{r}-\vec{r}_F^*\|}=\frac{e^{ikr}}{r}+\frac{e^{ikr^*}}{r^*}$$

and apply Green's second identity

$$\begin{aligned}\psi(\vec{r}_F) &= -\frac{1}{4\pi}\int_{\text{plane}}\left(\psi\frac{\partial G_N}{\partial n}-G_N\frac{\partial\psi}{\partial n}\right)da\\ &= \frac{1}{4\pi}\int_{\text{plane}}G_N\frac{\partial\psi}{\partial n}da\\ &= \frac{1}{2\pi}\int_{\text{plane}}\left(-\frac{\partial\psi}{\partial z}\right)\frac{e^{ikr}}{r}da\equiv\frac{1}{2\pi}\int_{\text{plane}}\sigma\frac{e^{ikr}}{r}da\end{aligned} \tag{B.5}$$

Since we've already known that $\partial\psi/\partial z=\partial\psi_s/\partial z$ on the aperture, then if we can also know the value $\partial\psi/\partial z$ on the screen, then the problem can be solved by integrating Eq.(B.5). In the following section, we'll illustrate a simple case in which the value $\partial\psi/\partial z$ on the screen can be solved analytically.

[2]Again we use the point source case for illustration, and the consequence is generally true, for we can assume an arbitrary functional form of the source wave ψ_s and its image ψ_s^*. It is straight forward to show that $\partial\psi_s/\partial z=-\partial\psi_s^*/\partial z$.

B.2 Clements and Love's Method for Circular Aperture

We consider a simple situation where the aperture is circular and the source distribution possesses azimuthal symmetry. Actually D. L. Clements and E. R. Love have solved the similar problem for electrostatic potentials [11]. Here we apply their method for Helmholtz equation. Before we proceed, we present two lemmas.

Lemma 1 *If a, b and k are real, then*

$$\int_0^{2\pi} \frac{\exp\left(ik\sqrt{a^2+b^2-2ab\cos\theta}\right)}{\sqrt{a^2+b^2-2ab\cos\theta}} d\theta = 4\int_{\max(a,b)}^{\infty} \frac{\exp\left(k\sqrt{\left(t^2-a^2\right)\left(t^2-b^2\right)}/t\right) dt}{\sqrt{\left(t^2-a^2\right)\left(t^2-b^2\right)}}.$$

Lemma 2 *(Abel's Integral Equation) $f \in C^1$ on $a < t < b$. If*

$$f(t) = \int_a^t \frac{g(\rho)}{\sqrt{t^2-\rho^2}} d\rho, \text{ then } g(t) = \frac{2}{\pi}\frac{d}{dt}\int_a^t \frac{\rho f(\rho)}{\sqrt{t^2-\rho^2}} d\rho.$$

The first lemma is actually the generalization of the electrostatic version (or called Copson's integral):

$$\int_0^{2\pi} \frac{d\theta}{\sqrt{a^2+b^2-2ab\cos\theta}} = 4\int_{\max(a,b)}^{\infty} \frac{dt}{\sqrt{\left(t^2-a^2\right)\left(t^2-b^2\right)}},$$

which is derived by Copson in 1947 [12]. The derivation for our modified version can be derived by the same method without difficulty. The second lemma is the famous Abel's integral equation, and the derivation can be found, for example, in [13]. Notice that a slightly change of the notation occurs here: from now on we'll use

ρ as the radial coordinate on the plane (instead of the distance from the aperture to the source). Also, the coordinate of the field point is $\vec{r}_F \equiv (s, \phi, z)$, so

$$\begin{aligned} r &\equiv \|\vec{r} - \vec{r}_F\| = \|(\rho, \theta, 0) - (s, \phi, z)\| \\ &= \sqrt{s^2 + z^2 + \rho^2 - 2s\rho\cos(\phi - \theta)} \end{aligned}$$

(Since our solution ψ now also possesses azimuthal symmetry, ϕ can be taken to be 0.) Assume the radius of the aperture is a, then for $s > a$ and $z = 0$,(on the screen) we have

$$\begin{aligned} 0 &= \frac{1}{2\pi}\int_{\text{plane}} \sigma \frac{e^{ikr}}{r}\bigg|_{s>a,\ z=0} da \\ &= \frac{1}{2\pi}\int_0^\infty \sigma(\rho)\,\rho d\rho \int_0^{2\pi} \frac{\exp\left(ik\sqrt{s^2+\rho^2-2s\rho\cos\theta}\right)}{\sqrt{s^2+\rho^2-2s\rho\cos\theta}} d\theta. \end{aligned}$$

By Lemma 1, the angular integral can be transformed:

$$\begin{aligned} 0 &= \frac{2}{\pi}\int_0^s \sigma(\rho)\,\rho d\rho \int_s^\infty \frac{\exp\left(k\sqrt{(t^2-s^2)(t^2-\rho^2)}/t\right) dt}{\sqrt{(t^2-s^2)(t^2-\rho^2)}} \\ &\quad + \frac{2}{\pi}\int_s^\infty \sigma(\rho)\,\rho d\rho \int_\rho^\infty \frac{\exp\left(k\sqrt{(t^2-s^2)(t^2-\rho^2)}/t\right) dt}{\sqrt{(t^2-s^2)(t^2-\rho^2)}} \\ &= \frac{2}{\pi}\int_s^\infty \frac{dt}{\sqrt{t^2-s^2}} \int_0^s \frac{\exp\left(k\sqrt{(t^2-s^2)(t^2-\rho^2)}/t\right)\sigma(\rho)\,\rho d\rho}{\sqrt{t^2-\rho^2}} \\ &\quad + \frac{2}{\pi}\int_s^\infty \frac{dt}{\sqrt{t^2-s^2}} \int_s^t \frac{\exp\left(k\sqrt{(t^2-s^2)(t^2-\rho^2)}/t\right)\sigma(\rho)\,\rho d\rho}{\sqrt{t^2-\rho^2}} \\ &= \frac{2}{\pi}\int_s^\infty \frac{dt}{\sqrt{t^2-s^2}} \int_0^t \frac{\exp\left(k\sqrt{(t^2-s^2)(t^2-\rho^2)}/t\right)\sigma(\rho)\,\rho d\rho}{\sqrt{t^2-\rho^2}} \\ &\equiv \frac{2}{\pi}\int_s^\infty \frac{f(t)+h(t)\,dt}{\sqrt{t^2-s^2}}, \end{aligned}$$

where

$$
\begin{aligned}
h(t) &\equiv \int_0^a \frac{\exp\left(k\sqrt{(t^2-s^2)(t^2-\rho^2)}/t\right)\sigma_0(\rho)}{\sqrt{t^2-\rho^2}}\rho d\rho \\
f(t) &\equiv \int_a^t \frac{\exp\left(k\sqrt{(t^2-s^2)(t^2-\rho^2)}/t\right)\sigma(\rho)}{\sqrt{t^2-\rho^2}}\rho d\rho.
\end{aligned}
$$

Use the Lemma 2 to solve Abel's integral equation, we have

$$
\begin{aligned}
f(t)+h(t) &= \int_0^a \frac{\exp\left(k\sqrt{(t^2-s^2)(t^2-\rho^2)}/t\right)\sigma_0(\rho)}{\sqrt{t^2-\rho^2}}\rho d\rho \\
&\quad + \int_a^t \frac{\exp\left(k\sqrt{(t^2-s^2)(t^2-\rho^2)}/t\right)\sigma(\rho)}{\sqrt{t^2-\rho^2}}\rho d\rho \\
&= 0.
\end{aligned}
$$

Note that $0 < a < s < t$. Now the equation holds for all $s < t$, so if we take the limit $s \to t$, then the equation reduces to

$$
\int_0^a \frac{\sigma_0(\rho)}{\sqrt{t^2-\rho^2}}\rho d\rho + \int_a^t \frac{\sigma(\rho)}{\sqrt{t^2-\rho^2}}\rho d\rho = 0.
$$

Solve Abel's integral equation again for $\sigma(t)\,t$, we have

$$
\begin{aligned}
\sigma(t)\,t &= -\frac{2}{\pi}\frac{d}{dt}\int_a^t \frac{\rho d\rho}{\sqrt{t^2-\rho^2}}\int_0^a \frac{\sigma_0(\rho')\,\rho' d\rho'}{\sqrt{\rho^2-\rho'^2}} \\
&= -\frac{2}{\pi}\int_0^a \sigma_0(\rho')\,\rho' d\rho' \frac{d}{dt}\int_a^t \frac{\rho d\rho}{\sqrt{t^2-\rho^2}\sqrt{\rho^2-\rho'^2}} \\
&= -\frac{2}{\pi}\int_0^a \sigma_0(\rho')\,\rho' d\rho' \left(\frac{t}{t^2-\rho'^2}\sqrt{\frac{a^2-\rho'^2}{t^2-a^2}}\right).
\end{aligned}
$$

By changing the dummy variable ρ' to ρ, we have

$$
\sigma(t) = -\frac{2}{\pi}\int_0^a \sigma_0(\rho)\,\rho d\rho \left(\frac{1}{t^2-\rho^2}\sqrt{\frac{a^2-\rho^2}{t^2-a^2}}\right). \tag{B.6}
$$

where

$$\sigma_0 = -\frac{\partial \psi_s}{\partial z}\bigg|_{z=0}$$

is supposed to be given on the aperture $0 < \rho < a$. So as long as ψ_s is known, we can integrate over the aperture to get $\sigma(t)$ for $t > a$. That is, we've successfully obtained the general solution of $\partial\psi/\partial z$ and thus transformed the mixed boundary condition problem into the Neumann's Type.

B.3 Normal Incidence

The simplest example for which ψ_s possesses azimuthal symmetry is the plane wave propagating perpendicular to the screen. Assume $\psi_s = e^{ikz}$, then $\sigma_0 = -ik$, and Eq.(B.6) can be evaluated explicitly:

$$\begin{aligned}
\sigma(t) &= -\frac{2}{\pi}\int_0^a \sigma_0(\rho)\,\rho d\rho\left(\frac{1}{t^2-\rho^2}\sqrt{\frac{a^2-\rho^2}{t^2-a^2}}\right) \\
&= \frac{2ik}{\pi}\int_0^a \rho d\rho\left(\frac{1}{t^2-\rho^2}\sqrt{\frac{a^2-\rho^2}{t^2-a^2}}\right) \\
&= \frac{2}{\pi}ik\left(\frac{a}{\sqrt{t^2-a^2}} - \tan^{-1}\left(\frac{a}{\sqrt{t^2-a^2}}\right)\right).
\end{aligned}$$

So the solution of ψ is

$$\psi(\vec{r}_F) = \frac{1}{2\pi}\int_{\text{plane}} \sigma\frac{e^{ikr}}{r}da$$

where

$$\sigma(\rho) = \begin{cases} -ik & \text{, on the aperture} \\ \dfrac{2}{\pi}ik\left(\dfrac{a}{\sqrt{\rho^2-a^2}} - \tan^{-1}\left(\dfrac{a}{\sqrt{\rho^2-a^2}}\right)\right) & \text{, on the screen} \end{cases}.$$

Since σ (or the gradient of ψ) on the whole plane is know, the wave function $\psi\left(\vec{r}_F\right)$ can be estimated by numerical methods.

Bibliography

[1] Arnold. Sommerfeld, *Lectures on Theoretical Physics* **Vol**.(IV) *Optics*, Translated by Otto Laporte & Peter A. Moldauer. (Academic Press Inc., Publishers, New York, 1954), p.311-312

[2] V. A. Rubinowicz, Ann. Phys. 53, 257 (1917).

[3] P. 311-318 of [1].

[4] P. 200 of [1].

[5] B. B. Baker & E. T. Copson, *The Mathematical Theory of Huygens' Principle*, 3rd ed. (Chelsea Publishing Company, New York, 1987), p. 98-101.

[6] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. (John Wiley & Sons, Inc., New York, 1999), p. 37-40.

[7] P. 316 of [1].

[8] J. S. Asvestas, J. Opt. Soc. Am. A2, 891 (1985).

[9] P. 249 of [1].

[10] M. Born & E. Wolf, *Principles of Optics*, 3rd ed. (Cambridge University Press, Cambridge, New York, 1999), p. 657-659.

[11] D. L. Clements & E. R. Love, Proc. Cambridge Phil. Sot., **76**, 313-325 (1974).

[12] E. T. Copson, Proc. Edinburgh Math. Soc., (II) **8**, 14-19, (1947)

[13] Rudolf Gorenflo, Sergio Vessella, *Abel integral equations : Analysis and Applications*, (Berlin, New York, Springer-Verlag, 1991)