Journal of Engineering Mathematics **43:** 255–279, 2002.

Asymptotic analysis and design of diffractive optical elements

SVETLANA RUDNAYA[1], FADIL SANTOSA[2], ALESSANDRA CHIARELI[3] and DAVID MISEMER[4]
[1]*Avant! Corporation, 46871 Bayside Parkway, Freemont, CA 94538, U.S.A.*
[2]*Minnesota Center for Industrial Mathematics, School of Mathematics, University of Minnesota, Vincent Hall, Minneapolis, MN 55455, U.S.A.*
[3]*Specialty Optical Fibers, 3M Company, St. Paul, MN 55144, U.S.A.*
[4]*Software, Electronics, and Mechanical Systems Technology Center, 3M Company, St. Paul, MN 55144, U.S.A.*

Received 18 June 2001; accepted in revised form 17 April 2002

Abstract. The problem of light diffraction by a micro-optical diffractive element is investigated. The method of stationary phase is applied to obtain approximate values of the integrals in the Kirchhoff approximation. The accuracy of the asymptotic approximation is studied in detail. As an application, the obtained approximate formulas are used to solve a design problem of constructing a diffractive optical element with a desired intensity distribution.

Key words: asymptotic analysis, design problem, diffraction, diffractive optical elements, micro-optical device

1. Introduction

In this work, we refer to a Diffractive Optical Element (DOE) as a micro-optical element which modifies incoming light so as to produce a desired intensity pattern on an image plane. We show a schematic of such a system in Figure 1. One application of this technology is in the 'writing' of fiber Bragg gratings. In this application, UV (ultra-violet) light is modified by a DOE, so as to produce a modulated periodic light intensity on a treated fiber [1].

The system we analyze consists of an opaque screen with an aperture over which a DOE is placed. The DOE could be made from a thin film whose thickness varies as a function of position. It is placed far from the light source, so that an incoming light can be assumed to be a plane wave. As the light goes through the DOE, its properties (the phase and/or the amplitude) are changed according to the principles of optics. The modified outgoing light, which is no longer a plane wave, produces a certain intensity pattern on the image plane. We refer to the problem of determining the intensity pattern given a complete information about the DOE as the *Forward Problem*.

Of our particular interest is the design of a DOE, also called an *Inverse Problem*. Given geometrical and physical parameters of the system and a target pattern, we would like to determine the features, such as a thickness variation, of the DOE that produces an intensity pattern that is as close to the target as possible. The inverse problem can be complicated by specific constraints coming from manufacturing. For example, the thickness of the film may be restricted to be piecewise constant, satisfy a given resolution and take on only one of l (≥ 2) given values, where l corresponds to discrete relief levels. The smaller the value of l, is the cheaper the manufacturing process will be.

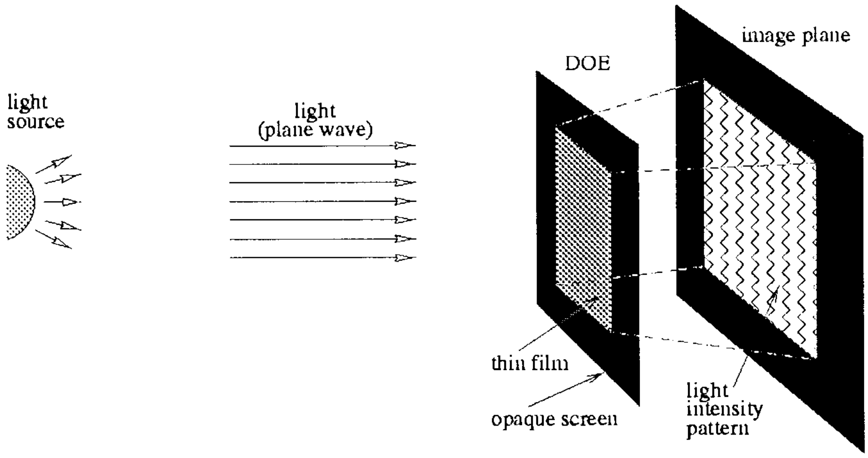

Figure 1. Using the DOE to produce a certain light intensity pattern.

In this work, we use formal asymptotics to analyze the relationship between a DOE and the intensity image it creates. We start with the Kirchhoff model of diffraction, which represents the scalar wave amplitude as an integral over the aperture. Because the DOE considered in this work has rapidly oscillatory thickness variation, with a smooth envelope, we can use the method of stationary phase (and also the method of multiscale expansion; see Appendix) to simplify the said relationship. The simplified formula is used to approximately design a DOE with a given intensity pattern.

The paper is organized as follows. In the next section, we describe the forward problem in the form of the Kirchhoff integral. Section 3 contains the stationary-phase analysis of the integral when the DOE has a multiscale structure. In Section 4, we show how the simplified formula can be used in a design process. The main ideas are illustrated throughout with numerical examples. We conclude with a discussion of our approach and its place in the design of a DOE. An appendix is provided which gives an alternate technique for the asymptotic analysis of the Kirchhoff integral.

2. The forward problem

In our model we use the *Rayleigh-Sommerfeld diffraction theory*, which treats light as a *scalar* phenomenon [2], considering only the scalar amplitude of one component of either the electric or magnetic field and neglecting any relation between them through Maxwell's equations. This assumption has been shown experimentally to give very accurate results, provided that the diffracting aperture is large compared with the wavelength and the diffracted fields are observed not too close to the aperture. These conditions are satisfied in our problem. Another assumption is that we consider only monochromatic waves.

2.1. KIRCHHOFF APPROXIMATION OF DIFFRACTION

We start with the Helmholtz equation for a complex wave amplitude $U(x)$

$$(\nabla^2 + k^2)U = 0, \tag{1}$$

where k is called the *wave number* and given by $k = 2\pi/\lambda$ (λ is wavelength). Thus, U is a scalar field representing either the electric or magnetic field.

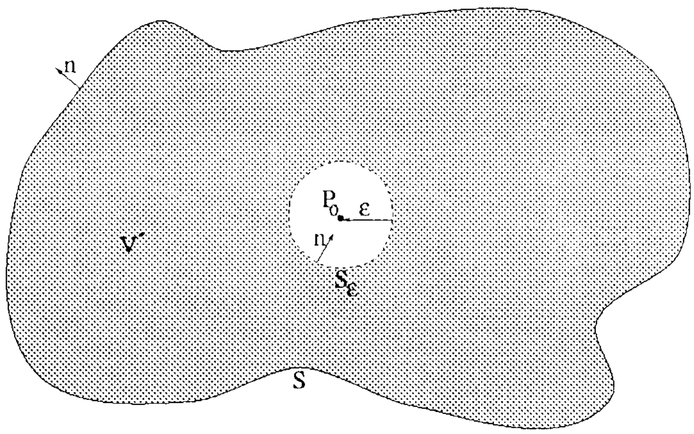

Figure 2. Contour of integration.

We next derive a 2-D version of the Kirchhoff approximation. Our development follows that in [3]. Let $P_0 = (x_0, z_0)$ be the point of observation, S be an arbitrary closed curve surrounding P_0 and V be the region bounded by S. Let S_ε be a circle of small radius ε about the point P_0, V' be the region lying between S and S_ε, $S' = S + S_\varepsilon$, see Figure 2.

If U satisfies the Helmholtz equation in V and G satisfies the Helmholtz equation in V' for any $\varepsilon > 0$, then

$$\int_{V'} (G\nabla^2 U - U\nabla^2 G)\mathrm{d}x\mathrm{d}z = -\int_{V'} (GUk^2 - UGk^2)\mathrm{d}x\mathrm{d}z \equiv 0.$$

Applying Green's formula over V' to the expression, we obtain

$$\int_{S'} \left(G\frac{\partial U}{\partial n} - U\frac{\partial G}{\partial n} \right) \mathrm{d}s = 0$$

or

$$\int_{S} \left(G\frac{\partial U}{\partial n} - U\frac{\partial G}{\partial n} \right) \mathrm{d}s = -\int_{S_\varepsilon} \left(G\frac{\partial U}{\partial n} - U\frac{\partial G}{\partial n} \right) \mathrm{d}s,$$

where the normal vector $\bar{n}$ is taken outward of P_0 on S, but toward P_0 on S_ε.

Choose G to be a constant multiple of the free-space Green's function. With $P_1 = (x, z)$, we have that $G(P_1) = \mathrm{H}_0^{(1)}(kr)$, where $\mathrm{H}_0^{(1)}$ is the Hankel function of the zeroth order [4, 5], and

$$r = \sqrt{(x - x_0)^2 + (z - z_0)^2}.$$

Then G satisfies the 2D Helmholtz equation (1) and

$$\frac{\partial G}{\partial n} = -k\cos(\bar{n}, \bar{r})\mathrm{H}_1^{(1)}(kr),$$

where $\mathrm{H}_1^{(1)}$ is the Hankel function of the first order, and $\bar{r}$ is the unit vector pointing from P_0 to P_1.

For P_1 on S_ε ($\varepsilon \ll 1$),

$$G \sim \frac{2\mathrm{i}}{\pi}\log\varepsilon, \qquad \frac{\partial G}{\partial n} \sim \frac{2\mathrm{i}}{\pi\varepsilon}.$$

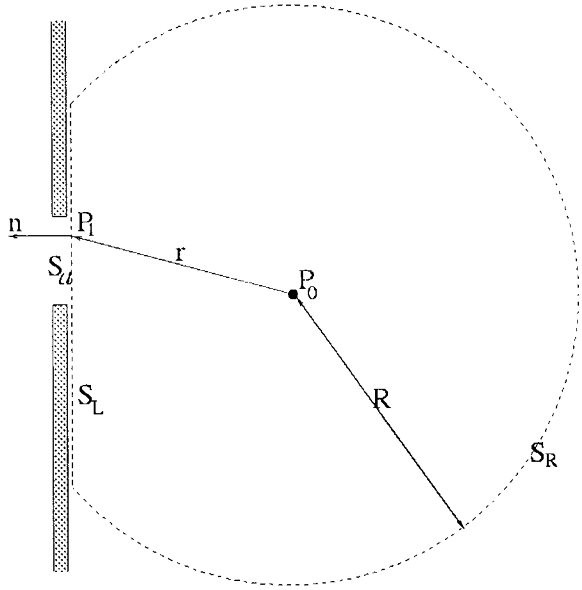

Figure 3. The choice of curve $S = S_L + S_R$.

Therefore

$$\lim_{\varepsilon\to 0}\int_{S_\varepsilon} G\mathrm{d}s = \lim_{\varepsilon\to 0} 2\pi\varepsilon\left(\frac{2\mathrm{i}}{\pi}\log\varepsilon\right) = 0,$$

$$\lim_{\varepsilon\to 0}\int_{S_\varepsilon} \frac{\partial G}{\partial n}\mathrm{d}s = \lim_{\varepsilon\to 0} 2\pi\varepsilon\left(-\frac{2\mathrm{i}}{\pi\varepsilon}\right) = -4\mathrm{i},$$

and we get [6]

$$U(P_0) = \frac{\mathrm{i}}{4}\int_S \left\{\frac{\partial U}{\partial n}G - U\frac{\partial G}{\partial n}\right\}\mathrm{d}s. \tag{2}$$

This result, called the *integral theorem of Helmholtz and Kirchhoff* [3], allows the field at any point P_0 to be expressed in terms of the ‘boundary values’ of the wave on any closed surface surrounding that point.

We now proceed by choosing the closed curve S. Let the closed curve S consist of two parts, S_L and S_R, as shown in Figure 3, so that the line segment, S_L, lies directly behind the diffracting screen and is joined by a large circular arc, S_R, of radius R and centered at the observation point P_0.

Using the Hankel functions properties [7] for large arguments ($\xi \to \infty$)

$$\mathrm{H}_0^{(1)}(\xi) \sim \sqrt{\frac{2}{\pi\xi}}\mathrm{e}^{\mathrm{i}(\xi-\frac{\pi}{4})}, \qquad \mathrm{H}_1^{(1)}(\xi) \sim \sqrt{\frac{2}{\pi\xi}}\mathrm{e}^{\mathrm{i}(\xi-\frac{3\pi}{4})},$$

and the 2D Sommerfeld radiation condition at infinity

$$\lim_{R\to\infty}\sqrt{R}\left(\frac{\partial U}{\partial n} - \mathrm{i}kU\right) = 0 \qquad \text{(uniformly in angle)},$$

we conclude that the integral over S_R vanishes, and we are left with

$$U(P_0) = \frac{\mathrm{i}}{4}\int_{S_L}\left\{\frac{\partial U}{\partial n}G - U\frac{\partial G}{\partial n}\right\}\mathrm{d}s. \tag{3}$$

To simplify formula (3) we modify the Green's function, G, so that it will vanish over the entire surface S_L. The development leading to (3) remains valid with this modification. The modified Green's function in this case is given by

$$\widetilde{G}(P_1) = \mathrm{H}_0^{(1)}(kr) - \mathrm{H}_0^{(1)}(k\tilde{r}),$$

where $r = |\bar{r}| = |\overrightarrow{P_0P_1}|$, $\tilde{r} = |\tilde{\bar{r}}| = |\overrightarrow{\widetilde{P}_0P_1}|$, $\widetilde{P}_0$ is the mirror image of P_0 on the opposite side of the screen. Computing the normal derivative of $\widetilde{G}$ and simplifying, we obtain

$$U(P_0) = \frac{\mathrm{i}k}{2}\int_{S_L} U(P_1)\frac{z_0}{r}\mathrm{H}_1^{(1)}(kr)\mathrm{d}s. \tag{4}$$

Now apply the *Kirchhoff boundary conditions* [8] to (4):

1. *Across the aperture S_a, the field distribution U is exactly the same as it would be in the absence of the screen;*
2. *Over $S_L \setminus S_a$ (the geometrical shadow of the screen), the field distribution U is identically zero.*

The integral in (4) reduces to

$$U(P_0) = \frac{\mathrm{i}k}{2}\int_{S_a} U(P_1)\frac{z_0}{r}\mathrm{H}_1^{(1)}(kr)\mathrm{d}s, \tag{5}$$

or, for kr large,

$$U(P_0) = \frac{\mathrm{e}^{-\mathrm{i}\frac{\pi}{4}}z_0}{\sqrt{\lambda}}\int_{S_a} U(P_1)\frac{\mathrm{e}^{\mathrm{i}kr}}{r^{3/2}}\mathrm{d}s.$$

Formula (5) is often referred to as the *Kirchhoff Approximation*.

The formula allows one to find the field at any point P_0 from knowledge of the field at the aperture S_a. Then the intensity is given by

$$I(P_0) = |U(P_0)|^2 = \frac{z_0^2}{\lambda}\left|\int_{S_a} U(P_1)\frac{\mathrm{e}^{\mathrm{i}kr}}{r^{3/2}}\mathrm{d}s\right|^2. \tag{6}$$

2.2. Lens model

In our problem the DOE serves as the 'aperture'. Formula (6) allows us to compute the light intensity in the near field of the DOE, provided that we know the amplitude of the light on the DOE.

Let a light source be placed far behind the DOE, so that the light is assumed to propagate as a planar wave and the complex amplitude before the DOE is represented by U_0. As the light goes through the DOE, it may change its amplitude and phase depending on the physical properties and geometrical structure of the DOE.

The DOE considered in this work is assumed to change only the phase of the light. The phase shift can be achieved by varying the thickness and/or the refractive index of the material of the DOE; see Figure 4.

For example, let $P_1(x, 0)$ be on the DOE and let the thickness at that point be $d(P_1)$. If the refractive index of the material be n, then the phase shift at this point is approximately

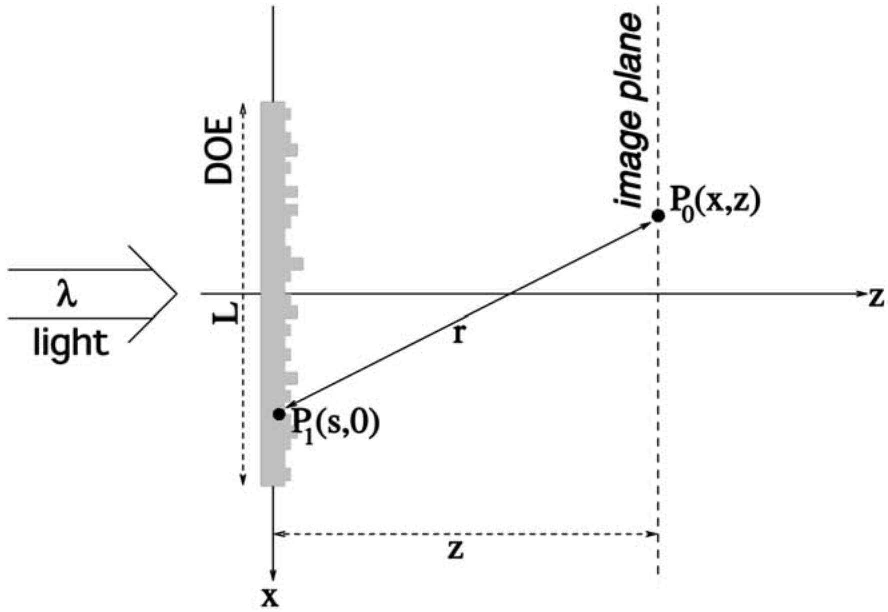

Figure 4. An example of a DOE.

$$\phi(P_1) = \frac{2\pi}{\lambda}(n-1)d(P_1), \tag{7}$$

where λ is the wavelength of the light source. In this case at point P_1, U_0 will be changed to $U(P_1) = U_0 e^{i\phi(P_1)}$. Therefore, the required phase shift can be achieved just by varying the thickness profile of the element. In any case, due to the simplicity of relation (7), from now on we will only consider the phase shift created by a DOE, not its thickness variation.

In general, if $U(P_1) = U_0\, p(P_1)$ then for $P_0(x, z)$, we have

$$U(P_0) = \frac{e^{-i\frac{\pi}{4}} z}{\sqrt{\lambda}} U_0 \int p(s) \frac{e^{ikr}}{r^{3/2}} ds, \tag{8}$$

where $r = \sqrt{(x-s)^2 + z^2}$ and

$$I(x) = |U_0|^2 \frac{z^2}{\lambda} \left| \int p(s) \frac{e^{ikr}}{r^{3/2}} ds \right|^2 .$$

Here $p(s)$ represents the change in the wave caused by the DOE. For instance, $p(s) = e^{i\phi(s)}$ would represent an element that changes only the phase. Since we are interested only in the shape of the intensity profile, we may set $U_0 = 1$.

Formula (8) describes the so-called *Forward problem*. Given $p(\cdot)$ we can determine $I(\cdot)$. In numerical calculations, integration in (8) is performed by quadrature with required accuracy.

3. Asymptotic analysis

According to the Kirchhoff approximation with the lens model, the relation between intensity $I(\cdot)$ and the DOE, represented by phase $p(x)$, is given by (8). The numerical integration involved in (8) can be very expensive computationally (especially in 3-D). We will look for approximate expressions for the integrals for the special case where the phase function, $p(\cdot)$, involves two scales.

We assume that $p(x)$ has compact support $[-w, w]$, where $w \gg \lambda$, $\lambda = 2\pi/k$ is the wavelength of the light source. By the two-scales, we mean that

$$p(x) = P\left(x, \frac{x}{\Lambda}\right),$$

where $\Lambda/w \ll 1$ and $P(x, y)$ is periodic in y.

We write $u(x, z)$ instead of $U(P_0)$. Rescale the problem

$$\tilde{x} = \frac{x}{w}, \qquad \tilde{z} = \frac{z}{w}, \qquad \tilde{p}(\tilde{x}) = p(w\tilde{x}), \qquad \tilde{u}(\tilde{x}, \tilde{z}) = u(w\tilde{x}, w\tilde{z}), \tag{9}$$

drop the 'tildes', and rewrite integral in (8) as

$$u(x, z) = \frac{z}{\sqrt{2\pi \mathrm{i}\varepsilon}} \int_{-1}^{1} p(s) \frac{\mathrm{e}^{\mathrm{i}r/\varepsilon}}{r^{3/2}} \mathrm{d}s, \tag{10}$$

where $r = \sqrt{(x-s)^2 + z^2}$, $\quad \varepsilon = \frac{1}{wk} = \frac{\lambda}{2\pi w} \ll 1$. We have set $U_0 = 1$ in arriving at (10).

After the rescaling, we note that $p(x) = P(x, x/\varepsilon)$ with $P(x, y)$ periodic of period $2\pi/\gamma$ in y, where $\gamma = \frac{\lambda}{\Lambda}$. Therefore, it can be written in the form

$$P(x, y) = \sum_{n=-\infty}^{\infty} P_n(x) \mathrm{e}^{\mathrm{i}\gamma n y},$$

where

$$P_n(x) = \frac{\gamma}{2\pi} \int_0^{2\pi/\gamma} P(x, y) \mathrm{e}^{-\mathrm{i}\gamma n y} \mathrm{d}y.$$

Then

$$p(x) = P\left(x, \frac{x}{\varepsilon}\right) = \sum_{n=-\infty}^{\infty} P_n(x) \mathrm{e}^{\mathrm{i}\gamma n x/\varepsilon}.$$

Since integral (10) is linear with respect to $p(x)$, the solution can be found in the form

$$u(x, z) = \sum_{n=-\infty}^{\infty} u_n(x, z),$$

where each $u_n(x, z)$ is a solution for (10) p equal to $P_n(x)\mathrm{e}^{\mathrm{i}\gamma n x/\varepsilon}$. So, we can reduce our analysis to the case $p(x) = A(x)\mathrm{e}^{\mathrm{i}Bx/\varepsilon}$.

3.1. METHOD OF STATIONARY PHASE

The method of stationary phase [9] describes the asymptotic behavior of the integral

$$J = \int_a^b f(t) \mathrm{e}^{\mathrm{i}\chi g(t)} \mathrm{d}t,$$

where χ is a large positive real parameter. We review the basic assumptions and the approximation properties of the method.

3.1.1. *Assumptions*

i. $-\infty < a < b \le \infty$.

ii. In (a,b), $g^{(m+1)}(t)$, $f^{(m)}(t)$ are continuous, $m \ge 0$, $g(t)$ is real-valued, $g'(t) > 0$, $g(b-0) < \infty$.

iii. As $t \to a+$,

$$f(t) \sim \sum_{s=0}^{\infty} f_s (t-a)^s, \qquad g(t) \sim g(a) + \sum_{s=0}^{\infty} g_s (t-a)^{s+\mu},$$

where f_0, g_0 are nonzero, $\mu > 0$. Moreover, the first of these expansions is differentiable m times and the second $m+1$ times.

iv. Let $P_s(t) = \left\{ \frac{1}{g'(t)} \frac{\mathrm{d}}{\mathrm{d}t} \right\}^s \frac{f(t)}{g'(t)}$. Then $\lim_{t \to b-} P_s(t) < \infty$, $s = 0, 1, \ldots, m$.

v. a, b, f, and g are independent of χ.

3.1.2. *Main results*

Let

i. $n \ge 0$, $m\mu - 1 \le n < (m+1)\mu$, and $\nu = \max(n, m\mu)$,

ii. $v = g(t) - g(a)$, and $F(v) = P_0(t)$.

For small v, $F(v)$ can be expanded in asymptotic series of the form

$$F(v) \sim \sum_{s=0}^{\infty} a_s v^{(s+1-\mu)/\mu},$$

where the coefficients a_s may be found by reverting series, in particular,

$$a_0 = \frac{f_0}{\mu g_0^{1/\mu}}, \qquad a_1 = \left\{ \frac{f_1}{\mu} - \frac{2 g_1 f_0}{\mu^2 g_0} \right\} \frac{1}{g_0^{2/\mu}}.$$

Then

$$J = \mathrm{e}^{\mathrm{i}\chi g(a)} \sum_{s=1}^{\nu} \mathrm{e}^{\mathrm{i}\frac{s\pi}{2\mu}} \Gamma\left(\frac{s}{\mu}\right) \frac{a_{s-1}}{\chi^{s/\mu}} - \mathrm{e}^{\mathrm{i}\chi g(b)} \sum_{s=1}^{m} P_{s-1}(b) \left(\frac{\mathrm{i}}{\chi}\right)^s + \mathrm{err}. \tag{11}$$

The error term is given by

$$\mathrm{err} = \left(\frac{\mathrm{i}}{\chi}\right)^m \int_a^b \mathrm{e}^{\mathrm{i}\chi g(t)} Q'_{m,n}(t)\mathrm{d}t + 0\left(\frac{1}{\chi^{m+1}}\right),$$

where

$$Q_{m,n}(t) = P_{m-1}(t) - \sum_{s=1}^{n} \frac{\Gamma\left(\frac{s}{\mu}\right)}{\Gamma\left(\frac{s}{\mu} - m + 1\right)} \frac{a_{s-1}}{\{g(t) - g(a)\}^{m-s/\mu}}.$$

The method of stationary phase is mostly known [5] by its first term that describes the leading behavior of the integral J for $\mu = 2$

$$J = \int f(t)\mathrm{e}^{\mathrm{i}\chi g(t)}\mathrm{d}t \sim f(t_0)\mathrm{e}^{\mathrm{i}\chi g(t_0)} \sqrt{\frac{2\pi\mathrm{i}}{\chi g''(t_0)}}, \qquad \chi \to \infty,$$

where t_0 is the stationary point of $g(t)$, $g''(t_0) > 0$.

3.2. Application to Kirchhoff approximation

Recall, that we investigate integral (10) with $p(x) = A(x)\mathrm{e}^{\mathrm{i}Bx/\varepsilon}$

$$u(x,z) = \frac{z}{\sqrt{2\pi \mathrm{i}\varepsilon}} \int_{-1}^{1} \frac{A(t)}{r^{3/2}} \mathrm{e}^{\mathrm{i}(Bt+r)/\varepsilon} \mathrm{d}t,$$

where $r = \sqrt{(x-t)^2 + z^2}$, $\varepsilon \ll 1$.

Let $A(x)$ be continuous and twice differentiable on $(-1, 1)$. Apply the method of stationary phase (11) to the integral

$$J = \int_{-1}^{1} \frac{A(t)}{r^{3/2}} \mathrm{e}^{\mathrm{i}(Bt+r)/\varepsilon} \mathrm{d}t,$$

with

$$f(t) = \frac{A(t)}{r^{3/2}}, \qquad g(t) = Bt + r, \qquad P_0(t) = \frac{A(t)}{r^{1/2}(Br + t - x)}, \qquad \chi = \frac{1}{\varepsilon}.$$

Since $g'(t) = B + \frac{t-x}{r}$, the stationary point t_0 exists (and is unique) iff $|B| < 1$,

$$t_0 = x - \frac{B}{\beta} z, \quad \text{and} \quad \beta = \sqrt{1 - B^2}.$$

In this case

$$f(t_0) = A(t_0)\left(\frac{\beta}{z}\right)^{3/2}, \qquad g(t_0) = Bx + \beta z, \qquad g''(t_0) = \frac{\beta^3}{z}.$$

Consider 4 possible cases regarding the stationary point t_0.

3.2.1. *Case I*

$|B| < 1$, $|t_0| < 1$, *i.e.* the stationary point t_0 is an inside point of the interval.

Apply the method of stationary phase for the intervals $(-1, t_0)$ and $(t_0, 1)$ with parameters

$$m = 2, \qquad \mu = 2, \qquad n = 3, \qquad \nu = 4,$$

$$g_0 = \frac{1}{2} g''(t_0), \qquad f_0 = f(t_0), \qquad a_0 = \frac{f_0}{2\sqrt{g_0}} = \frac{A(t_0)}{\sqrt{2z}}.$$

Then we get, as a final result,

$$u(x,z) = A(t_0)\mathrm{e}^{\mathrm{i}g(t_0)/\varepsilon} - \sqrt{\varepsilon}\frac{z\mathrm{e}^{\mathrm{i}\pi/4}}{\sqrt{2\pi}}\Big[P_0(1)\mathrm{e}^{\mathrm{i}g(1)/\varepsilon} - P_0(-1)\mathrm{e}^{\mathrm{i}g(-1)/\varepsilon}\Big] + \mathrm{O}(\varepsilon),$$

or in terms of x and z,

$$u(x,z) = A\left(x - \frac{B}{\beta} z\right)\mathrm{e}^{\mathrm{i}(Bx+\beta z)/\varepsilon} + \sqrt{\varepsilon}\Big[Q(x,z;1) - Q(x,z;-1)\Big] + \mathrm{O}(\varepsilon), \tag{12}$$

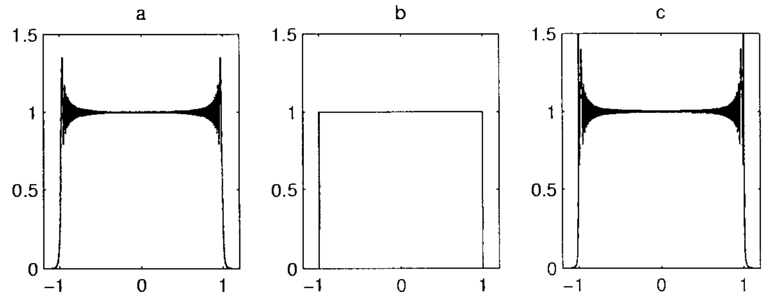

Figure 5. Numerical calculations of $|u(x,z)|$ (for $A(t) \equiv 1$, $B = 0$, $\varepsilon = 2 \times 10^{-3}$, $z = 0{\cdot}1$): (a) the exact integral, (b) O(1) approximation; (c) $O(\sqrt{\varepsilon})$ approximation.

where

$$Q(x,z;t) = -\frac{z\mathrm{e}^{\mathrm{i}\pi/4}}{\sqrt{2\pi}} \frac{A(t)}{r^{1/2}(Br + t - x)} \mathrm{e}^{\mathrm{i}(Bt+r)/\varepsilon}, \tag{13}$$

$$r = \sqrt{(x-t)^2 + z^2}.$$

Note that $Q(x,z;\pm 1)$ terms introduce disturbance due to the endpoints $t = \pm 1$ of the interval of integration.

Since $A(t)$ is defined and continuous only on $(-1,1)$, $A(1)$ and $A(-1)$, in (12), (13), are assumed in the sense of the limiting values $\lim_{t\to 1^-} A(t)$ and $\lim_{t\to -1^+} A(t)$.

3.2.2. *Case II*

$|B| < 1$, $t_0 = \pm 1$, *i.e.* the stationary point t_0 is an endpoint of the interval.

Apply the method of stationary phase for the interval $(-1,1)$ with the same parameters as in Case I. Then we get

$$u(x,z) = \frac{1}{2} A(t_0) \mathrm{e}^{\mathrm{i}g(t_0)/\varepsilon} \mp \sqrt{\varepsilon} \frac{z\mathrm{e}^{\mathrm{i}\pi/4}}{\sqrt{2\pi}} \left[\frac{f_1}{2g_0} \mathrm{e}^{\mathrm{i}g(t_0)/\varepsilon} - P_0(\mp 1) \mathrm{e}^{\mathrm{i}g(\mp 1)/\varepsilon} \right] + O(\varepsilon),$$

or, in terms of x and z, (note that here $x = \pm 1 + \frac{B}{\beta} z$)

$$u(x,z) = \frac{1}{2} A(\pm 1) \mathrm{e}^{\mathrm{i}(Bx+\beta z)/\varepsilon} \pm \sqrt{\varepsilon} \left[\widehat{Q}(x,z;\pm 1) - Q(x,z;\mp 1) \right] + O(\varepsilon), \tag{14}$$

where

$$\widehat{Q}(x,z;t) = -\frac{z\mathrm{e}^{\mathrm{i}\pi/4}}{\sqrt{2\pi}} \left[\frac{3}{2} \frac{A(t)B}{\sqrt{\beta z^3}} + \frac{A'(t)}{\sqrt{\beta^3 z}} \right] \mathrm{e}^{\mathrm{i}(Bt+r)/\varepsilon}.$$

Again, $A(t)$, $A'(t)$ at endpoints $t = \pm 1$ are assumed in the sense of the limiting values from inside of the interval.

3.2.3. *Case III*

$|B| < 1$, $|t_0| > 1$, *i.e.* the stationary point t_0 is outside of the interval.

Then, on $[-1,1]$ $g(t) > 0$ (or < 0). Apply the method of stationary phase for the interval $(-1,1)$ with

$$m = 1, \quad \mu = 1, \quad n = 0, \quad \nu = 0,$$

$$g_0 = g'(\mp 1), \qquad f_0 = f(\mp 1), \qquad a_0 = \frac{f_0}{g_0}.$$

Then

$$u(x,z) = \pm\sqrt{\varepsilon}\Big[Q(x,z;1) - Q(x,z;-1)\Big] + \mathrm{O}(\varepsilon^{3/2}). \tag{15}$$

3.2.4. *Case IV*
$|B| \geq 1$, *i.e.* there are no stationary points.

Apply the method of stationary phase for the interval $(-1, 1)$ with the same parameters as in Case III. Then

$$u(x,z) = \pm\sqrt{\varepsilon}\Big[Q(x,z;1) - Q(x,z;-1)\Big] + \mathrm{O}(\varepsilon^{3/2}). \tag{16}$$

3.2.5. *Leading-order approximation.*
Extend $A(x)$ on $(-\infty, \infty)$ as follows

$$A(-1) = \frac{1}{2}\lim_{t\to -1^+} A(t), \qquad A(1) = \frac{1}{2}\lim_{t\to 1^-} A(t), \qquad A(x) \equiv 0 \forall |x| > 1.$$

Then, taking into consideration only $\mathrm{O}(1)$ terms in (12)–(16), we get

$$u(x,z) \approx \begin{cases} A\left(x - \frac{B}{\beta}z\right)\mathrm{e}^{\mathrm{i}(Bx+\beta z)/\varepsilon}, & \text{if } |B| < 1 \\ 0, & \text{if } |B| \geq 1 \end{cases} \qquad \text{for } \varepsilon \ll 1. \tag{17}$$

3.2.6. *Edge effect*
Note that $O(1)$ approximation [17] is valid as long as $\mathrm{O}(\sqrt{\varepsilon})$ and higher terms are negligible. However, it does not describe a so-called *edge effect* observed in the exact integral (10) near the endpoints $x = \pm 1$ of the interval. Including $\mathrm{O}(\sqrt{\varepsilon})$ terms gives better approximation to that phenomena, see Figure 5.

To examine the effect of including $\mathrm{O}(\sqrt{\varepsilon})$ terms, consider in detail $Q(x, z; \pm 1)$. Note that it has a singularity at $x_{\pm 1} = \pm 1 + \frac{B}{\beta}z$ (when denominator equals 0). In particular,

$$|Q(x,z;\pm 1)| \sim \mathrm{O}\left(\frac{\sqrt{z}}{|x - x_{\pm 1}|}\right), \qquad \text{as } |x - x_{\pm 1}| \to 0.$$

We can summarize our findings as follows.

(i) For $|x - x_{\pm 1}| \sim \mathrm{O}(\sqrt{z})$ or more, the edge effect can be neglected and approximation (17) is valid.

(ii) For $|x - x_{\pm 1}| \sim \mathrm{O}(\sqrt{\varepsilon z})$, the edge effect cannot be neglected and $\mathrm{O}(\sqrt{\varepsilon})$ terms should be included into our approximation

$$u(x,z) \approx A\left(x - \frac{B}{\beta}z\right)\mathrm{e}^{\mathrm{i}(Bx+\beta z)/\varepsilon} \pm \sqrt{\varepsilon}\Big[Q(x,z;1) - Q(x,z;-1)\Big]. \tag{18}$$

(iii) For $|x - x_{\pm 1}| \ll \mathrm{O}(\sqrt{\varepsilon z})$, $\mathrm{O}(\sqrt{\varepsilon})$ terms do not describe the edge effect accurately. Indeed they blow up at $x \to x_{\pm 1}$.

(iv) At $x = x_{\pm 1}$, we use (14) as an approximation.

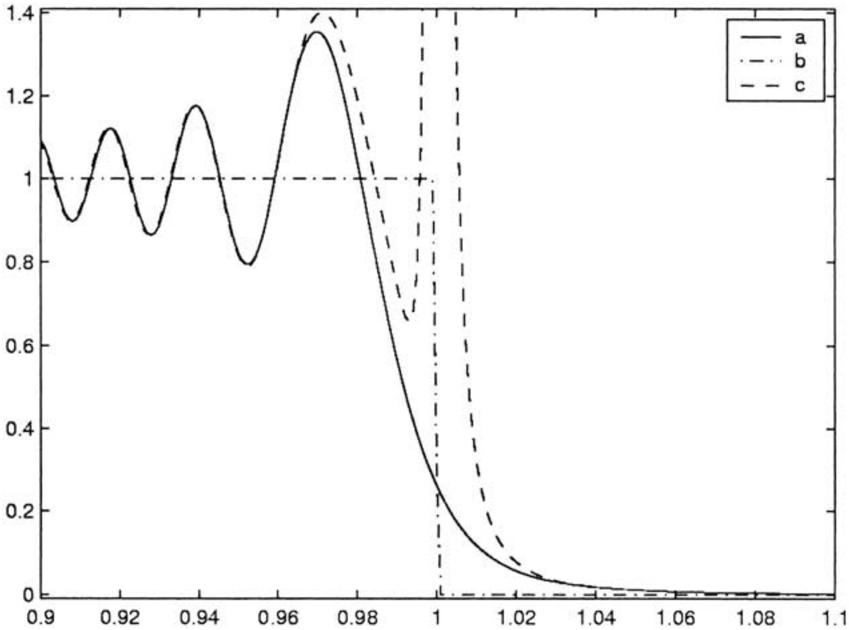

Figure 6. The edge effect in detail (for $A(t) \equiv 1$, $B = 0$, $\varepsilon = 2 \times 10^{-3}$, $z = 0{\cdot}1$): (a) the exact integral, (b) $\mathrm{O}(1)$ approximation; (c) $\mathrm{O}(\sqrt{\varepsilon})$ approximation.

See Figure 6 for a detailed picture of the exact integral and its approximations in the neighborhood of the endpoint $x = 1$. We can see that $\mathrm{O}(\sqrt{\varepsilon})$ approximation (18) gives a good approximation to the edge effect, except when $|x - 1| \leq 0{\cdot}03 \approx 2\sqrt{\varepsilon z}$, which agrees with the above conclusions. A different expansion is needed in the neighborhood of the edge.

3.3. General two-scale phases

Consider now a two-scale $p(x)$ which can be expanded as

$$p(x) = \sum_{n=-\infty}^{\infty} P_n(x) \mathrm{e}^{\mathrm{i}n\gamma x/\varepsilon}.$$

The solution for u can be found in the form

$$u(x, z) = \sum_{n=-\infty}^{\infty} u_n(x, z),$$

where each $u_n(x, z)$ is a Kirchhoff approximation (10) with the phase $P_n(x)\mathrm{e}^{\mathrm{i}\gamma n x/\varepsilon}$. Using the first-order approximation in (17), we can write

$$u_n(x, z) \approx P_n\left(x - \frac{n\gamma}{\beta_n} z\right) \mathrm{e}^{\mathrm{i}(n\gamma x + \beta_n z)/\varepsilon},$$

where $\beta_n = \sqrt{1 - (n\gamma)^2}$. Note that for $n > 1/\gamma$, β_n is imaginary and $u_n(x, z)$ corresponds to evanescent waves. Because of their exponential decay in z, we will simply ignore their contribution. Therefore, for $\varepsilon \ll 1$,

$$u(x, z) \approx \sum_{|n|<1/\gamma} P_n\left(x - \frac{n\gamma}{\beta_n} z\right) \mathrm{e}^{\mathrm{i}(n\gamma x + \beta_n z)/\varepsilon}. \tag{19}$$

Note that we are interested in the *near field* of the boundary $z = 0$ (where the z-range is small in comparison to the x-range). For small $|n|z$, we can simplify (19) to

$$u(x, z) \approx \sum_{|n|<1/\gamma} P_n(x) \mathrm{e}^{\mathrm{i}n\gamma x/\varepsilon} \mathrm{e}^{\mathrm{i}\beta_n z/\varepsilon}.$$

Going back to the original, 'physical', variables, which were rescaled in (9), we get the following result. Let $p(x)$ have representation

$$p(x) = P\left(\frac{x}{w}, \frac{x}{\Lambda}\right) = \sum_{n=-\infty}^{\infty} P_n\left(\frac{x}{w}\right) \mathrm{e}^{\mathrm{i}2\pi nx/\Lambda},$$

where

$$P_n(x) = \int_0^1 P(x, y)\mathrm{e}^{-\mathrm{i}2\pi ny}\mathrm{d}y.$$

Then

$$u(x, z) \approx \sum_{|n|<1/\gamma} P_n\left(\frac{x}{w}\right) \mathrm{e}^{\mathrm{i}2\pi nx/\Lambda} \mathrm{e}^{\mathrm{i}\beta_n z},$$

where $\beta_n = k\sqrt{1 - (n\gamma)^2}$ and the intensity is given by

$$\begin{aligned} I(x, z) &= |u(x, z)|^2 \left[\sum_n P_n \mathrm{e}^{\mathrm{i}2\pi nx/\Lambda} \mathrm{e}^{\mathrm{i}\beta_n z}\right] \left[\sum_m P_m^* \mathrm{e}^{-\mathrm{i}2\pi mx/\Lambda} \mathrm{e}^{-\mathrm{i}\beta_m z}\right] = \\ &= \underbrace{\sum_n |P_n|^2 + \sum_{n\neq 0} P_n P_{-n}^* \mathrm{e}^{\mathrm{i}4\pi nx/\Lambda}}_{I_1(x)} + \underbrace{\sum\sum_{|n|\neq|m|} P_n P_m^* \mathrm{e}^{\mathrm{i}2\pi(n-m)x/\Lambda} \mathrm{e}^{\mathrm{i}(\beta_n-\beta_m)z}}_{I_2(x,z)}. \end{aligned}$$

That is, we can separate $I(x, z)$ into two parts: $I_1(x)$, which does not depend on z and $I_2(x, z)$, which contains only terms periodic in z. This may help us to simplify the expression for $I(x, z)$. Consider two special cases:

3.3.1. *Case A*

$P_{-n} = P_n$. Then

$$\begin{aligned} u(x, z) &= P_0 \mathrm{e}^{\mathrm{i}\beta_0 z} + 2\sum_{n>0} P_n \cos\left(2\pi n \frac{x}{\Lambda}\right) \mathrm{e}^{\mathrm{i}\beta_n z}, \\ I_1(x) &= |P_0|^2 + 4\sum_{n>0} |P_n|^2 \cos^2\left(2\pi n \frac{x}{\Lambda}\right), \\ I_2(x, z) &= 4\sum_{n>0} \sin\left(2\pi n \frac{x}{\Lambda}\right) \mathfrak{Re}[P_0 P_n^* \mathrm{e}^{\mathrm{i}(\beta_0-\beta_n)z}] + \\ &\quad 8\sum\sum_{0<n<m} \sin\left(2\pi n \frac{x}{\Lambda}\right) \cos\left(2\pi m \frac{x}{\Lambda}\right) \mathfrak{Re}[P_n P_m^* \mathrm{e}^{\mathrm{i}(\beta_n-\beta_m)z}]. \end{aligned} \tag{20}$$

3.3.2. *Case B*

$P_{-n} = -P_n, n \neq 0$. Then

$$\begin{aligned} u(x,z) &= P_0 e^{i\beta_0 z} + 2i \sum_{n>0} P_n \sin\left(2\pi n \frac{x}{\Lambda}\right) e^{i\beta_n z}, \\ I_1(x) &= |P_0|^2 + 4 \sum_{n>0} |P_n|^2 \sin^2\left(2\pi n \frac{x}{\Lambda}\right), \\ I_2(x,z) &= 4 \sum_{n>0} \cos\left(2\pi n \frac{x}{\Lambda}\right) \mathfrak{Im}[P_0 P_n^* e^{i(\beta_0-\beta_n)z}] + \\ & \quad 8 \sum_{0<n<m}\sum \cos\left(2\pi n \frac{x}{\Lambda}\right) \sin\left(2\pi m \frac{x}{\Lambda}\right) \mathfrak{Re}[P_n P_m^* e^{i(\beta_n-\beta_m)z}]. \end{aligned} \tag{21}$$

3.4. AVERAGING THE INTENSITY OVER THE z-INTERVAL

In practice, we may be interested in the intensity not at a single plane $z = z_0$ but rather in the average intensity $I_{avr}(x)$ over some interval in $z \in (z_0 - \frac{\Delta z}{2}, z_0 + \frac{\Delta z}{2})$

$$I_{avr}(x) = \frac{1}{\Delta z} \int_{z_0-\frac{\Delta z}{2}}^{z_0+\frac{\Delta z}{2}} I(x,z) dz = I_1(x) + \hat{I}_2(x; z_0, \Delta z), \tag{22}$$

where

$$\hat{I}_2(x; z_0, \Delta z) = \frac{1}{\Delta z} \int_{z_0-\frac{\Delta z}{2}}^{z_0+\frac{\Delta z}{2}} I_2(x,z) dz.$$

We next investigate situations under which the z-dependent term, *i.e.* $\hat{I}_2(x; z_0, \Delta z)$ can be neglected. Note that

$$\hat{I}_2(x; z_0, \Delta z) = \sum_{|n|\neq|m|}\sum P_n P_m^* e^{i2\pi(n-m)x/\Lambda} \mathcal{M}_{nm},$$

where

$$\mathcal{M}_{nm} = \frac{1}{\Delta z} \int_{z_0-\frac{\Delta z}{2}}^{z_0+\frac{\Delta z}{2}} e^{i(\beta_n-\beta_m)z} dz = e^{i(\beta_n-\beta_m)z_0} \text{sinc}\left[\frac{(\beta_n-\beta_m)\Delta z}{2\pi}\right],$$

and $\text{sinc}(x) = \sin(\pi x)/(\pi x)$. Since

$$|\mathcal{M}_{nm}| \leq \left|\frac{2}{(\beta_n-\beta_m)\Delta z}\right| \leq \frac{4}{k\gamma^2 \Delta z |m^2-n^2|}, \qquad \forall m,n, |m| \neq |n|,$$

then

$$|\hat{I}_2(x; z_0, \Delta z)| \leq \frac{4}{k\gamma^2\Delta z} \sum_{|n|\neq|m|}\sum \frac{|P_n P_m|}{|m^2-n^2|}.$$

So, we can see that $|\hat{I}_2(x; z_0, \Delta z)| \to 0$ as $\Delta z \to \infty$.

Further simplifications are possible if we restrict to the case when $P_n = P_{-n}$. Then, from (20),

$$\hat{I}_2(x; z_0, \Delta z) = 4\sum_{n>0} \cos\left(2\pi n \frac{x}{\Lambda}\right) \mathfrak{Re}[P_0 P_n^* \mathcal{M}_{0n}] + 8\sum\sum_{0<n<m} \cos\left(2\pi n \frac{x}{\Lambda}\right) \cos\left(2\pi m \frac{x}{\Lambda}\right) \mathfrak{Re}[P_n P_m^* \mathcal{M}_{nm}],$$

and

$$|\hat{I}_2(x; z_0, \Delta z)| \le 4\sum_{n>0} |P_0 P_n \mathcal{M}_{n0}| + 8\sum\sum_{0<n<m} |P_n P_m \mathcal{M}_{nm}|.$$

Note that, since $P_n \to 0$ as $n \to \infty$, the biggest contribution to $\hat{I}_2(x; z_0, \Delta z)$ comes from the first few nonzero terms, in particular $[P_0 P_1^* \mathcal{M}_{01}]$, provided that $|P_0 P_1| \neq 0$. Since, for small γ,

$$\frac{2\pi}{\beta_0 - \beta_1} = \frac{\Lambda^2}{\lambda}(1 + \sqrt{1 - \gamma^2}) \approx \frac{2\Lambda^2}{\lambda} = T \text{ (the Talbot distance [10])},$$

then

$$\mathcal{M}_{01} \approx \frac{T}{\pi \Delta z} e^{i2\pi z_0/T} \sin\left(\pi \frac{\Delta z}{T}\right), \qquad |\mathcal{M}_{nm}| \le \frac{T}{\pi \Delta z} \frac{1}{|m^2 - n^2|}, \quad \forall |m| \neq |n|.$$

Rewrite

$$\hat{I}_2(x; z_0, \Delta z) = \frac{4}{\pi} \frac{T}{\Delta z} \left\{ \cos\left(2\pi \frac{x}{\Lambda}\right) \sin\left(\pi \frac{\Delta z}{T}\right) \mathfrak{Re}[P_0 P_1^* e^{i2\pi z_0/T}] + Q \right\}, \tag{23}$$

where

$$|Q| \le \sum_{n>1} \frac{|P_0 P_n|}{n^2} + 2\sum\sum_{0<n<m} \frac{|P_n P_m|}{m^2 - n^2}.$$

Now we can see that the expected biggest contribution to $\hat{I}_2(x; z_0, \Delta z)$ is minimized if either $\mathfrak{Re}[P_0 P_1^* e^{i2\pi z_0/T}] \approx 0$ or $\Delta z/T =$ integer *i.e.* the z-interval is a multiple of the Talbot distance T). However, if $|P_0 P_1| = 0$ or ≈ 0, other nonzero terms should be taken into consideration to estimate $\hat{I}_2(x; z_0, \Delta z)$.

The averaged intensity is to first approximation independent of z. When conditions do not allow us to ignore the z-dependence, an additional term is needed. In either case, the intensity is rapidly oscillatory with period $\Lambda/2$. In summary, we have

3.4.1. *First approximation (independent of z)*

If we can neglect $\hat{I}_2(x; z_0, \Delta z)$, (22) simplifies greatly to

$$I_{\text{avr}}(x) \approx I_1(x), \tag{24}$$

which does not contain direct dependence on th z-interval and is periodic with respect to the 'fast' variable with period $\frac{\Lambda}{2}$.

3.4.2. *Second approximation (z-dependent)*

If we cannot neglect $\hat{I}_2(x; z_0, \Delta z)$, then approximate it by the first, supposedly the biggest, term, *e.g.* in case $P_n = P_{-n}$, $P_0 P_1 \neq 0$,

$$I_{\text{avr}}(x) \approx I_1(x) + \frac{4}{\pi} \frac{T}{\Delta z} \cos\left(2\pi \frac{x}{\Lambda}\right) \sin\left(\pi \frac{\Delta z}{T}\right) \mathfrak{Re}[P_0 P_1^* e^{i2\pi z_0/T}].$$

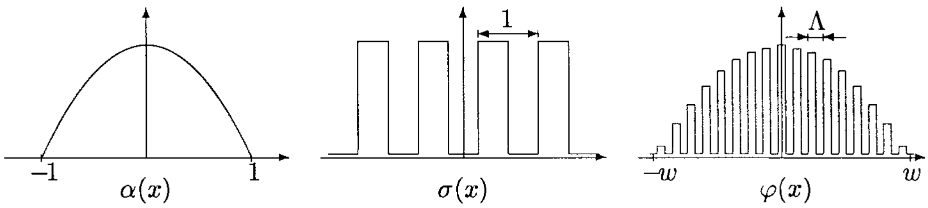

Figure 7. The phase function considered in the design of DOEs.

4. Design of diffractive optical elements

In this section we consider DOEs with quasi-periodic phase

$$p(x) = \mathrm{e}^{\mathrm{i}\varphi(x)}, \qquad \varphi(x) = \alpha\left(\frac{x}{w}\right)\sigma\left(\frac{x}{\Lambda}\right),$$

where $\frac{\Lambda}{w} \ll 1$, $\sigma(\cdot)$ is periodic with period 1; $\alpha(\cdot)$ is called an 'envelope' function, such as one shown in Figure 4. We can represent $p(x)$ as

$$p(x) = P\left(\frac{x}{w}, \frac{x}{\Lambda}\right) = \sum_{n=-\infty}^{\infty} P_n\left(\frac{x}{w}\right)\mathrm{e}^{\mathrm{i}2\pi n\frac{x}{\Lambda}},$$

where

$$P_n(X) = \int_0^1 \mathrm{e}^{\mathrm{i}\alpha(X)\sigma(Y)}\mathrm{e}^{-\mathrm{i}2\pi nY}\mathrm{d}Y. \tag{25}$$

In order to use the approximation formulas from the previous section, we need to know the coefficients P_n for $|n| < \frac{1}{\gamma}$. In general, they can be computed numerically by use of (25). We consider a few cases when they can be found analytically, such as in stepwise and sinusoidal structures. These simple periodic structure allow us to examine the averaged intensity in detail.

4.1. First approximation (independent of z)

Assume first that the dependence on z can be neglected so that the first approximation $I_{\mathrm{avr}}(x) \approx I_1(x)$ is valid.

4.1.1. *Piecewise constant DOEs with variable duty cycle*

Define $\sigma(x)$ on the interval $(-\frac{1}{2}, \frac{1}{2})$ as

$$\sigma(x) = \begin{cases} 1, & |x| < \theta/2, \\ 0, & |x| > \theta/2, \end{cases}$$

where $\theta \in [0, 1]$ is called the *duty cycle*. Then

$$P_n(X) = \begin{cases} 1 + \theta(\mathrm{e}^{\mathrm{i}\alpha} - 1), & n = 0, \\ \theta(\mathrm{e}^{\mathrm{i}\alpha} - 1)\mathrm{sinc}(n\theta), & n \neq 0, \end{cases} \tag{26}$$

where $\alpha = \alpha(x/w)$. Note that $P_{-n} = P_n$. Then, from (20), (24),

$$I_{\mathrm{avr}}(x) \approx 1 - 4\theta(1-\theta)\sin^2\left(\frac{\alpha}{2}\right) + \sin^2\left(\frac{\alpha}{2}\right)\Psi\left(\frac{x}{\Lambda}\right), \tag{27}$$

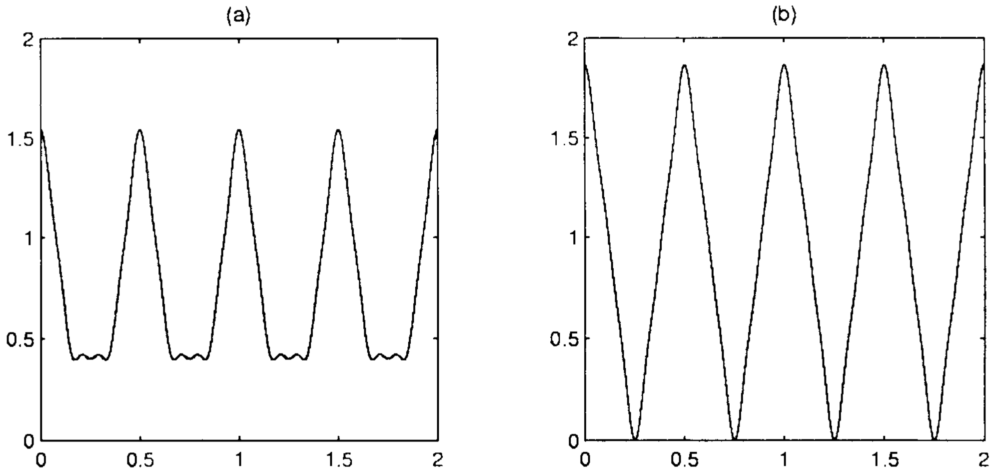

Figure 8. Function $\Psi(Y)$ for $\gamma = 0{\cdot}157$, (a) $\theta = 0{\cdot}3$, (b) $\theta = 0{\cdot}5$.

where

$$\Psi(Y) = \frac{16}{\pi^2} \sum_{1 \le n < \frac{1}{\gamma}} \frac{\sin^2(\pi n \theta)}{n^2} \cos^2(2\pi n Y). \tag{28}$$

An example of $\Psi(Y)$ is shown in Figure 8(a).

In (27), we can see that the intensity consists of a rapid oscillation, represented by $\Psi\left(\frac{x}{\Lambda}\right)$, within an envelope. The upper and lower curves of the envelope, $I_{\text{upper}}(x)$ and $I_{\text{lower}}(x)$, are given by

$$I_{\text{upper}}(x) = 1 - 4\theta(1-\theta)\sin^2\left(\frac{\alpha}{2}\right) + \sin^2\left(\frac{\alpha}{2}\right)\max_Y \Psi(Y),$$

$$I_{\text{lower}}(x) = 1 - 4\theta(1-\theta)\sin^2\left(\frac{\alpha}{2}\right) + \sin^2\left(\frac{\alpha}{2}\right)\min_Y \Psi(Y).$$

4.1.2. *Piecewise constant DOEs with duty cycle = 1/2*

Taking $\theta = 1/2$ in (26)–(28) we have

$$P_n(X) = \begin{cases} (e^{i\alpha} + 1)/2, & n = 0, \\ (-1)^{\frac{n-1}{2}} \frac{e^{i\alpha} - 1}{\pi n}, & n \text{ odd}, \\ 0, & n \text{ even}, \neq 0, \end{cases} \tag{29}$$

and

$$I_{\text{avr}}(x) \approx 1 + \sin^2\frac{\alpha}{2}\left(\Psi\left(\frac{x}{\Lambda}\right) - 1\right), \tag{30}$$

where

$$\Psi(Y) = \frac{16}{\pi^2} \sum_{\substack{n \text{ odd} \\ 1 \le n < \frac{1}{\gamma}}} \frac{1}{n^2} \cos^2(2\pi n Y),$$

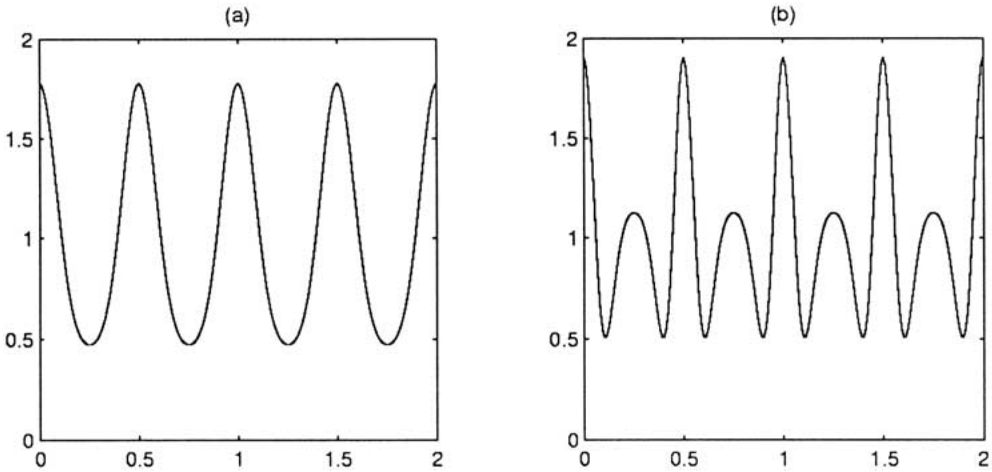

Figure 9. $I_{avr}(x)$ for $\gamma = 0{\cdot}157$, (a) $\alpha(x) \equiv \pi$, (b) $\alpha(x) \equiv 2\pi$.

see Figure 8(b), for an example. Since

$$\min_Y \Psi(Y) = 0, \qquad \max_Y \Psi(Y) = \frac{16}{\pi^2} \sum_{\substack{n \text{ odd} \\ 1 \le n < \frac{1}{\gamma}}} \frac{1}{n^2} = A,$$

the upper and lower curves of the envelope of $I_{avr}(x)$ are

$$I_{upper}(x) = 1 + (A-1)\sin^2\frac{\alpha}{2}, \qquad I_{lower}(x) = 1 - \sin^2\frac{\alpha}{2}. \tag{31}$$

Note that since $A = A(\gamma) < 2$, the envelope is slightly asymmetric. However, $A(\gamma) \to 2$ as $\gamma \to 0$.

4.1.3. *Sinusoidal DOEs*

In this case, $\sigma(x) = \sin^2(\pi x)$, and

$$P_n(X) = (-i)^n e^{i\alpha/2} J_n\left(\frac{\alpha}{2}\right), \tag{32}$$

where $J_n(\cdot)$ is the Bessel function of nth order. Since $J_{-n}(x) = (-1)^n J_n(x)$, $P_{-n} = P_n$ and, from (20), (24), we get

$$I_{avr}(x) \approx J_0^2\left(\frac{\alpha}{2}\right) + 4 \sum_{1 \le n < \frac{1}{\gamma}} J_n^2\left(\frac{\alpha}{2}\right) \cos^2\left(2\pi n \frac{x}{\Lambda}\right),$$

see Figure 9, for example, for $\alpha(x) \equiv \pi$ and $\alpha(x) \equiv 2\pi$.

So, the envelope of $I_{avr}(x)$ has upper and lower curves

$$I_{upper}(x) = J_0^2\left(\frac{\alpha}{2}\right) + 4 \sum_{1 \le n < \frac{1}{\gamma}} J_n^2\left(\frac{\alpha}{2}\right), \qquad I_{lower}(x) \approx J_0^2\left(\frac{\alpha}{2}\right) + 4 \sum_{1 \le n < \frac{1}{2\gamma}} J_{2n}^2\left(\frac{\alpha}{2}\right). \tag{33}$$

4.2. Second approximation (z-dependent)

To analyze in detail what happens to the intensity profile $I_{\text{avr}}(x)$, when the first approximation is no longer valid, consider two examples.

4.2.1. *Piecewise constant DOEs with duty cycle 1/2*

From (23), (29) we get

$$\hat{I}_2(x; z_0, \Delta z) \approx \frac{4}{\pi^2}\frac{T}{\Delta z}\left[\sin(\alpha)\cos\left(2\pi\frac{x}{\Lambda}\right)\sin\left(2\pi\frac{z_0}{T}\right)\sin\left(\pi\frac{\Delta z}{T}\right) + C\right]$$

where

$$|C| \leq 0{\cdot}2.$$

We can see that the effect of $\hat{I}_2(x; z_0, \Delta z)$ can be neglected and the first approximation (30) is valid only if at least one of the following conditions is satisfied

$$\begin{aligned} &\text{(a)}\quad \Delta z \gg T, \text{ or}\\ &\text{(b)}\quad \Delta z/T \in \mathbb{Z}, \text{ or}\\ &\text{(c)}\quad 2z_0/T \in \mathbb{Z}. \end{aligned} \tag{34}$$

For example, if Δz is predetermined by a particular application, $\hat{I}_2(x; z_0, \Delta z)$ still can be minimized by a proper centering of the z-interval (to satisfy condition (c)). In fact, if condition (c) is satisfied then Δz is not important, it can even be $= 0$, which means that the z-interval reduces to a point $z = z_0$ and $I_{\text{avr}}(x) \equiv I(x, z_0) \approx I_1(x)$. Otherwise, if neither of the conditions (34) is satisfied,

$$I_{\text{avr}}(x) \approx 1 + \sin^2\frac{\alpha}{2}\left(\Psi\left(\frac{x}{\Lambda}\right) - 1\right) + \sin(\alpha)\Phi\left(\frac{x}{\Lambda}\right), \tag{35}$$

where

$$\Psi(Y) = \frac{16}{\pi^2}\sum_{\substack{n \text{ odd}\\ 1\leq n<\frac{1}{\gamma}}}\frac{1}{n^2}\cos^2(2\pi n Y),$$

has period 1/2, and

$$\Phi(Y) = \frac{4}{\pi^2}\frac{T}{\Delta z}\sin\left(2\pi\frac{z_0}{T}\right)\sin\left(\pi\frac{\Delta z}{T}\right)\cos\left(2\pi Y\right),$$

has period 1. Expression (35) shows that the envelope of $I_{\text{avr}}(x)$ consists of three curves instead of the original two in (31)

$$\begin{aligned} I^{\pm}_{\text{up}}(x) &= 1 + (A-1)\sin^2\frac{\alpha}{2} \pm B\sin(\alpha),\\ I'_{\text{lower}}(x) &= 1 - \sin^2\frac{\alpha}{2}, \end{aligned} \tag{36}$$

where $B = \frac{4}{\pi^2}\frac{T}{\Delta z}\left|\sin\left(2\pi\frac{z_0}{T}\right)\sin\left(\pi\frac{\Delta z}{T}\right)\right|$.

The new envelope is defined by

$$I_{\text{upper}}(x) = \max_Y \left\{1 + \sin^2 \tfrac{\alpha}{2} (\Psi(Y) - 1) + \sin(\alpha)\Phi(Y)\right\} = I^+_{\text{up}}(x),$$

$$I_{\text{lower}}(x) = \min_Y \left\{1 + \sin^2 \tfrac{\alpha}{2} (\Psi(Y) - 1) + \sin(\alpha)\Phi(Y)\right\} \approx \min\left\{I^-_{\text{up}}(x), I'_{\text{lower}}(x)\right\}.$$

Note however that, if $\cos\left(\frac{\alpha}{2}\right) = 0$ or ≈ 0, $P_0 \approx 0$ and other nonzero terms in $\hat{I}_2(x; z_0, \Delta z)$ should be taken into consideration, *e.g.* $\mathfrak{Re}[P_1 P_3^* \mathcal{M}_{13}]$, to predict accurately $I_{\text{avr}}(x)$.

4.2.2. *Sinusoidal DOEs*
From (23), (32) we get

$$\hat{I}_2(x; z_0, \Delta z) \approx \frac{4}{\pi}\frac{T}{\Delta z}\left[-J_0\left(\frac{\alpha}{2}\right) J_1\left(\frac{\alpha}{2}\right) \cos\left(2\pi\frac{x}{\Lambda}\right) \sin\left(2\pi\frac{z_0}{T}\right) \sin\left(\pi\frac{\Delta z}{T}\right) + C\right],$$

where

$$|C| \leq \sum_{n>1} \frac{|J_0 J_n|}{n^2} + 2 \sum_{0<n<m} \frac{|J_n J_m|}{m^2 - n^2} \approx 0{\cdot}276.$$

Under the same conditions as in (34) $\hat{I}_2(x; z_0, \Delta z)$ can be neglected. Otherwise

$$I_{\text{avr}}(x) \approx J_0^2\left(\frac{\alpha}{2}\right) + 4 \sum_{1\leq n<\frac{1}{\gamma}} J_n^2\left(\frac{\alpha}{2}\right) \cos^2\left(2\pi n\frac{x}{\Lambda}\right) - BJ_0\left(\frac{\alpha}{2}\right) J_1\left(\frac{\alpha}{2}\right) \cos\left(2\pi\frac{x}{\Lambda}\right)$$

where

$$B = \frac{4}{\pi}\frac{T}{\Delta z}\left|\sin\left(2\pi\frac{z_0}{T}\right) \sin\left(\pi\frac{\Delta z}{T}\right)\right|.$$

4.3. Numerical results

To check how well the asymptotic formulas (31), (33), (36) approximate the envelope of the intensity, we calculated the exact intensity using (6), and overlayed the results with the approximate envelope. We ran numerical computations for the stepwise and sinusoidal profile DOEs with different envelope functions $\alpha(x)$. The physical parameters were

DOE length:	$2w = 100\mu$m
DOE period:	$\Lambda = 1\mu$m
wavelength:	$\lambda = 0{\cdot}15\mu$ m

so that $\gamma = \frac{\lambda}{\Lambda} = 0{\cdot}15$, $\varepsilon \approx 2{\cdot}4 \times 10^{-4}$.

To verify the **first approximation**, z_0, Δz were chosen to satisfy condition (34) ($z_0 = 6{\cdot}67\mu$m $\approx T/2$, $\Delta z = 12{\cdot}47\mu$m). Figures 10–13 show plots of intensity for stepwise and sinusoidal DOEs with different envelopes $\alpha(x)$ (linear, half-cosine, gaussian), where the 'exact solution' indicates integral (6), computed numerically, and the 'approximated envelope' is determined analytically by (31), (33). We can see that analytically predicted envelopes approximate quite well the exact solutions computed by numerical integration.

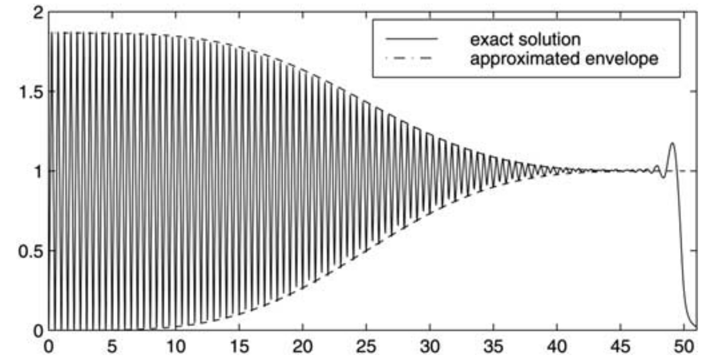

Figure 10. Comparison of the exact intensity and the approximate envelopes for a piecewise constant DOE with $\theta = \frac{1}{2},\ \alpha(x) = \pi\,(1 - |x|)$.

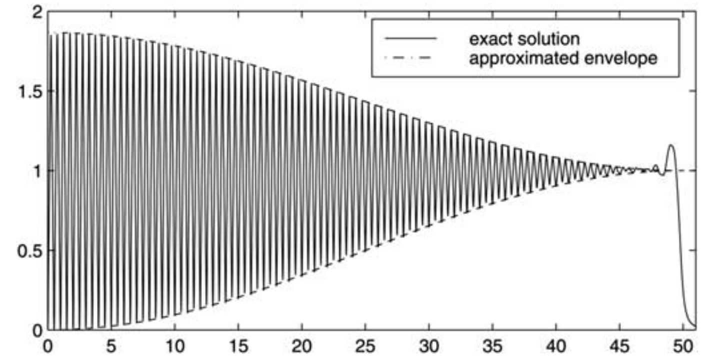

Figure 11. Comparison of the exact intensity and the approximate envelopes for a piecewise constant DOE with $\theta = \frac{1}{2},\ \alpha(x) = \frac{\pi}{2}(\cos(\pi x) + 1)$.

To verify the **second approximation**, z_0, Δz were chosen so that (34) is not satisfied ($z_0 = 10\mu\text{m} \approx 3T/4$, $\Delta z = 6{\cdot}27\mu\text{m} \approx T/2$) and the first approximation is no longer valid. Figure 14 shows that the 'split' envelope is well predicted by (36).

4.4. A DESIGN PROBLEM

To demonstrate how the inverse, or design, problem can be solved, consider a 'target' intensity profile to be a quasi-periodic function with an 'inside' periodic structure of period $\frac{\Lambda}{2}$ enclosed into a symmetric 'envelope' specified by $1 \pm f(x/w)$, where $0 \le f(x) \le 1$. Assuming that we can adjust z (the distance between the DOE and the image plane), we can use the first approximation for the piecewise constant DOE with the duty cycle $\theta = 1/2$ and period Λ given by

$$I(x) = 1 + \sin^2\frac{\alpha}{2}\left(\Psi\left(\frac{x}{\Lambda}\right) - 1\right), \qquad \alpha = \alpha\left(\frac{x}{w}\right),$$

which has the required period $= \frac{\Lambda}{2}$ and envelope

$$I_{\text{upper}} = 1 + (A - 1)\sin^2\frac{\alpha}{2},$$
$$I_{\text{lower}} = 1 - \sin^2\frac{\alpha}{2}.$$

To determine the envelope function $\alpha(x/w)$ for the DOE, solve the equation

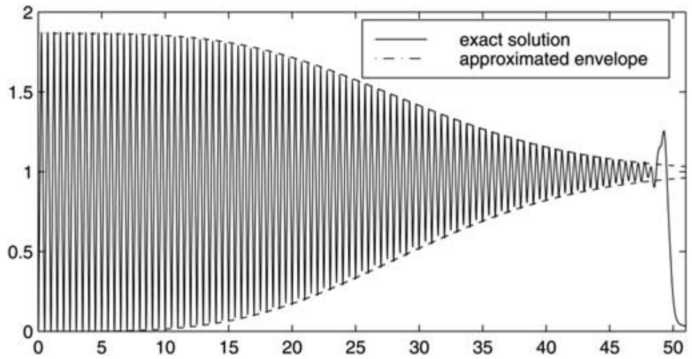

Figure 12. Comparison of the exact intensity and the approximate envelopes for a piecewise constant DOE with $\theta = \frac{1}{2},\ \alpha(x) = \pi e^{-2x^2}$.

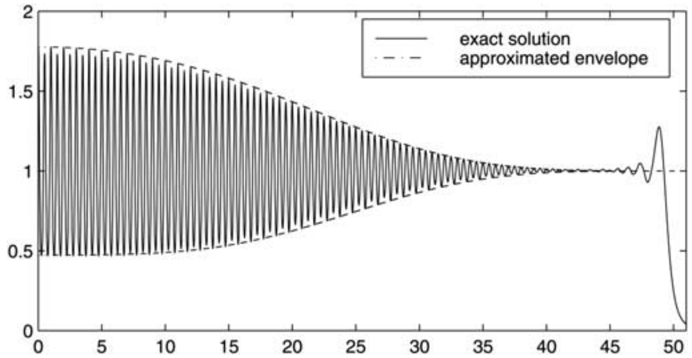

Figure 13. Comparison of the exact intensity and the approximate envelopes for a sinusoidal DOE with $\alpha(x) = \frac{\pi}{2}(\cos(\pi x) + 1)$.

$$1 - \sin^2 \frac{\alpha}{2} = 1 - f(x/w),$$

for α. We get that

$$\alpha(x) = 2 \arcsin \sqrt{f(x)}.$$

For example, if $f(x/w) = \cos^2\left(\frac{\pi x}{2w}\right)$ then $\alpha(x/w) = \pi\left(1 \pm \frac{x}{w}\right)$, *i.e.* linear function. Note, however, that the upper curve of the envelope in this case would be

$$I_{\text{upper}}(x) = 1 + (A - 1) f(x),$$

which is slightly different from $1 + f(x)$. As was shown before, the intensity envelope obtained by this kind of DOEs will always be slightly asymmetric due to the fact that $A = A(\gamma) < 2$.

The approach allows one to create a DOE which approximately meets the target intensity pattern by determining $\alpha(x)$ *directly* from the envelope of the intensity profile.

5. Discussion

In this work, we have described a procedure by which approximate expressions for the intensity of light caused by a DOE can be obtained. For DOEs which have two-scales, a rapid oscillation contained in a smooth envelope, we have obtained formulas which describe in

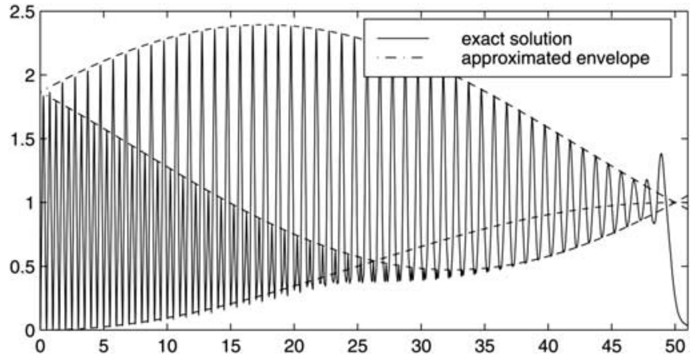

Figure 14. Comparison of the exact intensity and the approximate envelopes for a piecewise constant DOE with $\theta = \frac{1}{2},\ \alpha(x) = \pi\,(1 - |x|)$.

detail, and levels of accuracy, the intensity patterns on the image plane. The accuracy of the formulas were assessed in several examples. The framework can be used to design a DOE directly, given a desired target-intensity pattern.

While the asymptotic calculations were carried out in detail for a 2-D geometry, extension to 3-D is quite straightforward. Some results for 3-D DOEs were presented in Rudnaya's PhD thesis [11]. The key idea in this work is to rescale the problem so as to allow the use of the Method of Stationary Phase in evaluating the integral in the Kirchhoff approximation. Alternately, one can work directly with the rescaled Helmholtz equation and use the Method of Multiscales to solve the problem. We present this approach in the appendix.

The main open problem not addressed in this work is assessing the accuracy of the approximations. There are two sources of error, namely the Kirchhoff approximation and the asymptotic approximation that follows. It would be of great value to know how accurate the solutions we obtained are in comparison to the solution of the wave equation.

ACKNOWLEDGMENTS

This work was carried out while Svetlana Rudnaya was enrolled in graduate studies at the University of Minnesota. During this period, she also served as an intern for the 3M company under the supervision of David Misemer and Alessandra Chiareli. The collaboration was facilitated by the Minnesota Center for Industrial Mathematics. This research was supported in part by AFOSR through a MURI grant to the University of Delaware and by 3M Corporation.

Appendix. Multiscale analysis

We provide an alternate method by which the asymptotic formula (17) is obtained. We start with

$$\begin{cases} (\nabla^2 + k^2)u = 0, & -\infty < x < \infty, z > 0, \\ u(x,0) = p(x) = A(x)\mathrm{e}^{\mathrm{i}Bx/\varepsilon}, & -\infty < x < \infty. \end{cases} \tag{A-1}$$

which is the diffraction problem leading to the Kirchhoff approximation in (10).

Now we apply the multiple-scale analysis [12] to (A-1). Rescale the problem, using multiple scales in both x and z, we have

$$X_1 = x, \qquad X_2 = \frac{x}{\varepsilon}, \qquad Z_1 = z, \qquad Z_2 = \frac{z}{\varepsilon},$$
$$p(x) = A(X_1)\mathrm{e}^{\mathrm{i}BX_2}, \quad u(x,z) = U(X_1, X_2, Z_1, Z_2),$$
$$\frac{\partial^2 u}{\partial x^2} = U_{X_1X_1} + \frac{2}{\varepsilon}U_{X_1X_2} + \frac{1}{\varepsilon^2}U_{X_2X_2},$$
$$\frac{\partial^2 u}{\partial z^2} = U_{Z_1Z_1} + \frac{2}{\varepsilon}U_{Z_1Z_2} + \frac{1}{\varepsilon^2}U_{Z_2Z_2}. \tag{A-2}$$

Then (A-1) and (A-2) yield

$$\begin{cases} \varepsilon^2(U_{X_1X_1} + U_{Z_1Z_1}) + 2\varepsilon(U_{X_1X_2} + U_{Z_1Z_2}) + (U_{X_2X_2} + U_{Z_2Z_2} + U) = 0, \\ U(X_1, X_2, 0, 0) = A(X_1)\mathrm{e}^{\mathrm{i}BX_2}. \end{cases} \tag{A-3}$$

Expect $U(X_1, X_2, Z_1, Z_2)$ to be also periodic in X_2 in the form

$$U(X_1, X_2, Z_1, Z_2) = V(X_1, Z_1, Z_2)\mathrm{e}^{\mathrm{i}BX_2}. \tag{A-4}$$

Put (A-4) into (A-3) and get

$$\begin{cases} \varepsilon^2(V_{X_1X_1} + V_{Z_1Z_1}) + 2\varepsilon(\mathrm{i}BV_{X_1} + V_{Z_1Z_2}) + (V_{Z_2Z_2} + \beta^2 V) = 0, \\ V(X_1, 0, 0) = A(X_1), \end{cases} \tag{A-5}$$

where $\beta^2 = 1 - B^2$. Assume Taylor expansion of $V(X_1, Z_1, Z_2)$ in powers of ε

$$V = V_0 + \varepsilon V_1 + \varepsilon^2 V_2 + \ldots + \varepsilon^m V_m + \ldots, \tag{A-6}$$

Substitute (A-6) in (A-5) and collect powers of ε

$$\mathrm{O}(1): \qquad \begin{cases} (V_0)_{Z_2Z_2} + \beta^2 V_0 = 0 \\ V_0(X_1, 0, 0) = A(X_1) \end{cases}$$

$$\mathrm{O}(\varepsilon): \qquad \begin{cases} (V_1)_{Z_2Z_2} + \beta^2 V_1 = -2\mathrm{i}B(V_0)_{X_1} - 2(V_0)_{Z_1Z_2} \\ V_1(X_1, 0, 0) = 0 \end{cases}$$

$$\mathrm{O}(\varepsilon^2): \qquad \begin{cases} (V_2)_{Z_2Z_2} + \beta^2 V_2 = -2\mathrm{i}B(V_1)_{X_1} - 2(V_1)_{Z_1Z_2} - (V_0)_{X_1X_1} - (V_0)_{Z_1Z_1} \\ V_2(X_1, 0, 0) = 0 \end{cases}$$

and so on.

For $\beta^2 \neq 0$ ($|B| \neq 1$) the solution to $O(1)$ system is given by

$$V_0(X_1, Z_1, Z_2) = C(X_1, Z_1)\mathrm{e}^{\mathrm{i}\beta Z_2},$$

where

$$\beta = \begin{cases} \sqrt{1 - B^2} \quad (\text{real}), & \text{if } |B| < 1, \\ \mathrm{i}\sqrt{B^2 - 1} \quad (\text{imaginary}), & \text{if } |B| > 1. \end{cases}$$

This choice of β is justified by our problem, so that we deal only with outgoing (rather than incoming) waves in case $|B| < 1$ and our solution is not exponentially increasing in Z_2 in case $|B| > 1$. The initial condition on V_0 yields

$$C(X_1, 0) = A(X_1). \tag{A-7}$$

To exclude secular terms in the solution to $O(\varepsilon)$ system require

$$-2\mathrm{i}B(V_0)_{X_1} - 2(V_0)_{Z_1Z_2} = -2\mathrm{i}(BC_{X_1} + \beta C_{Z_1})\mathrm{e}^{\mathrm{i}\beta Z_2} \equiv 0,$$

or

$$BC_{X_1} + \beta C_{Z_1} \equiv 0. \tag{A-8}$$

Equations (A-7), (A-8) have the solution

$$C(X_1, Z_1) = A(X_1 - \frac{B}{\beta}Z_1).$$

Thus

$$V_0 = A\left(X_1 - \frac{B}{\beta}Z_1\right)\mathrm{e}^{\mathrm{i}\beta Z_2}, \qquad V_1 \equiv 0.$$

Note also that V_2 may have secular terms linear in Z_2. So,

$$U(X_1, X_2, Z_1, Z_2) = A\left(X_1 - \frac{B}{\beta}Z_1\right)\mathrm{e}^{\mathrm{i}BX_2}\mathrm{e}^{\mathrm{i}\beta Z_2} + \varepsilon^2 O(Z_2)$$

or

$$u(x, z) = A\left(x - \frac{B}{\beta}z\right)\mathrm{e}^{\mathrm{i}(Bx+\beta z)/\varepsilon} + \varepsilon O(z). \tag{A-9}$$

Note that, in case $|B| > 1$, $|U| \sim \mathrm{e}^{-|\beta|Z_2}$ is exponentially decreasing in Z_2 and therefore can be neglected. The case $|B| = 1$ requires different multiple scales, but this case, corresponding to $\Lambda/\lambda =$ integer, is quite rare in practice, and we do not consider it in this work.

The result (A-9) is consistent in the leading term with (17) derived by the method of stationary phase.

References

1. R. Kashyap, *Fiber Bragg Gratings*. San Diego: Academic Press (1999) 458 pp.
2. G. R. Fowles, *Introduction to Modern Optics*. New York: Holt, Rinehart and Winston (1968) 304 pp.
3. J. W. Goodman, *Introduction to Fourier Optics.* New York: McGraw-Hill (1968) 287 pp.
4. C. E. Pearson (ed.), *Handbook of Applied Mathematics. Selected Results and Methods*. New York: Van Nostrand Reinhold (1990) 1037 pp.
5. C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers*. New York: McGraw-Hill (1978) 593 pp.
6. A. Papoulis, *Systems and Transforms with Applications to Optics*. New York: McGraw-Hill (1968) 474 pp.
7. M. Abramowitz and I. A. Stegun (eds), *Handbook of Mathematical Functions*. New York: Dover Publications (1965) 1046 pp.
8. M. Born and E. Wolf, *Principles of Optics*. New York: Pergamon Press (1964) 803 pp.
9. F. W. J. Olver, Error bounds for stationary phase approximations. *SIAM J. Math. Anal* 5 (1974), 19–29.
10. J. T. Winthrop and C. R. Worthington, Theory of Fresnel images. I. Plane periodic objects in monochromatic light. *J. Opt. Soc. Am* 55 (1965) 373–381.
11. S. Rudnaya, *Analysis and Optimal Design of Diffrative Optical Elements*. PhD thesis, University of Minnesota, School of Mathematics (1999) 85 pp.
12. M. H. Holmes, *Introduction to Perturbation Methods*. New York: Springer-Verlag (1995) 337 pp.