Analytical evaluation of image formation in optical systems

1974 J. Phys. A: Math. Nucl. Gen. 7 505

(http://iopscience.iop.org/0301-0015/7/4/014)

J. Phys. A: Math., Nucl. Gen., Vol. 7, No. 4, 1974. Printed in Great Britain.

Analytical evaluation of image formation in optical systems

Eero Byckling

Department of Physics, University of Jyväskylä, Jyväskylä, Finland

Received 9 April 1973, in final form 17 September 1973

Abstract. The methods of modern theoretical physics, in particular eigenfunction expansions and complex integration, are used to analyse image formation in optical systems. We solve here the problem of the refraction of a wavefront in a spherical boundary between two media. The fields are expanded in terms of spherical harmonics, and the amplitude distribution in the image is then evaluated by the method of stationary phase. Both aberrations and diffraction are included.

1. Introduction

All practical methods for evaluating the imaging properties of optical systems are based on ray tracing techniques. In spite of their extreme usefulness, they involve certain drawbacks which are especially evident when designing complex systems and when the performance is near the diffraction limit. Firstly, the effect of a change of the system parameters on the image quality is highly nonlinear and lens design and optimization require both extensive experience and large computer programs. Secondly, ray tracing disregards the wave aspects of light and a fair amount of additional calculation is needed to obtain diffraction effects.

The above reasons give justification to trying to find alternative methods and pursuing these even if they appear at first fairly complex. Here we study the application of methods analogous to scattering theory in quantum physics and to those applied in microwave engineering.

A wave field in a homogeneous isotropic medium can be expanded in terms of a complete set of orthogonal functions and this expansion is valid everywhere outside the source of the radiation. If the expansion is in terms of spherical harmonics, the successive terms in the expansion can be regarded as due to multipole radiation of increasing order. Similarly, a converging wavefront can be expanded in multipoles with respect to an arbitrary point. We are then confronted by two problems.

(i) How is a multipole radiation field changed at a boundary of two media of different indices of refraction?

(ii) Is it possible to combine successive refractions so as to determine the wavefield emerging from a complex optical system?

In the following we develop the analytical techniques for solving the first of these problems in the case of a spherical boundary between two media. Problem (ii) will be treated in a subsequent paper.

The analogy with the quantum scattering problem as well as the following calculations show that it is *not* possible to have a complete understanding of a problem of this type in terms of intensity $I = |A|^2$ (ray tracing). It is necessary also to have the phase information, ie to work with the amplitude A.

We wish to stress that the results of this and the two subsequent papers form only an approach to the problem. At this stage it is not clear how this treatment, when developed to a practical computation scheme, would relate to the well known methods of calculating image formation in optical systems.

2. Single spherical boundary and point object on axis

We start with the simplest possible situation depicted in figure 1. The spherical surface with centre at O and radius r separates two media with refractive indices n and n'. Assume that a monochromatic point source is located at P. Let the magnitude of the wavevector in vacuum be $k = \omega/c$ so that in the two media it is nk and $n'k$.

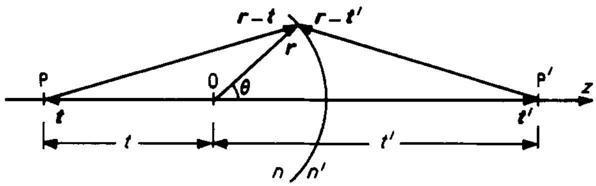

Figure 1.

Inside a region which is free of currents and charges, the amplitude at $\boldsymbol{r}$ is

$$\hat{A}(\boldsymbol{r}, t) = A(\boldsymbol{r})\,\mathrm{e}^{\mathrm{i}\omega t}. \tag{2.1}$$

If we disregard polarization, $A(\boldsymbol{r})$ is a scalar function (eg the value of the electric vector) which satisfies the wave equation

$$\nabla^2 A + n^2 k^2 A = 0. \tag{2.2}$$

If a complete orthogonal set of solutions of (2.2) is found, then any $A(\boldsymbol{r})$ can be expressed in terms of the set.

In order to satisfy the boundary conditions later on, we now expand all the wavefields of the problem in terms of spherical harmonics centred at O. Then the solutions of (2.2) are in spherical coordinates (r, θ, ϕ) (van Bladel 1964)

$$A(\boldsymbol{r}) = \sum_{l,m} B_{lm} h_l^{(2)}(nkr) Y_l^m(\theta, \phi) + \sum_{l,m} C_{lm} h_l^{(1)}(nkr) Y_l^m(\theta, \phi). \tag{2.3}$$

Here (see Arfken 1970) $h_l^{(1)}(x)$ and $h_l^{(2)}(x)$ are the spherical Hankel functions of order l, $Y_l^m(\theta, \phi)$ are spherical harmonics, and B_{lm}, C_{lm} are complex coefficients. In terms of Legendre polynomials $P_l^m(\cos\theta)$ we have

$$Y_l^m(\theta, \phi) = \left(\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}\right)^{1/2} P_l^m(\cos\theta)\,\mathrm{e}^{\mathrm{i}m\phi}. \tag{2.4}$$

The asymptotic expressions of $h_l^{(1)}$ and $h_l^{(2)}$ are

$$\begin{aligned} h_l^{(1)}(x) &\to (-\mathrm{i})^{l+1}\,\mathrm{e}^{\mathrm{i}kx}/kx \qquad x \to \infty \\ h_l^{(2)}(x) &\to \mathrm{i}^{l+1}\,\mathrm{e}^{-\mathrm{i}kx}/kx \qquad x \to \infty. \end{aligned} \tag{2.5}$$

If (2.3) is an outgoing wave, then only the second type of asymptotic behaviour is allowed, and all C_{lm} must be zero (van Bladel 1964). Further, the situation in figure 1 is symmetric around the z axis, and therefore all the coefficients B_{lm} with $m \neq 0$ are zero. We find

$$A(\boldsymbol{r}) = \sum_{l=0}^{\infty} \left(\frac{2l+1}{4\pi}\right)^{1/2} B_l h_l^{(2)}(nkr) P_l(\cos\theta). \tag{2.6}$$

In particular, the simple spherical wave

$$A(\boldsymbol{r}) = \frac{\exp(-ink|\boldsymbol{r}-\boldsymbol{t}|)}{|\boldsymbol{r}-\boldsymbol{t}|} \tag{2.7}$$

originating at a point source at $\boldsymbol{t}$ can be written in the form (2.6). The coefficients B_l are most simply found by recalling the well known formula (van Bladel 1964)

$$\frac{\exp(-ink|\boldsymbol{r}-\boldsymbol{t}|)}{|\boldsymbol{r}-\boldsymbol{t}|} = -ink \sum_{l=0}^{\infty} (2l+1) P_l(-\cos\theta) j_l(nkr) h_l^{(2)}(nkt) \tag{2.8}$$

which is valid for $r < t$. The minus sign in $\cos\theta$ is because the angle between $\boldsymbol{t}$ and $\boldsymbol{r}$ in figure 1 is $\pi - \theta$.

We now find the amplitude at point P′,$\boldsymbol{t}'$, in the second medium. If $\boldsymbol{r}$ is a point at the boundary surface and $A(\boldsymbol{r})$ the amplitude at that point, the effect caused by this element on the amplitude at point $\boldsymbol{t}'$ is given by the Green function (propagator) (van Bladel 1964)

$$G(\boldsymbol{r},\boldsymbol{t}') = \frac{\exp(-in'k|\boldsymbol{r}-\boldsymbol{t}'|)}{|\boldsymbol{r}-\boldsymbol{t}'|}. \tag{2.9}$$

The total amplitude at point P′ is then the integral

$$A(\boldsymbol{t}') = \int_{\text{surface}} \mathrm{d}^2 r\, G(\boldsymbol{r},\boldsymbol{t}') A(\boldsymbol{r}). \tag{2.10}$$

The surface is assumed to have a finite aperture so that only rays for $\theta \leqslant \hat{\theta}$ can pass through. We then have

$$A(\boldsymbol{t}') = \int_0^{2\pi} \mathrm{d}\phi \int_0^{\hat{\theta}} \mathrm{d}\theta \frac{\exp(-in'k|\boldsymbol{r}-\boldsymbol{t}'|)}{|\boldsymbol{r}-\boldsymbol{t}'|} \frac{\exp(-ink|\boldsymbol{r}-\boldsymbol{t}|)}{|\boldsymbol{r}-\boldsymbol{t}|}. \tag{2.11}$$

Using (2.8), integrating over ϕ and writing $z = \cos\theta$, $\hat{z} = \cos\hat{\theta}$, we get

$$A(t') = -nn'k^2 2\pi \sum_{l,l'} \int_{\hat{z}}^{1} \mathrm{d}z (2l+1)(2l'+1) P_l(-z) P_{l'}(z) j_l(nkr) j_{l'}(n'kr) h_l^{(2)}(nkt) \times h_{l'}^{(2)}(n'kt'). \tag{2.12}$$

The integral over z is evaluated in appendix 1 (see (A.7)). Further, we use

$$j_l(x) = \left(\frac{\pi}{2x}\right)^{1/2} J_{l+\frac{1}{2}}(x) \tag{2.13}$$

$$h_l^{(2)}(x) = \left(\frac{\pi}{2x}\right)^{1/2} H_{l+\frac{1}{2}}^{(2)}(x) \tag{2.14}$$

$$P_l(-z) = (-1)^l P_l(z) = \mathrm{e}^{-i\pi l} P_l(z) \tag{2.15}$$

and the asymptotic forms of $J_{l+\frac{1}{2}}(x)$ and $H^{(2)}_{l+\frac{1}{2}}(x)$ which are derived in appendix 2 (see (A.13), (A.14)). Then we find

$$A(t') = 8\frac{1}{(r^2tt')^{1/2}}\sum_{l,l'}(ll')^{1/2}\,\mathrm{e}^{-\mathrm{i}\pi l}\frac{\sin(l-l')\hat{\theta}}{l-l'}[(n^2k^2r^2-l^2)(n'^2k^2r^2-l'^2)(n^2k^2t^2-l^2)$$
$$\times(n'^2k^2t'^2-l'^2)]^{-1/4}\cos\Phi_l\,\cos\Phi'_{l'}\exp(-\mathrm{i}\Psi_l-\mathrm{i}\Psi'_{l'}). \tag{2.16}$$

The phase factors are

$$\Phi_l = (n^2k^2r^2-l^2)^{1/2}-l\cos^{-1}\left(\frac{l}{nkr}-\frac{\pi}{4}\right) \tag{2.17}$$

$$\Psi_l = (n^2k^2t^2-l^2)^{1/2}-l\cos^{-1}\left(\frac{l}{nkt}-\frac{\pi}{4}\right) \tag{2.18}$$

and $\Phi'_{l'}$ and $\Psi'_{l'}$ are obtained by substituting

$$n\to n',\qquad t\to t',\qquad l\to l'. \tag{2.19}$$

The asymptotic formulae for P_l, J_l, H_l are correct to order l^{-1}. In (2.16) we have also replaced $l+\frac{1}{2}$ by l, $l'+\frac{1}{2}$ by l', and thus neglected terms of order l^{-1}. Now the main contribution to (2.16) comes when $l, l' \simeq kd$ where d is of the magnitude of the macroscopic dimensions r, t, t' of the system. In practice $kd \simeq 2\pi d/\lambda$ is of the order 10^5 and thus the results are accurate to some 10^{-5}. Also because l is large and all functions vary slowly in l, all sums over l can be replaced with integrals over l. The resulting errors are of the order l^{-1}. For the validity of this method when l is regarded as a continuous variable, it is essential to interpret $(-1)^l$ as $\mathrm{e}^{-\mathrm{i}\pi l}$ in (2.15); this procedure may be justified but will not be discussed here.

In (2.16) the function $\sin(l-l')\hat{\theta}/(l-l')$ vanishes rapidly for $|l-l'| > \pi/4\hat{\theta}$. In practice $\hat{\theta}$ is between 10^{-2} and $\pi/2$. Then only a strip of width less than 100 in the l,l' plane contributes to (2.16) (see figure 2). Using the formula (Gradshteyn and Ryzhik 1965)

$$\int_{-\infty}^{\infty}\mathrm{d}v\frac{\sin\hat{\theta}v}{v}\mathrm{e}^{\mathrm{i}\chi v} = \begin{cases}\pi & |\chi| < \hat{\theta}\\ 0 & |\chi| > \hat{\theta}\end{cases} \tag{2.20}$$

we see that there are possibly upper and lower limits for l set by $\hat{\theta}$, and we get

$$A(t') = 8\pi\frac{1}{(r^2tt')^{1/2}}\sum_{l_1<l<l_2} l[(n^2k^2r^2-l^2)(n'^2k^2r^2-l^2)(n^2k^2t^2-l^2)(n'^2k^2t'^2-l^2)]^{-1/4}$$
$$\times\mathrm{e}^{-\mathrm{i}\pi l}\cos\Phi_l\cos\Phi'_{l'}\exp(-\mathrm{i}\Psi_l-\mathrm{i}\Psi'_{l'}). \tag{2.21}$$

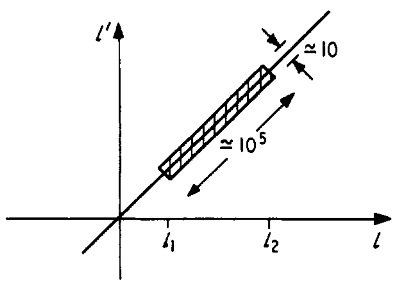

Figure 2. The region of integration in the l, l' plane.

Assuming for a moment that l_1 and l_2 are not important, it is seen that in summing over l in (2.21), a sizable contribution results only from an interval where the phase

$$\phi_l = \pm\Phi_l \pm \Phi'_l - \Psi_l - \Psi'_l - \pi l \tag{2.22}$$

varies slowly. This can be interpreted as *phase coherence* within an interval in l. One has

$$\frac{\partial \Phi_l}{\partial l} = -\cos^{-1}\left(\frac{l}{nkr}\right) \tag{2.23}$$

and thus

$$\frac{\partial \phi_l}{\partial l} = \mp\cos^{-1}\left(\frac{l}{nkr}\right) \mp\cos^{-1}\left(\frac{l}{n'kr}\right) + \cos^{-1}\left(\frac{l}{nkt}\right) + \cos^{-1}\left(\frac{l}{n'kt'}\right) - \pi. \tag{2.24}$$

In a situation of the type depicted in figure 1 one has $n > n', t, t' > r$. Then $\partial\phi_l/\partial l = 0$ only when the first and second signs in (2.24) are $+$ and $-$, respectively. If $l = l_0$ is the solution of $\partial\phi_l/\partial l = 0$, then calculation of $\partial^2\phi_l/\partial l^2$ and use of equation (A.9) gives

$$A(t') = |A(t')| \exp[\mathrm{i}(\phi_{l_0} - \tfrac{1}{4}\pi)] \tag{2.25}$$

$$|A(t')| = 2\pi\left(\frac{2\pi}{-Br^2tt'}\right)^{1/2} l_0[(n^2k^2r^2 - l_0^2)(n'^2k^2r^2 - l_0^2)(n^2k^2t^2 - l_0^2)$$
$$\times (n'^2k^2t'^2 - l_0^2)]^{-1/4} \tag{2.26}$$

$$B = \frac{\partial^2\phi}{\partial l_0^2} = (n^2k^2r^2 - l_0^2)^{-1/2} - (n'^2k^2r^2 - l_0^2)^{-1/2} + (n^2k^2t^2 - l_0^2)^{-1/2}$$
$$+ (n'^2k^2t'^2 - l_0^2)^{-1/2}. \tag{2.27}$$

The phase ϕ_{l_0} is given by (2.22) for $l = l_0$.

Let us now consider the situation when l is much smaller than its maximum value ($l \ll 10^6$) but large enough that the asymptotic formulae are valid ($l \gg 10$). The limit

$$\cos^{-1} x \underset{x\to 0}{\to} \tfrac{1}{2}\pi - x - \tfrac{1}{6}x^3 \tag{2.28}$$

means for $\partial\phi_1/\partial l = 0$ that

$$\frac{1}{nkr} - \frac{1}{n'kr} + \frac{1}{nkt} + \frac{1}{n'kt'} = \frac{1}{6}l_0^2\left(\frac{1}{(nkr)^3} - \frac{1}{(n'kr)^3} + \frac{1}{(nkt)^3} + \frac{1}{(n'kt')^3}\right). \tag{2.29}$$

This implies that l_0 is zero if

$$\frac{1}{nr} - \frac{1}{n'r} + \frac{1}{nt} + \frac{1}{n't'} = 0. \tag{2.30}$$

This equation is the familiar formula $(n - n')/r = n/s + n'/s'$ written for $t = s - r$ and $t' = s' + r$. Now the small values of l get their amplitude mainly from small values of θ, so that it is easy to understand that the intensity at the gaussian image point (2.30) is due to phase coherence at small l.

When t' increases from the gaussian value, the left-hand side of (2.29) becomes negative, l_0 does not exist and the intensity vanishes. When t' decreases, (2.29) gives very roughly $l_0 \sim [(t'_0 - t')/t'_0]^{1/2}$. Because B goes as l_0^2, $A(t')$ starts from a constant. With further change in t', the limit l_2 in (2.21) is met and $A(t')$ goes to zero. An accurate description of both the amplitude and phase of $A(t')$ can be obtained by solving l_0 numerically from $\partial\phi_l/\partial l = 0$ and substituting into (2.25)–(2.27).

The preceding considerations apply for an aberration limited image. To include diffraction, we must consider the double summation in the l, l' plane accurately also in the $l-l'$ direction. The phase in the l, l' plane is

$$\phi(l, l') = \Phi_l - \Phi'_{l'} - \Psi_l - \Psi'_{l'} - \pi l. \tag{2.31}$$

The derivative with respect to $v = l-l'$ at $l = l'$ is

$$\chi = \frac{\partial\phi}{\partial v} = \frac{1}{2}\left[\cos^{-1}\left(\frac{l}{nkr}\right) + \cos^{-1}\left(\frac{l}{n'kr}\right) + \cos^{-1}\left(\frac{l}{nkt}\right) - \cos^{-1}\left(\frac{l}{n'kt'}\right) - \pi\right]. \tag{2.32}$$

When $l \to 0$, also $\chi \to 0$, and thus in (2.21) $l_1 = 0$. The upper limit l_2 results from

$$\chi = \frac{\partial\phi}{\partial v} = \hat{\theta} \qquad \text{at } l = l_2. \tag{2.33}$$

The exact (to order $1/nkr$) value of $A(t')$ can be obtained by summing (integrating) (2.21) numerically from 0 to l_2.

The width of the region where ϕ_l is stationary is seen from $\phi_l \simeq \phi_{l_0} + \frac{1}{2}B(l-l_0)^2$ to be $2(2/B)^{1/2}$. Thus the image is diffraction-limited if from (2.33) $l_2 \ll 2(2/B)^{1/2}$ and aberration-limited if $l_2 \gg 2(2/B)^{1/2}$. These, of course, lead to the well known criteria.

The intensity in an off-axis point $\boldsymbol{t}'$ (caused by an on-axis object point) is treated almost similarly. The argument z in $P_{l'}(z)$ in (2.12) is the cosine of the angle between $\boldsymbol{r}$ and $\boldsymbol{t}'$. If α', β' are the azimuthal and polar angles of $\boldsymbol{t}'$ (see figure 3), then one can use the formula

$$P_{l'}(\cos\gamma') = \frac{4\pi}{2l'+1}\sum_{m'} Y_{l'}^{m'}(\alpha', \beta') Y_{l'}^{m'}(\theta, \phi). \tag{2.34}$$

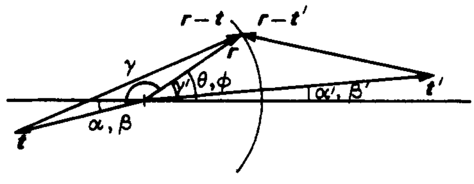

Figure 3.

Integration over ϕ now gives $2\pi\delta_{m',0}$ and we find instead of (2.12),

$$A(t') = -nn'k^2 2\pi \sum_{l,l'} \int_{z}^{1} dz(2l+1)(2l'+1)P_l(-z)P_{l'}(z)P_{l'}(\cos\alpha')j_l(nkr)$$
$$\times j_{l'}(n'kr)h_l^{(2)}(nkt)h_{l'}^{(2)}(n'kt'). \tag{2.35}$$

In the phase ϕ_l, (2.22) there will then be an additional term $\pm\alpha' l$. There are now two terms in $A(\boldsymbol{t}')$, which have saddle-points for $\partial\phi_l/\partial l = \pm\alpha$, (see (2.24)). Also the limits l_1 and l_2 are changed due to the additional term $\pm\alpha'$ in (2.32). Otherwise the computation of $A(\boldsymbol{t}')$ is as earlier.

3. Off-axis object point

If $\boldsymbol{t}$ and $\boldsymbol{t}'$ are not on the z axis, the amplitude at $\boldsymbol{t}'$ is still given by (2.10) and (2.7). The relevant expansion is then

$$\frac{\exp(-\mathrm{i}nk|\boldsymbol{r}-\boldsymbol{t}|)}{|\boldsymbol{r}-\boldsymbol{t}|} = -\mathrm{i}nk\sum_{l=0}^{\infty}(2l+1)P_l(\cos\gamma)j_l(nkr)h_l^{(2)}(nkt) \tag{3.1}$$

and the similar primed expression for the second medium. The notation is that shown in figure 3. One can then write $P_l(\cos\gamma)$ in terms of the polar and azimuthal angles α, β for $\boldsymbol{t}$ and θ, ϕ for $\boldsymbol{r}$ using the formula

$$P_l(\cos\gamma) = \frac{4\pi}{2l+1}\sum_{m=-l}^{l} Y_l^m(\alpha,\beta)Y_l^m(\theta,\phi) \tag{3.2}$$

and similarly for $P_l(\cos\gamma')$. Then $A(\boldsymbol{t}')$ becomes

$$\begin{aligned} A(\boldsymbol{t}') = {} & -nn'k^2 2\pi\sum_{l,l'}\sum_{m}(2l+1)(2l'+1)\frac{(l-m)!(l'+m)!}{(l+m)!(l'-m)!} \\ & \times\int_{\hat{z}}^{1}\mathrm{d}z P_l^m(z)P_{l'}^{-m}(z)P_l^m(\cos\alpha)P_{l'}^{-m}(\cos\alpha')\,\mathrm{e}^{\mathrm{i}m(\beta-\beta')}j_l(nkr)j_l(n'kr) \\ & \times h_l^{(2)}(nkt)h_{l'}^{(2)}(n'kt'). \end{aligned} \tag{3.3}$$

In (3.3) we have made use of (2.4) and integrated over ϕ which gives $2\pi\delta_{m,m'}$. After this the sum over m' has been taken.

The z integral is evaluated in (A.6). The use of (2.13), (2.14), (A.13), (A.14) then gives

$$\begin{aligned} A(\boldsymbol{t}') = {} & \frac{8}{(r^2tt')^{1/2}}\sum_{l,l',m}(ll')^{1/2}\left(\frac{(l-m)!(l'-m)!}{(l+m)!(l'+m)!}\right)^{1/2}\left(\frac{l^2-m^2}{l^2\sin^2\hat{\theta}-m^2}\,\frac{l'^2-m^2}{l'^2\sin^2\hat{\theta}-m^2}\right)^{1/4} \\ & \times\frac{(l^2\sin^2\hat{\theta}-m^2)^{1/2}+(l'^2\sin^2\hat{\theta}-m^2)^{1/2}}{l+l'}\,\frac{\sin(\omega_l-\omega_{l'})}{l-l'}P_l^m(\cos\alpha)P_{l'}^m(\cos\alpha') \\ & \times\mathrm{e}^{\mathrm{i}m(\beta-\beta')}[(n^2k^2r^2-l^2)(n'^2k^2r^2-l^2)(n^2k^2t^2-l^2) \\ & \times(n'^2k^2t'^2-l^2)]^{-1/4}\cos(\Phi_l+\Phi'_{l'})\exp(-\mathrm{i}\Psi_l-\Psi'_{l'}). \end{aligned} \tag{3.4}$$

The phases ω_l and $\omega_{l'}$, due to the θ integral, are given by

$$\omega_l = \omega_l(\hat{\theta}) = l\cos^{-1}\left(\frac{l}{(l^2-m^2)^{1/2}}\cos\hat{\theta}\right) - m\cos^{-1}\left(\frac{m}{(l^2-m^2)^{1/2}}\cot\hat{\theta}\right) - \frac{\pi}{4}. \tag{3.5}$$

The derivative is simply

$$\frac{\partial\omega_l(\hat{\theta})}{\partial l} = \cos^{-1}\left(\frac{l}{(l^2-m^2)^{1/2}}\cos\hat{\theta}\right) \tag{3.6}$$

and thus for $v = l-l'$ we get

$$\frac{\partial(\omega_l-\omega_{l'})}{\partial v} = \frac{1}{2}\cos^{-1}\left(\frac{l}{(l^2-m^2)^{1/2}}\cos\hat{\theta}\right) + \frac{1}{2}\cos^{-1}\left(\frac{l'}{(l'^2-m^2)^{1/2}}\cos\hat{\theta}\right). \tag{3.7}$$

We see again that only small values of $|l-l'|$ contribute. Denoting by χ the derivative of the phase (see later equations (3.22), (3.23)) we see that, due to (2.20) we get the constraint

$$\chi < \cos^{-1}\left(\frac{l\cos\hat{\theta}}{(l^2-m^2)^{1/2}}\right). \tag{3.8}$$

Then we can make the following replacement in (3.4):

$$\sum_{l,l',m} \frac{\sin(\omega_l-\omega_{l'})}{l-l'}\chi \to \pi\sum_{l,m}\chi, \qquad \chi < \cos^{-1}\left(\frac{l\cos\hat{\theta}}{(l^2-m^2)^{1/2}}\right). \tag{3.9}$$

The phase of the quantity to be summed over l and m in (3.4) is, on the strip $l = l'$, equal to

$$\phi(l,m) = -\Phi_l+\Phi_l'-\Psi_l-\Psi_l'\pm\omega_l(\alpha)\pm\omega_l(\alpha')+m\beta-m\beta' \tag{3.10}$$

with

$$\omega_l(\alpha) = l\cos^{-1}\left(\frac{l}{(l^2-m^2)^{1/2}}\cos\alpha\right)-m\cos^{-1}\left(\frac{m}{(l^2-m^2)^{1/2}}\cot\alpha\right)-\frac{\pi}{4}. \tag{3.11}$$

If we write

$$A(t') = \sum_{l,m}|A_{l,m}(t')|\,e^{i\phi(l,m)} \tag{3.12}$$

the coefficient is

$$|A_{l,m}(t')| = \frac{2^4}{(r^2tt')^{1/2}l}\frac{l^2-m^2}{(l^2\sin^2\alpha-m^2)^{1/4}(l^2\sin^2\alpha'-m^2)^{1/4}}[(n^2k^2r^2-l^2)$$
$$\times(n'^2k^2r^2-l^2)(n^2k^2r^2-l^2)(n'^2k^2t'^2-l^2)]^{-1/4}. \tag{3.13}$$

The derivatives of $\phi(l,m)$ with respect to l and m are

$$\frac{\partial\phi(l,m)}{\partial l} = +\cos^{-1}\left(\frac{l}{nkr}\right)-\cos^{-1}\left(\frac{l}{n'kr}\right)+\cos^{-1}\left(\frac{l}{nkt}\right)+\cos^{-1}\left(\frac{l}{n'kt'}\right)$$
$$\pm\cos^{-1}\left(\frac{l}{(l^2-m^2)^{1/2}}\cos\alpha\right)\pm\cos^{-1}\left(\frac{l}{(l^2-m^2)^{1/2}}\cos\alpha'\right) \tag{3.14}$$

$$\frac{\partial\phi(l,m)}{\partial m} = \mp\cos^{-1}\left(\frac{m}{(l^2-m^2)^{1/2}}\cot\alpha\right)\mp\cos^{-1}\left(\frac{m}{(l^2-m^2)^{1/2}}\cot\alpha'\right)+\beta-\beta'. \tag{3.15}$$

The signs in (3.14) result from the reasoning given after equation (2.24).

The equations

$$\frac{\partial\phi(l,m)}{\partial l} = 0, \qquad \frac{\partial\phi(l,m)}{\partial m} = 0 \tag{3.16}$$

give the point l_0, m_0 where the phase is stationary.

The second derivatives are

$$B = \frac{\partial^2\phi(l,m)}{\partial l^2} = (n^2k^2r^2-l^2)^{-1/2}-(n'^2k^2r^2-l^2)^{-1/2}+(n^2k^2t^2-l^2)^{-1/2}$$
$$+(n'^2k^2t'^2-l^2)^{-1/2}\pm\frac{m^2\cos\alpha}{(l^2-m^2)^{3/2}\{1-[l^2\cos^2\alpha/(l^2-m^2)]\}^{1/2}}$$
$$\pm\frac{m^2\cos\alpha'}{(l^2-m^2)^{3/2}\{1-[l^2\cos^2\alpha'/(l^2-m^2)]\}^{1/2}} \tag{3.17}$$

$$C = \frac{\partial^2\phi(l,m)}{\partial l\,\partial m} = \mp\frac{lm\cos\alpha}{(l^2-m^2)^{3/2}\{1-[l^2\cos^2\alpha/(l^2-m^2)]\}^{1/2}}$$
$$\mp\frac{lm\cos\alpha'}{(l^2-m^2)^{3/2}\{1-[l^2\cos^2\alpha'/(l^2-m^2)]\}^{1/2}} \tag{3.18}$$
$$D = \frac{\partial^2\phi(l,m)}{\partial m^2} = \pm\frac{l^2\cos\alpha}{(l^2-m^2)^{3/2}\{1-[l^2\cos^2\alpha/(l^2-m^2)]\}^{1/2}}$$
$$\pm\frac{l^2\cos\alpha'}{(l^2-m^2)^{3/2}\{1-[l^2\cos^2\alpha'/(l^2-m^2)]\}^{1/2}}. \tag{3.19}$$

The integrals over l and m are then evaluated near the stationary point by the well known formula

$$\int_{-\infty}^{\infty} \mathrm{d}x\,\mathrm{d}y \exp[-\tfrac{1}{2}(Bx^2+2Cxy+Dy^2)] = \frac{2\pi}{(BD-C^2)^{1/2}}. \tag{3.20}$$

Collecting our results we have

$$A(t') = |A(t')|\exp(\mathrm{i}\phi(l_0,m_0))$$
$$|A(t')| = 2^5\pi(r^2tt')^{-1/2}(BD-C^2)^{-1/2}\frac{l_0^2-m_0^2}{l_0}(l^2\sin^2\alpha-m^2)^{-1/4}(l^2\sin^2\alpha'-m^2)^{-1/4}$$
$$\times[(n^2k^2r^2-l_0^2)(n'^2k^2r^2-l_0^2)(n^2k^2t^2-l_0^2)(n'^2k^2t'^2-l_0^2)]^{-1/4} \tag{3.21}$$
$$\phi(l_0,m_0) = -\Phi_{l_0}+\Phi'_{l_0}-\Psi_{l_0}-\Psi'_{l_0}\pm\omega_{l_0}(\alpha)\pm\omega_{l_0}(\alpha')+m_0\beta-m_0\beta'.$$

The quantities Φ, Ψ and ω are defined in (2.17), (2.18) and (3.11), and B, C and D are defined in (3.17)–(3.19). The pair l_0, m_0 is a solution of (3.16) (see also (3.14)–(3.15)). In equations (3.14)–(3.19), (3.21) there are four sign combinations. Each of these may give a solution l_0, m_0, and if there are several solutions their contributions must be added together.

Finally, we must consider the constraint (3.8). The derivative of the complete phase in (3.4) is

$$\frac{\partial\phi(l,l',m)}{\partial(l-l')} = \frac{1}{2}\left[\cos^{-1}\left(\frac{l}{nkr}\right)+\cos^{-1}\left(\frac{l'}{n'kr}\right)+\cos^{-1}\left(\frac{l}{nkt}\right)-\cos^{-1}\left(\frac{l'}{n'kt'}\right)\right.$$
$$\left.\pm\cos^{-1}\left(\frac{l}{(l^2-m^2)^{1/2}}\cos\alpha\right)\mp\cos^{-1}\left(\frac{l'}{(l'^2-m^2)^{1/2}}\cos\alpha'\right)\right]. \tag{3.22}$$

This is evaluated at $l = l'$,

$$\chi = \frac{\partial\phi(l,l',m)}{\partial(l-l')} \qquad \text{at } l = l'. \tag{3.23}$$

In the sum over l, m, only those pairs l, m are included which satisfy (3.8). When the method of stationary phase is used to evaluate this sum, the pair l_0, m_0 must satisfy (3.8). When the stationary point l_0, m_0 moves outside the region (3.8), the intensity decreases exponentially.

4. Discussion

We have evaluated analytically the complex amplitude $A(\mathbf{r}')$ in the image space due to a point source at $\mathbf{r}$ in object space. To be able to discuss the successive refractions in a real system, we must be able to image also a source which has a finite extension. This can be done using the present results by simply integrating over the intensity distribution in the object. This, however, requires so much computing that no practical applications would be expected. A completely different situation arises, when one realizes that the amplitude distribution in the intermediate image space is not needed, because the expansion in spherical harmonics is equivalent to it and can be used directly to describe the wavefront incident on the next surface. The case of successive refractions will be treated, using the present results, in a separate paper.

Acknowledgments

The author is greatly indebted to Dr Pekka Rautala for bringing these problems of optics to his attention and for many inspiring and encouraging discussions. Acknowledgment is also due to Dr C Cronström for suggesting the use of the wronskian in evaluating the integral (A.1).

Appendix 1. The integral $\int \mathrm{d}z P_l^m(z) P_{l'}^m(z)$

Because $P_l^m, P_{l'}^m$ are solutions of a self-adjoint second-order differential equation, the integral

$$\int_{\hat{z}}^{1} \mathrm{d}z\, P_l^m(z) P_{l'}^m(z) \tag{A.1}$$

can be evaluated in terms of the wronskian. The Legendre differential equation gives the equality

$$I = \int_{\hat{z}}^{1} \mathrm{d}z P_l^m(z)\left[\frac{\mathrm{d}}{\mathrm{d}z}\left((1-z^2)\frac{\mathrm{d}P_{l'}^m(z)}{\mathrm{d}z}\right) - \frac{m^2}{1-z^2}P_{l'}^m(z)\right] = -l'(l'+1)\int_{\hat{z}}^{1} \mathrm{d}z P_l^m(z) P_{l'}^m(z). \tag{A.2}$$

Partial integration of the left-hand side of (A.2) results in

$$I = \int_{\hat{z}}^{1} P_l^m(z)\left((1-z^2)\frac{\mathrm{d}P_{l'}^m(z)}{\mathrm{d}z}\right) - \int_{\hat{z}}^{1} \mathrm{d}z\left(\frac{\mathrm{d}}{\mathrm{d}z}P_l^m(z)\right)(1-z^2)\left(\frac{\mathrm{d}}{\mathrm{d}z}P_{l'}^m(z)\right)$$
$$-\int_{\hat{z}}^{1} \mathrm{d}z P_l^m(z)\frac{m^2}{1-z^2}P_{l'}^m(z). \tag{A.3}$$

Interchanging l and l' in (A.2) and (A.3) will give a similar expression for

$$-l(l+1)\int \mathrm{d}z\, P_l^m P_{l'}^m.$$

Taking the difference then results in

$$\int_{\hat{z}}^{1} \mathrm{d}z\, P_l^m(z) P_{l'}^m(z) = (1-\hat{z}^2)\frac{P_l^{m\prime}(\hat{z})P_{l'}^m(\hat{z}) - P_l^m(\hat{z})P_{l'}^{m\prime}(\hat{z})}{l(l+1)-l'(l'+1)} \tag{A.4}$$

In the limit of large l, m the derivative of $P_l^m(z)$ is, according to (A.28), equal to

$$\frac{\mathrm{d}}{\mathrm{d}z}P_l^m(z) = \left(\frac{2}{\pi l}\frac{(l+m)!}{(l-m)!}\right)^{1/2}\left(\frac{l^2-m^2}{l^2\sin^2\theta-m^2}\right)^{1/4}$$
$$\times\left(\frac{\frac{1}{2}\cos\theta}{\sin^2\theta-(m^2/l^2)}\cos(\quad)+\frac{\frac{1}{2}+(l/\sin^2\theta)[\sin^2\theta-(m^2/l^2)]}{[\sin^2\theta-(m^2/l^2)]^{1/2}}\sin(\quad)\right). \quad (A.5)$$

Substituting into (A.4) and approximating $l+\frac{1}{2} \simeq l$, we get

$$(l-l')(l+l')\int_{\hat{z}}^{1}\mathrm{d}z\, P_l^m(z)P_{l'}^m(z)$$
$$= \frac{1}{\pi(ll')^{1/2}}\left(\frac{(l+m)!(l'+m)!}{(l-m)!(l'-m)!}\right)^{1/2}\left(\frac{l^2-m^2}{l^2\sin^2\theta-m^2}\,\frac{l'^2-m^2}{l'^2\sin^2\theta-m^2}\right)^{1/4}$$
$$\times[(l^2\sin^2\theta-m^2)^{1/2}+(l'^2\sin^2\theta-m^2)^{1/2}]$$
$$\times\sin\left[l\cos^{-1}\left(\frac{l}{(l^2-m^2)^{1/2}}\cos\theta\right)-l'\cos^{-1}\left(\frac{l'}{(l'^2-m^2)^{1/2}}\cos\theta\right)\right.$$
$$\left.-m\cos^{-1}\left(\frac{m}{(l^2-m^2)^{1/2}}\cot\theta\right)+m\cos^{-1}\left(\frac{m}{(l'^2-m^2)^{1/2}}\cot\theta\right)\right]. \quad (A.6)$$

In the case $m = 0$ we get

$$\int_{\hat{z}}^{1}\mathrm{d}z\, P_l(z)P_{l'}(z) \simeq \frac{1}{\pi(ll')^{1/2}}\frac{\sin(l-l')\hat{\theta}}{l-l'}. \quad (A.7)$$

Appendix 2. Asymptotic forms of J_l, H_l, P_l^m

The principle of stationary phase is here used to evaluate $H_l^{(2)}(z)$, $J_l(z)$ and $P_l^m(\cos\theta)$ for large order. Consider the integral

$$\int_{-\Delta}^{\Delta}\mathrm{d}z\, g(z)\,\mathrm{e}^{\mathrm{i}sf(z)} \quad (A.8)$$

in the limit $s \to \infty$. If $f'(z)$ has in the interval $-\Delta, \Delta$ a zero at $z = z_0$, then at large s only the region near z_0 contributes to (A.8). Taking the Taylor expansion of $f(z)$ near $z = z_0$ gives (Born and Wolf 1964)

$$\int\mathrm{d}z\, g(z)\,\mathrm{e}^{\mathrm{i}sf(z)} \underset{s\to\infty}{\to} \left(\frac{-2\pi}{sf''(z_0)}\right)^{1/2} g(z_0)\,\mathrm{e}^{-\mathrm{i}\pi/4}\,\mathrm{e}^{\mathrm{i}sf(z_0)}. \quad (A.9)$$

For the real part we have

$$\int\mathrm{d}z\, g(z)\cos(sf(z)) \underset{s\to\infty}{\to} \left(\frac{-2\pi}{sf''(z_0)}\right)^{1/2} g(z_0)\cos(sf(z_0)-\tfrac{1}{4}\pi). \quad (A.10)$$

The Bessel function $J_l(z)$ and $H_l(z)$ have the integral representations (Abramowitz 1965)

$$J_l(z) = \frac{1}{\pi}\int_0^{\pi}\cos(z\sin\phi-l\phi)\,\mathrm{d}\phi \quad (A.11)$$

$$H_l^{(2)}(z) = \frac{1}{\pi \mathrm{i}} \int_{L_2} \exp[\tfrac{1}{2}z(\mathrm{e}^t - \mathrm{e}^{-t}) - lt] \, \mathrm{d}t. \tag{A.12}$$

The path of integration is shown in figure 4. Using (A.10) and (A.9) one easily obtains

$$J_{z \cos \beta}(z) \underset{z \to \infty}{\to} \left(\frac{2}{\pi z \sin \beta}\right)^{1/2} \cos(z \sin \beta - z\beta \cos \beta - \tfrac{1}{4}\pi) \tag{A.13}$$

$$H^{(2)}_{z \cos \beta}(z) \underset{z \to \infty}{\to} \left(\frac{2}{\pi z \sin \beta}\right)^{1/2} \exp[-\mathrm{i}(z \sin \beta - z\beta \cos \beta - \tfrac{1}{4}\pi)]. \tag{A.14}$$

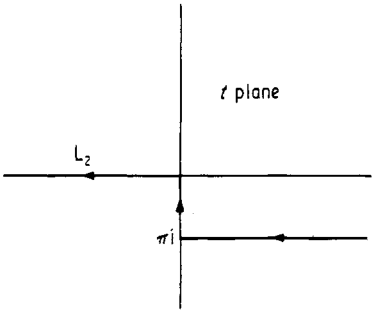

Figure 4. Contour of integration for $H_l^{(2)}(z)$.

The asymptotic expressions of $J_l(z)$ and $H_l^{(2)}(z)$ are also given in the literature (Watson 1948). There it is further shown that when for fixed l the argument z grows, then for $z > l$ both $J_l(z)$ and $H_l^{(2)}(z)$ decrease exponentially.

We now consider $P_l^m(\cos \theta)$ in the limit $l, m \to \infty$ with m/l fixed. There seems to be no previous treatment of this case in the literature. If m is an integer, $P_l^m(\cos \theta)$ has the integral representation (Abramowitz 1965)

$$P_l^m(\cos \theta) = \frac{(l+m)!}{l!} \frac{(-\mathrm{i})^m}{2\pi} \int_{-\pi}^{\pi} (\cos \theta + \mathrm{i} \sin \theta \cos t)^l \, \mathrm{e}^{\mathrm{i}mt} \, \mathrm{d}t. \tag{A.15}$$

We use the Condon–Shortley phase for P_l^m, which makes all P_l^m real, and thus (A.15) differs by the factor $(-\mathrm{i})^m$ from the expression in Abramowitz (1965). Let us denote $x = m/l (\leqslant 1)$ and consider the integral

$$I = \int_{-\pi}^{\pi} (\cos \theta + \mathrm{i} \sin \theta \cos t)^l \, \mathrm{e}^{\mathrm{i}lxt} \, \mathrm{d}t \tag{A.16}$$

$$\equiv \int_{-\pi}^{\pi} \mathrm{e}^{lf(t)} \, \mathrm{d}t. \tag{A.17}$$

Then

$$f(t) = \mathrm{i}xt + \ln(\cos \theta + \mathrm{i} \sin \theta \cos t) \tag{A.18}$$

$$f'(t) = \mathrm{i}x + \frac{-\mathrm{i} \sin \theta \sin t}{\cos \theta + \mathrm{i} \sin \theta \cos t}. \tag{A.19}$$

For convenient notation we define β through

$$\tanh \beta = x. \tag{A.20}$$

Then $f'(t)$ vanishes at a point $t = t_0$ where

$$\sin(t_0 - i\beta) = \sinh\beta \cot\theta. \tag{A.21}$$

The real and imaginary parts of t_0 are

$$\begin{aligned} \operatorname{Re} t_0 &= \sin^{-1}(\sinh\beta\cot\theta) \\ \operatorname{Im} t_0 &= \beta. \end{aligned} \tag{A.22}$$

Thus the integrand in (A.17) has a saddle-point provided that $|\sinh\beta\cot\theta| \leqslant 1$ or

$$|x| \leqslant |\sin\theta|. \tag{A.23}$$

For $|x| < |\sin\theta|$ there are two saddle-points t_0^+, t_0^- situated symmetrically with respect to the line $t = \pi/2$ (figure 5).

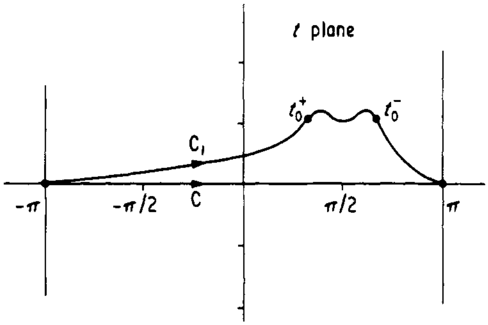

Figure 5. Path of integration and saddle-points in the evaluation of $P_l^m(\cos\theta)$ for large l, m.

The path of integration can be modified into C_1 in figure 5, and then the contributions outside the immediate vicinity of t_0^+ and t_0^- vanish for $l \to \infty$. We then use (A.9) after first computing

$$\begin{aligned} \exp(f(t_0^+)) &= \cosh\beta\, e^{-\beta\tanh\beta} \exp[i\cos^{-1}(\cosh\beta\cos\theta) \\ &\qquad + i\tanh\beta \operatorname{Arc\,sin}(\sinh\beta\cot\theta)] \\ \exp(f(t_0^-)) &= \cosh\beta\, e^{-\beta\tanh\beta} \exp[-i\cos^{-1}(\cosh\beta\cos\theta) \\ &\qquad + i\pi\tanh\beta - i\tanh\beta \operatorname{Arc\,sin}(\sinh\beta\cot\theta)]. \end{aligned} \tag{A.24}$$

Here Arc sin x means the branch of $\sin^{-1} x$ which gives function values between 0 and $\pi/2$. Also we find

$$f''(t_0^\pm) = \frac{(1-\cosh^2\beta\cos^2\theta)^{1/2}}{\cosh\beta} \exp\{\mp i[\cos^{-1}(\cosh\beta\cos\theta) + \tfrac{1}{2}\pi]\}. \tag{A.25}$$

Then (A.9) gives, for m even

$$\begin{aligned} I \simeq (\cosh\beta\, e^{-\beta\tanh\beta})^l \frac{[(2\pi/l)\cosh\beta]^{1/2}}{(1-\cosh^2\beta\cos^2\theta)^{1/4}} 2\cos[(l+\tfrac{1}{2}) \\ \times \cos^{-1}(\cosh\beta\cos\theta) + l\tanh\beta \operatorname{Arc\,sin}(\sinh\beta\cot\theta) - \tfrac{1}{4}\pi] \end{aligned} \tag{A.26}$$

and for m odd one has 2i sin( ) instead of 2 cos( ). Collecting the preceding formulae we get

$$P_l^m(\cos\theta) = \left(\frac{2}{\pi}\right)^{1/2}\frac{(l+m)!}{l!}(l^2\sin^2\theta - m^2)^{-1/4}\left(1-\frac{m^2}{l^2}\right)^{-1/2}\left(\frac{l-m}{l+m}\right)^{m/2}$$
$$\times\cos\left[(l+\tfrac{1}{2})\cos^{-1}\left(\frac{l}{(l^2-m^2)^{1/2}}\cos\theta\right)\right.$$
$$\left.-m\cos^{-1}\left(\frac{m}{(l^2-m^2)^{1/2}}\cot\theta\right)-\frac{\pi}{4}\right]. \tag{A.27}$$

Using Stirling's formula we get

$$P_l^m(\cos\theta) = \left(\frac{2}{\pi l}\frac{(l+m)!}{(l-m)!}\right)^{1/2}\left(\frac{l^2-m^2}{l^2\sin^2\theta - m^2}\right)^{1/4}$$
$$\times\cos\left[(l+\tfrac{1}{2})\cos^{-1}\left(\frac{l}{(l^2-m^2)^{1/2}}\cos\theta\right)\right.$$
$$\left.-m\cos^{-1}\left(\frac{m}{(l^2-m^2)^{1/2}}\cot\theta\right)-\frac{\pi}{4}\right]. \tag{A.28}$$

The special case $m = 0$ gives

$$P(\cos\theta) = \left(\frac{2}{\pi l\sin\theta}\right)^{1/2}\cos[(l+\tfrac{1}{2})\theta - \tfrac{1}{4}\pi]. \tag{A.29}$$

Equation (A.28) is valid for $|m| \leqslant l|\sin\theta|$. When $|m|$ increases past $l|\sin\theta|$, the saddle-points disappear and the value of $P_l^m(\cos\theta)$ decreases exponentially. In the limit $l \to \infty$ the region $|m| > l|\sin\theta|$ does not contribute and it can be neglected in the m summations.

References

Abramowitz M and Stegun I A 1965 *Handbook of Mathematical Functions* (New York: Dover)
Arfken G 1970 *Mathematical Methods for Physicists* (New York: Academic Press)
van Bladel J 1964 *Electromagnetic Fields* (New York: McGraw-Hill)
Born M and Wolf E 1964 *Principles of Optics* (Oxford: Pergamon)
Gradshteyn I S and Ryshik I M 1965 *Tables of Integrals, Series, and Products* (New York: Academic Press)
Watson G N 1948 *Theory of Bessel Functions* (Cambridge: Cambridge University Press)