Optica Acta: International Journal of Optics

ISSN: 0030-3909 (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tmop19

A Generalized Diffraction Grating Equation

A. Takahashi & T. Katayama

OPTICA ACTA, 1983, VOL. 30, NO. 12, 1735–1742

A generalized diffraction grating equation

A. TAKAHASHI and T. KATAYAMA
Department of Mathematics and Physics,
National Defense Academy, Yokosuka 239, Japan

(*Received 11 April 1983; revision received 6 June 1983*)

Abstract. A vector method of ray-tracing through unified gratings is presented. The method does not require any iterative calculation in the ray-transfer operations or any modification to the diffraction operations. The figuring of optical blanks is assumed to be quadratic or toroidal and rulings are considered to be represented by the intersections of a blank surface with a family of groove-generating surfaces. The generalized diffraction grating equation is expressed in terms of the gradient vector of the groove function.

1. Introduction

There are two ways of deriving the equations for tracing rays through diffraction gratings. One is the method developed by Welford [1], based on the application of Fermat's principle to the light path function associated with a grating, and the other is the so-called ray-vector method presented by Spencer and Murty [2]. Noda *et al.* [3] extended Welford's ray-tracing method to make it applicable to holographic gratings. The formulae are also applicable to mechanically ruled, variable-spacing gratings. Ludwig [4] developed a unified grating ray-tracing procedure using the Spencer and Murty's vector method, and Goto *et al.* [5] derived the diffracted-wavefront normal vector of an elementary grating.

The formulae given by Noda *et al.* [3] need some mathematical modification before they can be applied to mechanically ruled gratings. On the other hand, the vector forms given by Ludwig require two iterative computations to find the ray-surface intersection points and to obtain the local grating spacing of a holographic grating. Furthermore, the ruling-separation vector on the local plane grating must be determined from the vector relations between the groove-generating surface normal and the local surface normal.

The purpose in the present paper is to improve these deficiencies and to represent the generalized diffraction grating equation in vector form. When the blank surface is expressed by a quadratic or toroidal equation, the conventional transfer operations to this surface lead to an analytical solution of the point of incidence. It will also be shown that the diffracted ray vector is obtained by a linear combination of the vectors of the incident ray, the optical surface normal and the gradient of the ruling number expressed by the groove function $n=g(X, Y, Z)$.

2. Sign conventions and definitions

Let a surface S be defined by the equation $F(X, Y, Z)=0$ in a right-handed coordinate system (X, Y, Z) having its origin O at the vertex of S; let the Z axis be along the surface normal; and let the Y axis be parallel to the ruling at the origin (see figure 1). The radii of curvature in the meridional and sagittal planes at O are positive

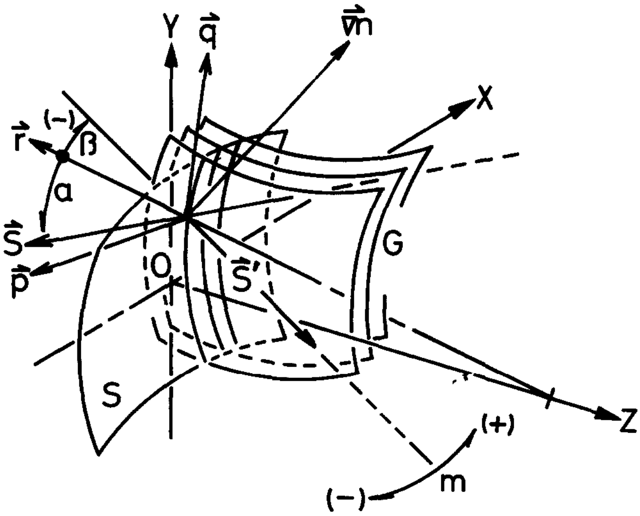

Figure 1. Sign conventions and definitions. The rulings are assumed to be defined by the intersections of the optical surface S with a family of groove-generating surfaces G.

if their centres of curvature are located on the positive Z axis, and the unit vector normal to the surface at the point of incidence is determined from a set of direction numbers whose components are given by the surface derivatives.

The unit vectors **S** and **S**$'$, with the direction cosines (I, J, K) and (I', J', K') of the incident and diffracted rays, will be taken in the direction of their progress. Angles of incidence, refraction, reflection and diffraction are taken to be positive when the Z axis must be rotated counterclockwise through an angle of less than $\pi/2$ to bring it into coincidence with the ray. The ratio $\mu = N/N'$ of the refractive indices of two media, N for the incident-ray side of the refracting surface and N' for the refracted-ray side, is taken to be unity in the case of reflection.

The sign of the mth order of diffraction follows a convention proposed by Wilson [6] (for details, see §4).

3. Transfer operations

3.1. *Quadratic surface*

We use the canonical form of quadric to express the surface figure of an optical blank:

for an ellipsoid

$$F(X, Y, Z,) = \frac{X^2}{A^2} + \frac{Y^2}{B^2} + \frac{(Z-C)^2}{C^2} - 1 = 0, \tag{1 a}$$

for an elliptic paraboloid

$$F(X, Y, Z) = \frac{X^2}{A^2} + \frac{Y^2}{B^2} - 2CZ = 0, \tag{1 b}$$

for a hyperboloid of two sheets

$$F(X, Y, Z) = \frac{X^2}{A^2} + \frac{Y^2}{B^2} - \frac{(Z+C)^2}{C^2} + 1 = 0. \tag{1 c}$$

In equations (1), A, B and C are determined from the specification of the optical surface which is defined in the region $Z \gtrless 0$ according to whether $C \gtrless 0$. Equations (1) can then be expressed as

$$F(X, Y, Z) = aX^2 + bY^2 + cZ^2 + uZ = 0, \tag{2}$$

where $a = 1/A^2$ and $b = 1/B^2$, and c and u are, respectively, the coefficients of Z^2 and Z in equations (1). By using the matrix method, as was done by Kastner and Neupert [7], the distance Δ between a point P′ on one surface S′, at which diffraction, reflection or refraction takes place, and a point P at which the ray from P′ intersects with the next surface S defined by equation (2), is determined from

$$(\mathbf{S}^{\mathrm{T}}\mathbf{Q}\mathbf{S})\Delta^2 + 2(\mathbf{S}^{\mathrm{T}}\mathbf{Q}\mathbf{W}_0)\Delta + (\mathbf{W}_0^{\mathrm{T}}\mathbf{Q}\mathbf{W}_0) = 0, \tag{3}$$

where the superscript T signifies a transposed vector. Here, we have used the two extended vectors $\mathbf{S}$ and $\mathbf{W}_0$, and the matrix $\mathbf{Q}$, defined in terms of the coefficients of equation (2):

$$\mathbf{S} = \begin{bmatrix} I \\ J \\ K \\ 0 \end{bmatrix}, \quad \mathbf{W}_0 = \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \\ 1 \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & u \\ 0 & 0 & u & 0 \end{bmatrix}, \tag{4}$$

where $\mathbf{W}_0$ is the vector directed from the vertex of the surface S to the point P′ on the previous surface S′ or to the object point. The coordinate vector $\mathbf{W}$ joins the vertex of the surface S to the incident point P on S and is given by

$$\mathbf{W} = \mathbf{W}_0 + \Delta\mathbf{S}, \quad \Delta > 0. \tag{5}$$

3.2. *Toroidal surface*

An elliptic toroidal surface [2, 8, 9] is generated by revolving an ellipse lying in the Y–Z plane about an axis Y' which is located in the Y–Z plane at a distance R from the origin and parallel to the Y axis (figure 2). Let the equation for the ellipse be $\rho^2(Z-r)^2 + r^2Y^2 = (r\rho)^2$, r and ρ being the semi-major and semi-minor axes, respectively. The toroidal equation is then given by

$$F(X, Y, Z) = \rho^2\{[(Z+u)^2 + X^2]^{1/2} - (R-r)\}^2 + r^2(Y^2 - \rho^2) = 0, \tag{6}$$

in which u is the parameter for generating toric surfaces of four types. As shown in figure 2, if $r < R < 2r$, one of $S_1(u = -R)$, $S_2(u = R - 2r)$, $S_3(u = 2r - R)$ and $S_4(u = R)$ is chosen for the blank surface S; when $0 < R < r$, one of $S_1(u = R - 2r)$, $S_2(u = -R)$, $S_3(u = R)$ and $S_4(u = 2r - R)$ is used.

We have derived the analytical ray-transfer equation to this surface to be

$$\begin{aligned}(\mathbf{S}^{\mathrm{T}}\mathbf{S})^2\Delta^4 &+ 4(\mathbf{S}^{\mathrm{T}}\mathbf{S})(\mathbf{W}_0^{\mathrm{T}}\mathbf{S})\Delta^3 \\ &+ [4(\mathbf{W}_0^{\mathrm{T}}\mathbf{S})^2 + 2(\mathbf{W}_0^{\mathrm{T}}\mathbf{W}_0)(\mathbf{S}^{\mathrm{T}}\mathbf{S}) - 2\xi(\mathbf{S}^{\mathrm{T}}\mathbf{S} + 2(\xi+\eta)\varepsilon^2J^2]\Delta^2 \\ &+ [4(\mathbf{W}_0^{\mathrm{T}}\mathbf{S})(\mathbf{W}_0^{\mathrm{T}}\mathbf{W}_0) - 4\xi(\mathbf{W}_0^{\mathrm{T}}\mathbf{S}) + 4(\xi+\eta)\varepsilon^2JY_0]\Delta \\ &+ [(\mathbf{W}_0^{\mathrm{T}}\mathbf{W}_0)^2 - 2\xi(\mathbf{W}_0^{\mathrm{T}}\mathbf{W}_0) + 2(\xi+\eta)\varepsilon^2Y_0^2 + \eta^2] = 0,\end{aligned} \tag{7}$$

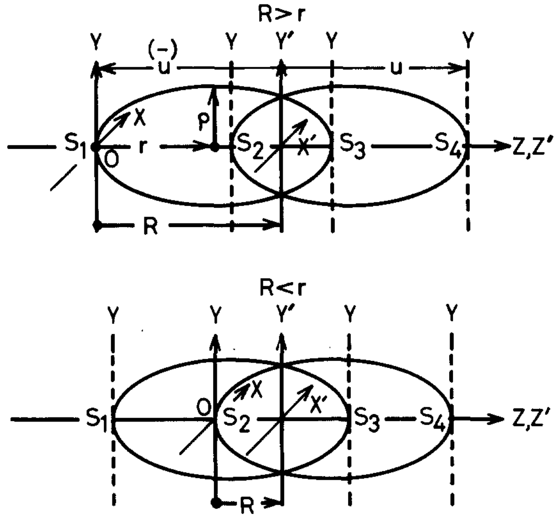

Figure 2. Generation of elliptic toroidal surfaces. Revolving an ellipse (of semi-major axis r and semi-minor axis ρ) about an axis Y' lying in the Y–Z plane at a distance R from the origin O, we have four elliptic toroidal surfaces whose vertices are at $S_1, \ldots, S_4$. These surfaces are represented by setting $u = \pm R$ and $u = \pm(2r - R)$ in equation (6).

where $\varepsilon = r/\rho$, $\xi = (R-r)^2 + r^2$, $\eta = (R-r)^2 - r^2$, and

$$\mathbf{S} = \begin{bmatrix} I \\ \varepsilon J \\ K \end{bmatrix}, \quad \mathbf{W}_0 = \begin{bmatrix} X_0 \\ \varepsilon Y_0 \\ Z_0 + u \end{bmatrix} \tag{8}$$

The coordinate vector $\mathbf{W}$ on the toroidal surface is also obtained from equation (5).

3.3. *Proper value of* Δ

We choose the proper value of Δ in a manner slightly different from the automatic ray-intersection method [10]. The procedure we adopted is as follows.

(1) We find all the positive roots Δ_i of equation (3) or(7), and, using equation (5), we calculate the coordinates (X_i, Y_i, Z_i) of the intersection point P_i $(1 \leqslant i \leqslant 4)$ of the incident ray with the quadratic or toroidal surface.

(2) We find the value of Δ_i for which the intersection points P_i $(1 \leqslant i \leqslant 4)$ fall within the actual dimensions of the surface.

(3) If we find only one value of Δ_i in stage (2), we take this value to be the proper value. In other cases, we proceed to the next step.

(4) We calculate $U = uK$, the product of the component K of the direction cosines of ray vector $\mathbf{S}$ and the parameter u given by equation (2) or (7).

(5) If the sign of U is positive (the blank surface is concave towards the incoming ray), we take the largest value of Δ_i as the proper value; if the sign of U is negative (convex), we take smallest value.

4. Diffraction operations

The grooves of any grating are considered to be the intersections of the blank surface $F(X, Y, Z) = 0$ with a family of groove-generating surfaces $G(X, Y, Z; n) = 0$, in which n is specified by the groove number counted from the origin (see figure 1). If the groove function $n=g(X, Y, Z)$ is assumed to be continuous and differentiable, the direction of the gradient vector $\nabla\mathbf{n}$ is in the direction of the groove-generating surface normal.

Now, we shall incorporate the vector $\nabla\mathbf{n}$ into the well-known diffraction grating equation of the local plane grating [2] at a point P on the blank, which is given by

$$(\mathbf{S}' \times \mathbf{r}) = \mu(\mathbf{S} \times \mathbf{r}) + (m\lambda/N'd)\mathbf{q}. \tag{9}$$

In equation (9), λ is the wavelength, m the order of diffraction, d the local plane grating spacing, $\mu = N/N'$ the ratio of the refractive indices, $\mathbf{q}$ the unit vector parallel to the ruling and pointing to the region $Y > 0$, and $\mathbf{r}$ the unit vector normal to the plane grating. If the vector product $(N'\mathbf{S}' - N\mathbf{S}) \times \mathbf{r}$ and the unit vector $\mathbf{q}$ in equation (9) point in the same direction at the vertex O, the sign of m is positive. The unit vector $\mathbf{r}$ is determined from the normalization of a set of direction numbers with components (F_X, F_Y, F_Z). We then define a unit vector $\mathbf{p}$, $\mathbf{p} = \mathbf{q} \times \mathbf{r}$.

The gradient $\nabla\mathbf{n}$ at the point P is related to the unit vectors $\mathbf{q}$ and $\mathbf{r}$ through

$$\mathbf{q} = \frac{\nabla\mathbf{n} \times \mathbf{r}}{|\nabla\mathbf{n} \times \mathbf{r}|} \quad \text{and} \quad \frac{1}{d} = |\nabla\mathbf{n} \times \mathbf{r}|. \tag{10}$$

Substituting equation (10) into equation (9) and rearranging, we obtain the generalized diffraction grating equation

$$\mathbf{S}' = \mu\mathbf{S} + \Lambda\nabla\mathbf{n} + \Gamma\mathbf{r}, \tag{11}$$

where Γ is an undetermined multiplier and $\Lambda = m\lambda/N'$. Squaring equation (11), we have that

$$\Gamma^2 + 2p\Gamma + q = 0, \tag{12 a}$$

where

$$p = \mu(\mathbf{S} \cdot \mathbf{r}) + \Lambda(\mathbf{r} \cdot \nabla\mathbf{n}), \tag{12 b}$$

and

$$q = \Lambda^2(\nabla\mathbf{n})^2 + 2\mu\Lambda(\mathbf{S} \cdot \nabla\mathbf{n}) + \mu^2 - 1. \tag{12 c}$$

According to whether $p > 0$ or $p < 0$, the negative or positive root of equation (12 *a*) must be taken for a reflection grating and a mirror. For a transmission grating and a lens, the sign of the proper root of equation (12 *a*) should be opposite to the above case.

5. Applications

To calculate the diffracted ray vector $\mathbf{S}'$ for gratings with variable spacing and curved grooves, it is necessary to determine the gradient vector $\nabla\mathbf{n}$.

5.1. *Holographic gratings*

As shown in figure 3 (*a*), the nth groove at the point $P(X, Y, Z)$ is expressed by

$$n = g(X, Y, Z) = \frac{1}{\lambda_0}\{[\langle CP\rangle - \langle DP\rangle] - [\langle CO\rangle - \langle DO\rangle]\}, \tag{13}$$

where λ_0 is the wavelength of laser light from the two point sources C and D, and

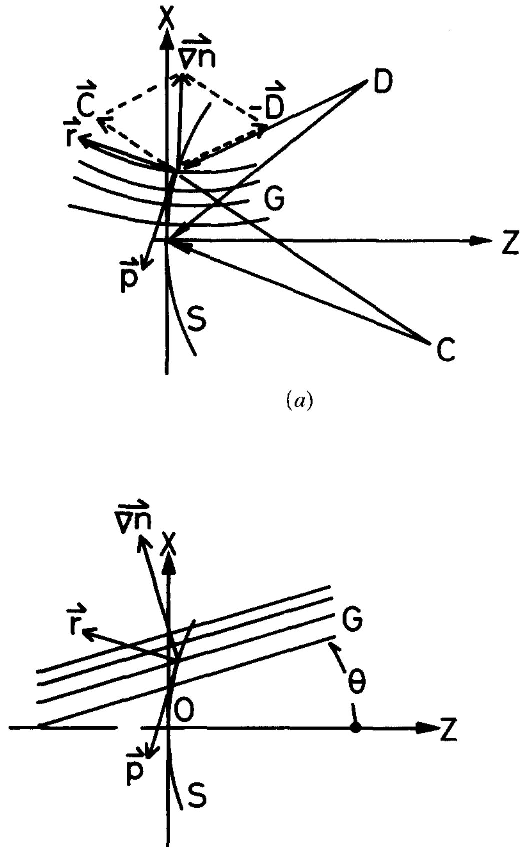

Figure 3. The formation of grooves, indicated in the meridional plane. Three vectors **p**, **r** and $\nabla\mathbf{n}$ are in the meridional plane. (*a*) Holographic grating whose grooves are formed in accordance with the interference pattern G on the blank S due to two point sources C and D. (*b*) A mechanically ruled grating consisting of the intersections of the blank surface S with a set of parallel planes G having a tilt angle θ and variable spacing.

$\langle CP\rangle$, $\langle DP\rangle$, $\langle CO\rangle$ and $\langle DO\rangle$ are the distances between the various points. Thus, the gradient vector $\nabla\mathbf{n}$ for the holographic grating is

$$\nabla\mathbf{n}=(1/\lambda_0)(\mathbf{C}-\mathbf{D}), \tag{14}$$

C and **D** being the unit vectors of the rays CP and DP, respectively. Therefore, the diffraction equation of the holographic grating is

$$\mathbf{S}'=\mu\mathbf{S}+(m\lambda/\lambda_0)(\mathbf{C}-\mathbf{D})+\Gamma\mathbf{r}. \tag{15}$$

This equation agrees with equation (35) in [3] if the difference in the sign convention with respect to μ is taken into account (see also the errata [11]).

5.2. *Mechanically ruled gratings*

Aberration-corrected concave gratings have been produced by Harada and Kita [12], using a numerically controlled ruling engine. The nth groove, as shown in figure 3 (*b*), can be represented by the equation

$$n=\frac{1}{d_0}\sum_{i=1}^{\infty} g_i(X-Z\tan\theta)^i, \tag{16}$$

where θ is the angle at which the groove-generating planes are inclined to the Y–Z plane, g_i the ruling parameters for controlling the variable spacing and d_0 the grating spacing at the vertex of the blank. The components of $\nabla\mathbf{n}$ are obtained from the derivatives of equation (16) with respect to the coordinates (X, Y, Z).

If the values $g_1=1$, $g_2=0,\ldots,g_i=0$, and $\theta=0$ are inserted in equation (11), the diffraction equation for mechanically ruled gratings having constant spacing and straight grooves is obtained.

5.3. *Variable spacing grating with concentric circular rulings*

In this case, the grating rulings are defined by the intersections of the blank surface with a set of concentric cylinders [2, 4]. The *n*th ruling is given by

$$n=\frac{1}{\rho_0}\sum_{i=1}^{\infty} \rho_i(X^2+Y^2)^{i/2}, \tag{17}$$

ρ_i being an adjustable parameter for the radial separation and ρ_0 the radial grating constant at the vertex of the blank.

6. Conclusions

The ray-transfer matrix method is applicable not only to the image plane, but also to quadratic and toroidal optical surfaces. This enables one to calculate the incident points on the aspherical surface without using the iteration technique [2, 4, 10].

The diffraction grating equation (11) can easily be applied to gratings of any type, without modification and with no need to perform iterative calculations of the local grating constant, simply by defining the groove function.

Acknowledgments

One of the authors (A. T.) wishes to thank Dr. S. Naito for helpful discussions. The present work was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education.

On présente une méthode vectorielle de tracé des rayons au travers de réseaux unifiés. Le méthode ne nécessite aucun calcul itératif dans les opérations de transfert de rayon et aucune modification dans les opérations de diffraction. La forme des supports optiques est supposée quadratique ou torique et les réseaux de traits sont considérés être représentés par les intersections d'une surface support avec une famille de surfaces génératrices de sillons. L'équation du réseau de diffraction généralisé est exprimée en utilisant le vecteur gradient de la fonction sillon.

Es wird eine Vektormethode zur Strahldurchrechnung durch vereinheitlichte Gitter präsentiert. Die Methode benötigt keinerlei iterative Rechnung bei den Strahlübertragungsoperationen und keinerlei Modifizierung bei den Beugungsoperationen. Die Form der Rohlinge wird als quadratisch oder torisch angenommen und Gitterteilungen werden durch Schnitte einer glatten Oberfläche mit einer Schar von furchenerzeugenden Flächen repräsentiert. Die verallgemeinerte Gittergleichung wird durch den Gradientenvektor der Furchenfunktion ausgedrückt.

References

[1] Welford, W. T., 1962, *Optica Acta*, **9**, 389.
[2] Spencer, G. H., and Murty, M. V. R. K., 1962, *J. opt. Soc. Am.*, **52**, 672.
[3] Noda, H., Namioka, T., and Seya, M., 1974, *J. opt. Soc. Am.*, **64**, 1037.
[4] Ludwig, U. W., 1973, *J. opt. Soc. Am.*, **63**, 1105.
[5] Goto, K., Kato, Y., and Togawa, K., 1979, *Optica Acta*, **26**, 841.
[6] Wilson, M. E., 1982, *Appl. Optics*, **21**, 537.
[7] Kastner, S. O., and Neupert, W. M., 1963, *J. opt. Soc. Am.*, **53**, 1180.
[8] Haber, H., 1950, *J. opt. Soc. Am.*, **40**, 153.
[9] Singh, K., and Chopra, K. N., 1970, *Optik*, **31**, 241.
[10] Howell, B. J., and Wilson, M. E., 1982, *Appl. Optics*, **21**, 2184.
[11] Noda, H., Namioka, T., and Seya, M., 1975, *J. opt. Soc. Am.*, **65**, 474.
[12] Harada, T., and Kita, T., 1980, *Appl. Optics*, **19**, 3987.