Convergence performance of the coupled-wave and the differential methods for thin gratings

Philippe Lalanne

Convergence performance of the coupled-wave and the differential methods for thin gratings

Philippe Lalanne

Institut d'Optique Théorique et Appliquée, Centre National de la Recherche Scientifique, BP 147,
91403 Orsay Cedex, France

The new formulation of the coupled-wave analysis recently proposed by Lalanne and Morris [J. Opt. Soc. Am. A **13**, 779 (1996)] and by Granet and Guizal [J. Opt. Soc. Am. A **13**, 1019 (1996)] that drastically improves the convergence performance of the method for lamellar gratings and for TM polarization is shown to be badly conditioned for gratings with a small thickness. Numerical evidence obtained with the coupled-wave analysis and with the differential methods for several grating diffraction problems shows that, in some cases that we identify, the convergence of the conventional formulation can be faster than that of the new one. The discussion includes lamellar, multilevel binary, and continuous-profile geometries. © 1997 Optical Society of America [S0740-3232(97)00407-9]

Key words: coupled-wave method, differential method, rigorous grating analysis, diffractive optics.

1. INTRODUCTION

Recently a striking improvement of the convergence rate of the coupled-wave method for TM polarization (magnetic-field vector perpendicular to the grating vector) was reported by several authors.[1,2] This improvement was obtained by reformulation of the eigenvalue problem in the grating region. In the conventional formulation,[3] the infinite set of second-order differential equations for the magnetic field is written as

$$k_0{}^{-2}[\mathbf{U}''] = [\mathbf{E}(\mathbf{K}_x \mathbf{E}^{-1} \mathbf{K}_x - \mathbf{I})][\mathbf{U}]. \tag{1}$$

In the new formulation, Eq. (1) is replaced by

$$k_0{}^{-2}[\mathbf{U}''] = [\mathbf{A}^{-1}(\mathbf{K}_x \mathbf{E}^{-1} \mathbf{K}_x - \mathbf{I})][\mathbf{U}]. \tag{2}$$

In Eqs. (1) and (2), $\mathbf{I}$ is the identity matrix; $\mathbf{K}_x$, a diagonal matrix; $\mathbf{E}$, a Toeplitz matrix formed by the permittivity harmonic coefficients of the relative permittivity ε_m; and $\mathbf{A}$, a Toeplitz matrix formed by the inverse-permittivity harmonic coefficients a_m. $\mathbf{U}$ is a vector formed by the space-harmonic coefficients U_m of the magnetic field, and k_0 represents the magnitude of the incident plane-wave vector in vacuum. These are rather standard notations used in Refs. 1 and 3, and they are used in Ref. 2 with minor changes. The difference between the conventional and the new formulations is small: The new formulation uses the matrix $\mathbf{A}^{-1}$ instead of the matrix $\mathbf{E}$. To my knowledge, two interpretations are available to explain why the use of matrix $\mathbf{A}^{-1}$ produces a striking difference of performance in terms of convergence rate. A simple intuitive argument was given in Ref. 1 based on the study of the quasi-static limit (grating period infinitely small compared with the wavelength). A similar discussion can be found in Ref. 4. In the quasi-static limit, gratings are equivalent to thin films, and the diffracted amplitudes and the electromagnetic fields can be derived analytically;

no evanescent orders have to be taken into account. The authors of Ref. 1 showed that, with the conventional formulation, an accurate description of the quasi-static limit requires that all the evanescent orders be retained in the computation. In contrast, as can be expected from the thin-film analogy, it is sufficient to retain only the zeroth orders to describe adequately the quasi-static limit properties of the grating with the new formulation. The authors of Ref. 1 concluded that the conventional formulation presents some kind of bad conditioning, and that is why they reformulated the eigenvalue problem. To avoid any confusion with the inverse-problem literature, "bad conditioning" in this paper refers to a formulation in which an infinite number of orders are necessary to modelize properly a situation (asymptotic limit) that can in principle be modelized by retaining only the zeroth orders. Similarly, "good conditioning" refers to a formulation in which only the zeroth orders are sufficient to modelize the situation. A more mathematical explanation of the performance improvement of the new formulation was given by Li in a recent paper.[5] The author showed that the new formulation converges faster because it uniformly satisfies the boundary conditions in the grating region, whereas the conventional formulation does so nonuniformly. This is obviously a strong and surprising argument. It has interesting consequences; if one is interested in visualizing the electromagnetic fields inside the grating region, the new formulation offers a good framework to compute accurately the electric and magnetic vectors and the electric displacement. At first sight, the intuitive argument used in the quasi-static limit and the mathematical derivation of Li are not correlated. For instance, the intuitive argument predicts that the new formulation provides better performance even for nonlamellar gratings with continuous relative permittivi-

ties; Li's analysis does not consider this situation. The fact that the two interpretations both argue in favor of the new formulation is happily a good coincidence.

In this paper it is shown that, in some situations, the new formulation exhibits slow convergence rates, which are even slower than those obtained with the conventional formulation. In Section 2 the TM-polarization case of gratings is analytically studied in the small depth limit (grating depth h much smaller than the wavelength, $h k_0 \ll 1$). It is shown that, with the new formulation, an infinite number of orders has to be retained in the computation to modelize correctly the small depth limit. On the contrary, the conventional formulation is shown to be well conditioned. This situation is similar to the quasistatic limit, but opposite conclusions are obtained about the performances of the conventional and the new formulations. In Section 3 a highly conductive grating is considered with different depth values, and the convergence rates of the new and the conventional formulations are compared in each case. Section 4 provides numerical evidence that, for multilevel binary and continuous-profile gratings, the grating depth is a parameter of primary importance for the convergence rates of the two formulations. Section 5 concludes the paper and includes a brief discussion of other polarization problems, such as the conical case.

2. SMALL DEPTH LIMIT OF GRATINGS FOR TRANSVERSE-MAGNETIC POLARIZATION

The one-dimensional grating diffraction problem considered in Sections 2 and 3 is depicted in Fig. 1. The grating region is composed of a one-dimensional periodic structure along the x axis with an arbitrary permittivity profile $\varepsilon(x)$. It is bound by two different media with refractive indices n_1 and n_3. The z axis is perpendicular to the grating boundaries, and the diffraction problem is invariant in the y direction. A temporal dependence of $\exp(j\omega t)$ of the incident wave is assumed ($j^2 = -1$). Magnetic effects are not considered in this paper, and the constant μ_0 denotes the permeability of the periodic structure and of the surrounding regions. We denote by
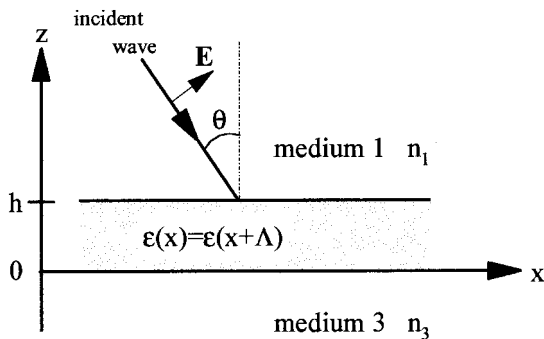

Fig. 1. Geometry for the nonconical grating diffraction problem analyzed in Sections 2 and 3 for TM polarization. The relative permittivity is assumed to be independent of the z variable. The refractive indices of the incident medium and of the substrate are n_1 and n_3, respectively. The grating depth is h, and θ denotes the angle of incidence.

c the velocity of light in vacuum. The following notations are those of Ref. 1. Normal incidence ($\theta = 0$) is assumed in this section. The grating period is denoted by Λ, and the length of the grating vector K is equal to $2\pi/\Lambda$. Using the Floquet theorem, we can express the x component E_x of the electric field and the y component H_y of the magnetic field in the grating region as

$$E_x = \sum_m S_m(z) \exp jKmx, \tag{3a}$$

$$H_y = \frac{1}{\mu_0 c} \sum_m U_m(z) \exp jKmx. \tag{3b}$$

For the conventional formulation, the first-order differential equations for E_x and H_y are

$$\frac{U'_m}{k_0} = -j \sum_p \varepsilon_{m-p} S_p, \tag{4a}$$

$$\frac{S'_m}{k_0} = jm\alpha^2 \sum_p p \varepsilon_{m,p}^{(-1)} U_p - jU_m. \tag{4b}$$

Similarly, for the new formulation, we have

$$\frac{U'_m}{k_0} = -j \sum_p a_{m,p}^{(-1)} S_p, \tag{5a}$$

$$\frac{S'_m}{k_0} = jm\alpha^2 \sum_p p \varepsilon_{m,p}^{(-1)} U_p - jU_m. \tag{5b}$$

In Eqs. (4) and (5), α is the wavelength-to-period ratio ($\alpha = K/k_0$), and $\varepsilon_{m,p}^{(-1)}$ and $a_{m,p}^{(-1)}$ are the (m, p)th coefficients of matrices $\mathbf{E}^{-1}$ and $\mathbf{A}^{-1}$, respectively. The difference between Eqs. (5) and (4) is rather small: Equations (5b) and (4b) are identical, and the coefficients ε_{m-p} in Eq. (4a) are replaced by the coefficients $a_{m,p}^{(-1)}$ in Eq. (5a).

The y components $H_{1,y}$ and $H_{3,y}$ of the magnetic fields in regions 1 and 3 are given by

$$\mu_0 c H_{1,y} = \exp[jn_1 k_0(z - h)] + \sum_m R_m$$
$$\times \exp[-jk_0(n_1^2 - m^2\alpha^2)^{1/2}(z - h)]$$
$$\times \exp(jmKx), \tag{6a}$$

$$\mu_0 c H_{3,y} = \sum_m T_m \exp[jk_0(n_3^2 - m^2\alpha^2)^{1/2}z]$$
$$\times \exp(jmKx), \tag{6b}$$

respectively. In Eqs. (6), $\mathrm{Re}(n_i^2 - m^2\alpha^2)^{1/2} - \mathrm{Im}(n_i^2 - m^2\alpha^2)^{1/2} > 0$ for $i = 1, 3$.

We now proceed with the solution of the grating diffraction problem in the limit in which the grating depth is small compared with the wavelength ($h k_0 \ll 1$). To avoid any confusion with related work,[6] in this paper the expression small depth limit is related to a situation such that the grating thickness is made infinitely small while all the other grating parameters (k_0, K, n_1, n_3, and ε_i, with $i = -\infty...\infty$) are held constant. We assume that the mth space-harmonic field components in the grating region can be expanded into a power series in terms of h ($h k_0 \ll 1$). Consequently, we have

$$S_m(h) = S_m(0) + hk_0 \frac{S'_m(0)}{k_0} + O(h^2 k_0{}^2), \quad (7a)$$

$$U_m(h) = U_m(0) + hk_0 \frac{U'_m(0)}{k_0} + O(h^2 k_0{}^2). \quad (7b)$$

Equations (7) provide a linear relationship between the electromagnetic-field components at the two boundaries of the grating region. We also assume that the reflection and the transmission coefficients, R_m and T_m, can be expanded in a similar way:

$$R_m = R_m^{(0)} + hk_0 R_m^{(1)} + O(h^2 k_0{}^2), \quad (8a)$$

$$T_m = T_m^{(0)} + hk_0 T_m^{(1)} + O(h^2 k_0{}^2). \quad (8b)$$

The quantities $R_m^{(0)}$, $R_m^{(1)}$, $T_m^{(0)}$, $T_m^{(1)}$, $S_m(0)$, and $U_m(0)$ are unknown constants to be determined from the boundary conditions (continuity of the tangential electric- and magnetic-field components at the grating boundaries). $R_m^{(0)}$ and $T_m^{(0)}$ are the reflection and the transmission coefficients of the mth space-harmonic field components for a null grating thickness. $R_m^{(1)}$ and $T_m^{(1)}$ are dimensionless coefficients.

A. Solutions with the Conventional Formulation
By means of Eqs. (4), Eqs. (7) become

$$S_m(h) = S_m(0) + hk_0 \left[jm\alpha^2 \sum_p p\varepsilon_{m,p}^{(-1)} U_p(0) - jU_m(0) \right] + O(h^2 k_0{}^2), \quad (9a)$$

$$U_m(h) = U_m(0) + hk_0 \left[-j \sum_p \varepsilon_{m-p} S_p(0) \right] + O(h^2 k_0{}^2). \quad (9b)$$

By matching the boundary conditions and neglecting the terms $O(h^2 k_0{}^2)$, we obtain

$$U_m(0) + hk_0 \left[-j \sum_p \varepsilon_{m-p} S_p(0) \right]$$
$$= \delta_{m,0} + R_m^{(0)} + hk_0 R_m^{(1)}, \quad (10a)$$

$$U_m(0) = T_m^{(0)} + hk_0 T_m^{(1)}, \quad (10b)$$

$$S_m(0) = -\frac{1}{n_3{}^2} \sqrt{n_3{}^2 - m^2\alpha^2} [T_m^{(0)} + hk_0 T_m^{(1)}], \quad (10c)$$

$$S_m(0) + hk_0 \left[jm\alpha^2 \sum_p p\varepsilon_{m,p}^{(-1)} U_p(0) - jU_m(0) \right]$$
$$= \frac{1}{n_1{}^2} \{ -n_1 \delta_{m,0} + \sqrt{n_1{}^2 - m^2\alpha^2}$$
$$\times [R_m^{(0)} + hk_0 R_m^{(1)}] \}. \quad (10d)$$

$\delta_{m,0}$ is the Kronecker symbol. By eliminating $U_m(0)$ and $S_m(0)$ between Eqs. (10), we find that

$$\delta_{m,0} + R_m^{(0)} - T_m^{(0)} + hk_0$$
$$\times \left[-j \sum_p \varepsilon_{m-p} \frac{1}{n_3{}^2} \sqrt{n_3{}^2 - p^2\alpha^2} T_p^{(0)} + R_m^{(1)} - T_m^{(1)} \right] = 0, \quad (11a)$$

$$-\frac{1}{n_3{}^2} \sqrt{n_3{}^2 - m^2\alpha^2} T_m^{(0)} + \frac{1}{n_1} \delta_{m,0}$$
$$- \frac{1}{n_1{}^2} \sqrt{n_1{}^2 - m^2\alpha^2} R_m^{(0)}$$
$$+ hk_0 \left[-\frac{1}{n_3{}^2} \sqrt{n_3{}^2 - m^2\alpha^2} T_m^{(1)} \right.$$
$$+ jm\alpha^2 \sum_p p\varepsilon_{m,p}^{(-1)} T_p^{(0)} - jT_m^{(0)}$$
$$\left. - \frac{1}{n_1{}^2} \sqrt{n_1{}^2 - m^2\alpha^2} R_m^{(1)} \right] = 0. \quad (11b)$$

Solving the system of Eqs. (11) is done straightforwardly. Identifying terms of order zero, we obtain

$$R_m^{(0)} = \frac{n_3 - n_1}{n_3 + n_1} \delta_{m,0}, \quad (12a)$$

$$T_m^{(0)} = \frac{2n_3}{n_3 + n_1} \delta_{m,0}. \quad (12b)$$

Equations (11) allow us to derive $T_m^{(1)}$ and $R_m^{(1)}$ for any value of m, but we restrict the computation to the derivation of $T_0^{(1)}$ and $R_0^{(1)}$ for clarity. Identifying terms of order one in Eqs. (11) and incorporating Eqs. (12), we find that

$$R_0^{(1)} = j \left(\frac{\varepsilon_0}{n_3} - \frac{n_3 n_1 + \varepsilon_0}{n_3 + n_1} \right) T_0^{(0)}, \quad (13a)$$

$$T_0^{(1)} = -j \frac{n_3 n_1 + \varepsilon_0}{n_3 + n_1} T_0^{(0)}. \quad (13b)$$

B. Solutions with the New Formulation
We straightforwardly derive the calculus of the reflection and the transmission coefficients with the new formulation from the previous calculation by noting that Eqs. (5) are identical to Eqs. (4) if the coefficient ε_{m-p} in Eq. (4a) is replaced by the coefficient $a_{m,p}^{(-1)}$ in Eq. (5a). Consequently, the expression of the reflection and the transmission coefficients [Eqs. (12) and (13)] holds with the new formulation, except that ε_0 has to be replaced by $a_{0,0}^{(-1)}$ in Eqs. (13).

Of course, when an infinite number of orders is retained in the computation, ε_0 and $a_{0,0}^{(-1)}$ are equal,[5] and the two formulations provide the same (exact) expressions for $R_0^{(1)}$ and $T_0^{(1)}$. However, for numerical purposes, only a finite number M of orders are retained in practice. We

denote by $a_{M,0,0}^{(-1)}$ [$a_{M,0,0}^{(-1)} \to a_{0,0}^{(-1)}$ when $M \to \infty$] the (0, 0)th coefficient of matrix $\mathbf{A}^{-1}$ when M orders are retained for the computation. ε_0 and $a_{M,0,0}^{(-1)}$ are different in general. Consequently, when using the new formulation for gratings with a small depth ($hk_0 \ll 1$), one needs to retain all the orders to compute the exact values of coefficients $R_0^{(1)}$ and $T_0^{(1)}$. In contrast, when using the conventional formulation, one obtains exact values whatever the number of retained orders is ($M \geqslant 1$). Although it was not considered in the previous derivations for purposes of clarity, similar conclusions hold for the mth transmitted and reflected amplitude coefficients, which are accurately derived by use of the new formulation only if an infinite number of orders are retained. The situation is similar to that of the quasi-static-limit case, except that, in the small depth limit, opposite conclusions are obtained: The conventional formulation is well conditioned, and the new one is not.

The previous discussion can be stated in a different way. From the expansion series of Eqs. (8) and from the expressions of $R_0^{(1)}$ and $T_0^{(1)}$ derived in this section for the new and conventional formulations, it is straightforwardly shown that for normal incidence

Proposition (1): For any truncation rank M, it is possible to define a depth h_0 so that, for any depth smaller than the depth h_0, the conventional formulation provides more-accurate computational results of the zeroth-order normalized electric-field amplitudes R_0 and T_0 than does the new formulation.

In fact, Proposition (1) holds for any normalized electric-field amplitudes R_m and T_m. For arbitrary angles of incidence, the simple approach used in this section does not permit us to derive mathematically that one formulation will be more accurate than the other one. It just can be shown that

Proposition (2): For any truncation rank M, it is possible to choose a domain of incidence near $\theta = 0$, so that for any depth smaller than a given depth h_0, the conventional formulation provides more-accurate computational results of the zeroth-order normalized electric-field amplitudes R_0 and T_0 than does the new formulation.

Thus Proposition (1) does not generalize to arbitrary angles of incidence, but, as is shown in the next two sections, even for large angles of incidence, a significant degradation of the convergence rate of the new formulation is observed for small grating thicknesses.

Before considering numerical examples, let us summarize the results obtained in this section and in previous related studies.[1,7] The new formulation is well conditioned for small period-to-wavelength ratios, is badly conditioned for small depth-to-wavelength ratios, and uniformly matches the boundary conditions for lamellar gratings. When thin lamellar gratings are considered the new formulation raises a conflict. In contrast, the conventional formulation is well conditioned for small depth-to-wavelength ratios, is badly conditioned for small period-to-wavelength ratios, and does uniformly match the boundary conditions for lamellar gratings. In Sec-

tion 3 the convergence rates of the conventional and the new formulations are compared for lamellar gratings and for different grating depths.

3. NUMERICAL EXAMPLES: LAMELLAR GRATINGS

For numerical purposes, we consider a difficult diffraction problem that has been extensively discussed in the context of convergence rates of the coupled-wave method.[1,2,8] The diffraction configuration is a 30° incidence angle and a highly conductive grating etched in a gold substrate ($n_3 = 3.18 + j4.41$) is considered. The grating period is 1 μm, the duty cycle (defined as the groove-width-to-period ratio) is 0.5, and the wavelength is chosen equal to 1 μm, so that the grating depth is also equal to the depth-to-wavelength ratio. Only the negative first and zeroth reflected orders are propagating.

Figures 2–4 show the total reflected intensities (the deviation from 1 represents the absorption) as a function of the number of retained orders for $h = 0.01\lambda$, 0.05λ, 0.5λ, respectively. The solid curves are obtained with the new formulation. Plus signs correspond to the conventional formulation. For $h = 0.01\lambda$ (Fig. 2), the new formulation converges slowly; from 3 to 40 retained orders, and some oscillations are observed. The amplitude of the oscillations decreases as the number of retained orders in-
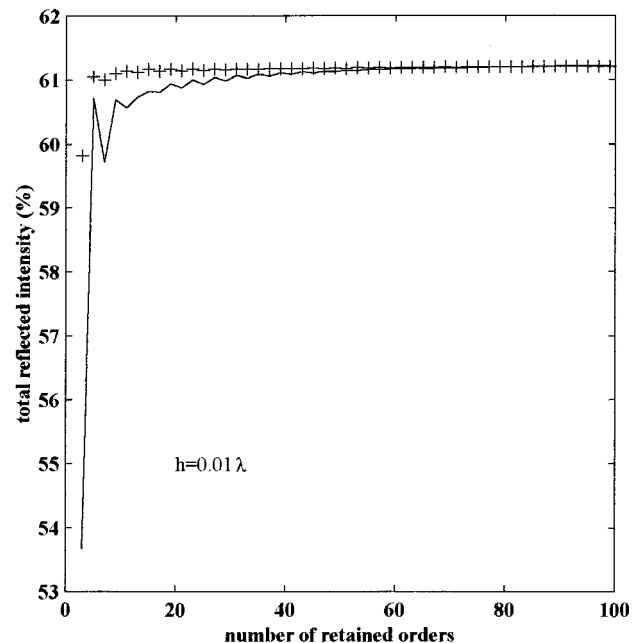

Fig. 2. Total reflected intensity (including the zeroth and the minus-first orders) as a function of the number of retained orders for a metallic grating and TM-polarized light. The solid curve is obtained with the new eigenproblem formulation. The plus signs are data provided by the conventional eigenproblem formulation. The lamellar grating considered for the computation is etched in the substrate (the relative permittivity is either $n_1{}^2$ or $n_3{}^2$) and has a fill factor of 0.5 (the fill factor is defined as the groove-width-to-period ratio). $n_1 = 1$, $n_3 = 3.18 + j4.41$, $\theta = 30°$, $\Lambda = \lambda$, and $h = 0.01\lambda$.
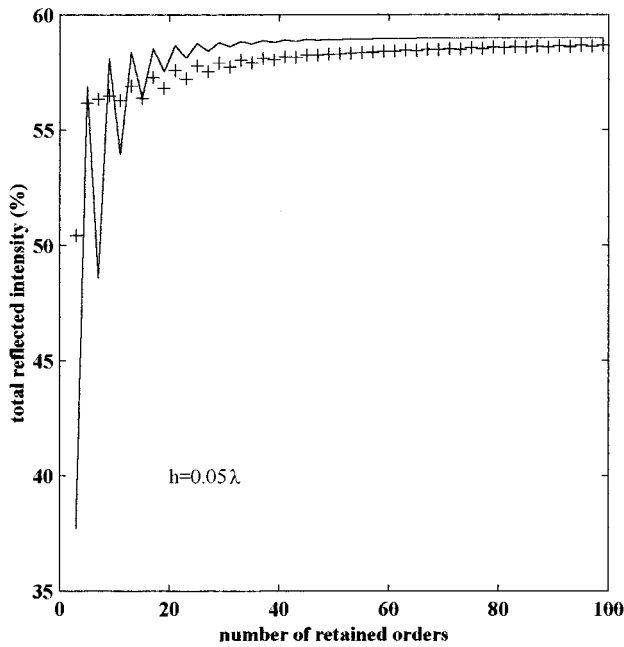
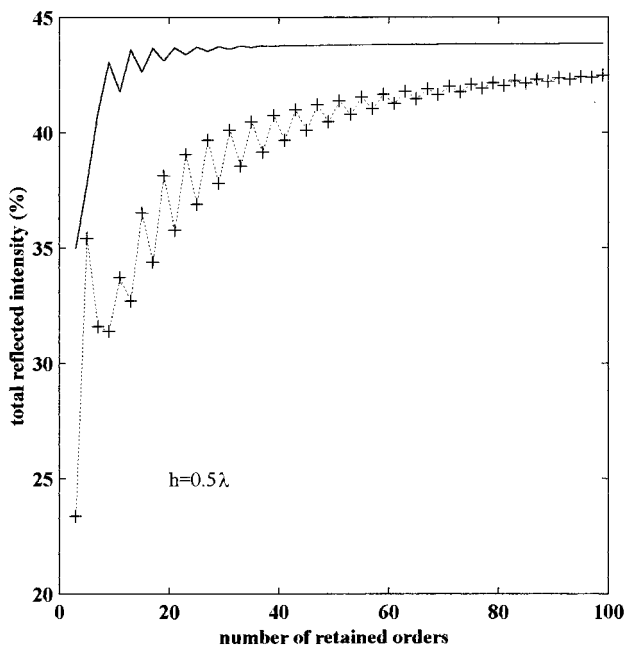Fig. 3.   Same as in Fig. 2, except that $h = 0.05\lambda$.

Fig. 4.   Same as in Fig. 2, except that $h = 0.5\lambda$.

creases.   In contrast, the conventional formulation provides a faster convergence rate; no oscillations are observed, and an accurate value of the total reflected intensity is obtained even with a small number of retained orders ($M > 10$).   It is concluded that the conventional formulation provides better performance than does the new one.   Consequently, the convergence rate is more sensitive to the bad conditioning of the new formulation than to the nonuniform boundary matching provided by the conventional formulation.   In Fig. 3 ($h = 0.05\lambda$), an intermediate situation is depicted.   Both formulations provide oscillations.   For $M < 20$, the am-

plitude of the oscillations is much stronger with the new formulation than with the conventional one.   For $20 < M < 40$, the amplitudes of the oscillations are approximately the same with the two formulations.   For $M > 50$, weak oscillations remain visible with the conventional formulation; no oscillations are observed with the new formulation.   Moreover, for $M > 60$, the total reflected intensity keeps on slowly increasing with the conventional formulation, whereas a plateau is observed with the new formulation.   The observation of a plateau in Fig. 3 indicates a better performance of the new formulation, even for a small depth value of 0.05λ.   For $h = 0.5\lambda$ (Fig. 4), the situation is clear; no small depth limit considerations can be applied, and the new formulation outperforms the conventional one.   The convergence-rate performances are similar to those observed in Refs. 1 and 2, in which this diffraction problem was analyzed for a depth equal to one wavelength.

In Figs. 2–4, the vertical scales are not the same, and a visual comparison cannot be directly operated.   When a small number of orders are retained for the computation, the error observed for large depth (Fig. 4) with the conventional formulation is much larger than the error observed for small depth (Fig. 2) with the new formulation. This result is not surprising because, as the grating depth is becoming smaller and smaller, the influence of the grating itself on the diffracted intensities vanishes.   For the purpose of illustration, let us consider the case of Fig. 2 ($h = 0.01\lambda$).   The reflected intensities of the zeroth and the minus-first orders are 61.0% and 0.2%, respectively. In the ultimate limit in which the grating thickness is null, no diffraction occurs.   The zeroth-order normalized electric-field amplitude R_0 is simply given by Fresnel formulas [this is the signification of Eqs. (12)], and the reflected intensity is 61.4%.   This value slightly differs from the zeroth-order reflected intensity obtained for $h = 0.01\lambda$ (deviation of 0.4%).   Thus if we consider that the diffraction problem of Fig. 2 can be roughly seen as resulting from a reflection at a planar air–metal interface, the oscillations observed with the new formulation for a small number of retained orders are surprisingly large.   In my opinion, this illustrates the effect of the bad conditioning on the convergence rate.

The situations shown in Figs. 2–4 clearly show that the grating depth has a significant effect on the convergence performance of the two formulations.   This confirms the simple prediction obtained in Section 2 for normal incidence.   In Section 4 a similar study is reported for non-lamellar gratings with multilevel binary and continuous profiles.

4.   NUMERICAL EXAMPLES: MULTILEVEL BINARY AND CONTINUOUS PROFILES

In this section we consider the diffraction problem depicted in Fig. 5.   The angle of incidence is $\theta = 30°$, the grating period is $\Lambda = \lambda$ ($\lambda = 1\ \mu m$), and the incident medium is air, $n_1 = 1$.   Two grating geometries are shown. In the first case (heavy lines in Fig. 5), the grating has a triangular profile.   The grooves are isosceles triangles with a continuously varying fill factor from 0 at the air–grating interface to 1 at the substrate–grating interface.
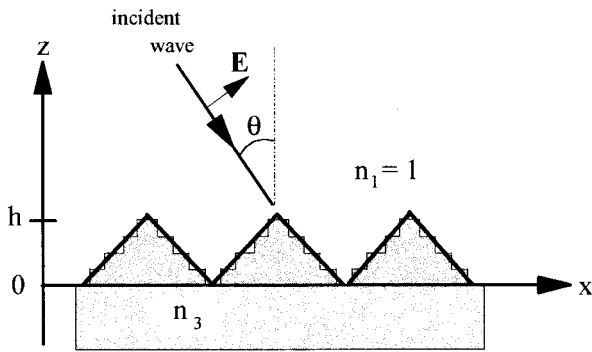
Fig. 5. Diffraction problems analyzed in Section 4. Two grating geometries are shown: the grating with a triangular profile (heavy lines), and its associated multilevel binary version for N = 4 grating layers. $n_1 = 1$ and $\theta = 30°$. h denotes the grating depth, and a TM-polarized light is considered.

Below, this grating geometry is referred to as the continuous grating profile geometry. For this geometry, the comparison between the convergence rates of the conventional and the new formulations is provided by use of the differential method.[9] For the numerical computation, the classical fourth-order Runge–Kutta algorithm is used, and for deep gratings the R-matrix propagation algorithm[10] is considered to eliminate numerical instabilities. We denote by N the number of slices. According to the terminology of Ref. 10, N also represents the stratification order. The second geometry shown in Fig. 5 corresponds to a multilevel binary grating. This geometry is related to the continuous grating profile geometry, in the sense that one obtains the fill factor of each grating layer by clipping the fill factor of the continuous grating profile. In Fig. 5, a four-layer grating geometry is depicted; each grating-layer thickness is $h/4$. Consistently with the continuous-profile case, the number of grating layers is denoted by N. For multilevel binary gratings, the numerical results reported below are obtained by use of the coupled-wave method. For deep gratings, the numerically stable enhanced transmittance approach[11] is used. For the two geometries, the grating is assumed to be etched into the substrate. Thus the relative permittivity in the grating region is either $n_1{}^2$ or $n_3{}^2$. Note that the following results are obtained without any modification of the R-matrix algorithm and of the enhanced transmittance approach. So the effect on the convergence-rate performance is due only to the use of the conventional or the new formulations for solving the differential equation or the eigenvalue problem in each slice or grating layer. Different numerical results are now reported.

Let us start with a simple diffraction problem for which nonconductive gratings are considered ($n_3 = 2.42$). Four transmitted and two reflected orders are nonevanescent. In Fig. 6 the zeroth-order transmitted intensity is plotted as a function of the number of retained orders for a small depth equal to one tenth of the wavelength. As above, the solid curve and the plus signs correspond to the new and the conventional formulations, respectively. This choice is consistent throughout the paper. Figure

6(a) is obtained with the coupled-wave method. For the computation, ten grating layers are considered. For M = 49, the zeroth-order transmitted intensities computed with the new and the conventional formulations have approximately identical values of 80.939% and 80.944%, respectively. For Fig. 6(b), the differential method is used with ten slices. Here again, the convergence is fast, and for 49 retained orders nearly identical diffraction efficiencies are obtained. The zeroth-order
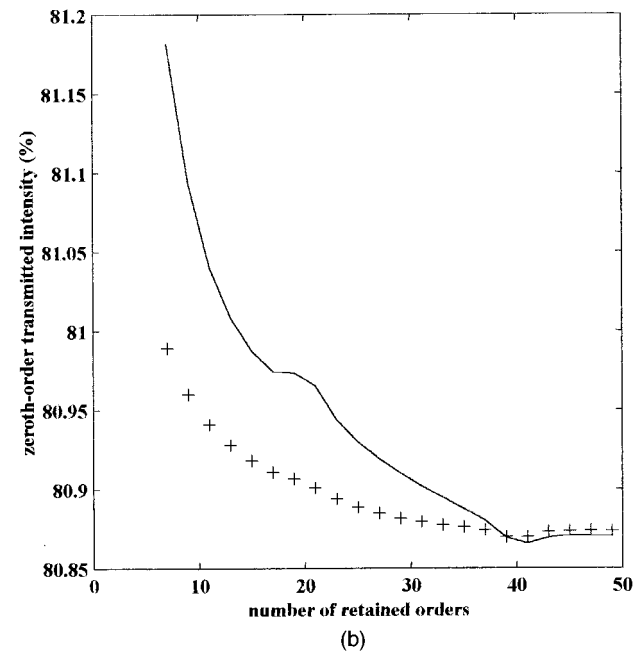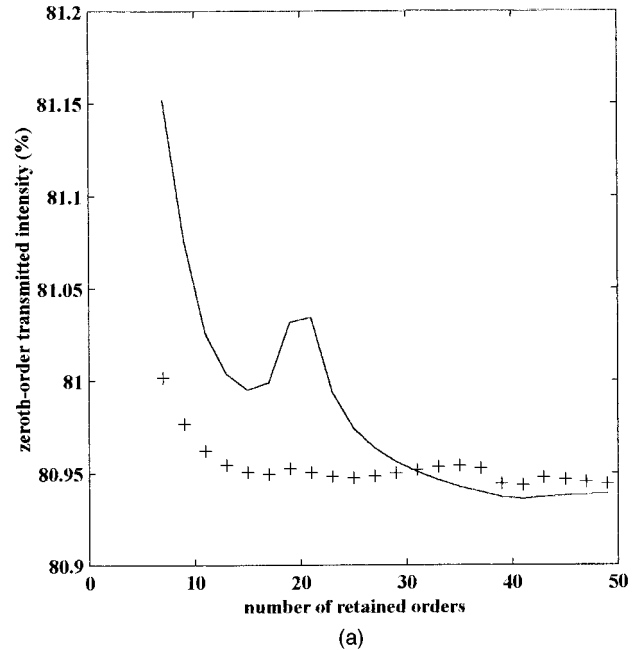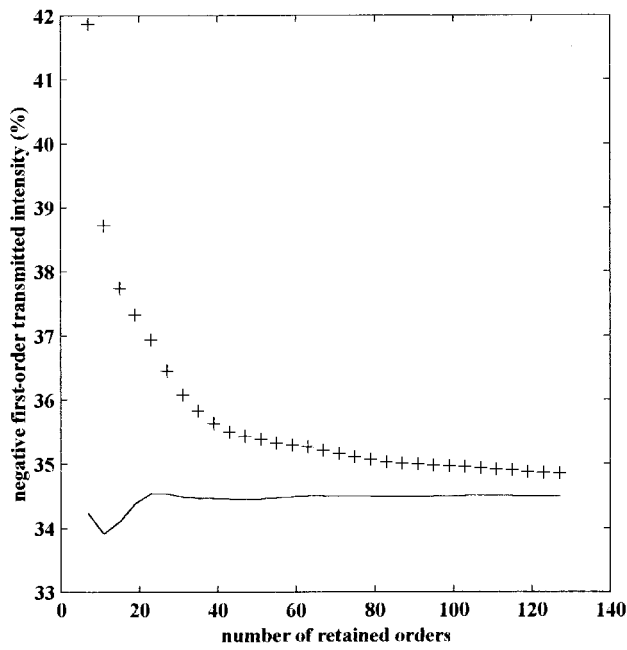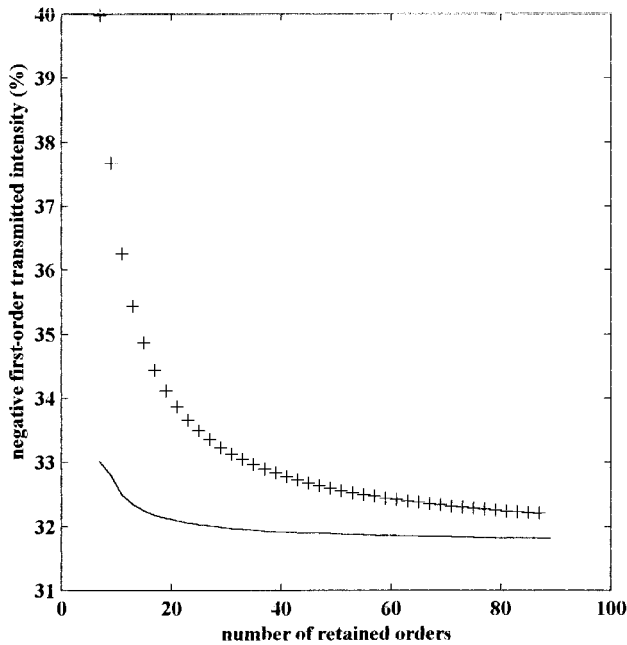

(a)

(b)

Fig. 6. Comparison between the new and the conventional formulations for the dielectric ($n_3 = 2.42$) gratings shown in Fig. 5 and for $h = 0.1\lambda$. The solid curves and the plus signs correspond to the new and the conventional formulations, respectively. (a) Results obtained with the coupled-wave method and $N = 10$, (b) results obtained with the differential method and $N = 10$.

(a)

(b)

Fig. 7.   Same as in Fig. 6, except that $h = \lambda$ and $N = 100$ slices are used with the differential method.

transmitted intensities computed with the new and the conventional formulations are 80.870% and 80.873%, respectively.   These are values close to those obtained with the coupled-wave method.   For this simple diffraction problem in which a dielectric grating is considered, the two formulations perform well, and convergence rates are fast.   However, from Fig. 6 it can be concluded that the conventional formulation provides the fastest convergence rate.   This conclusion holds for both the differential and the coupled-wave methods.   We now consider the

same dielectric grating ($n_3 = 2.42$), but with a depth equal to one wavelength.   Figures 7(a) and 7(b) show the negative first-order transmitted intensities computed with the two formulations, respectively, as a function of the number of retained orders.   The results obtained with the coupled-wave method are shown in Fig. 7(a).   Ten grating layers ($N = 10$) are used for the computation.   Clearly, the new formulation outperforms the conventional one.   As shown in Fig. 7(b), in which 100 slices are considered to approximate the continuous profile, a similar conclusion holds for the differential method.   The opposite performances obtained in Figs. 6 and 7 show that the grating depth has a significant effect on the convergence rates of the two formulations, not only for lamellar and metallic gratings as was shown in Section 3, but also for dielectric gratings with multilevel binary and continuous-profile geometries.

In Fig. 7, although the fastest convergence rates are obtained with the new formulation, the effect is rather small since both formulations perform well for dielectric substrates.   This is why we now consider a highly conductive grating situation for which the refractive index n_3 is equal to $3.18 + j4.41$, as in Section 3.   Only the minus-first and zeroth orders are propagating.   The depth h is equal to 0.1λ.   Figure 8(a) is obtained with the coupled-wave method, and $N = 20$ grating layers are considered for the computation.   For 179 retained orders, the total reflected intensities computed with the conventional and the new formulations are equal to 60.357% and 60.391%, respectively.   If we consider the average value 60.37% as the exact value, we note from Fig. 8(a) that it is sufficient to retain 25 orders with the conventional formulation to compute an accurate value of the reflected intensity (the deviation does not exceed 0.3% for $25 < M < 179$).   With the new formulation, more than 100 orders are necessary to obtain the same accuracy.   In Fig. 8(b), results obtained with the differential method are shown.   Twenty slices are used in the computation.   For 179 retained orders, the total reflected intensities computed with the conventional and the new formulations are 60.592% and 60.446%, respectively.   These are results consistent with those found with the coupled-wave method.   Moreover, the convergence performances shown in Fig. 8(b) are similar to those observed in Fig. 8(a).   However, we note that in neither Fig. 8(a) nor Fig. 8(b) is a plateau observed with the conventional formulation; in Fig. 8(a) a weakly wavy behavior of the total reflected intensity remains visible, and in Fig. 8(b) a small bumpy variation is observed for approximately 160 retained orders.   For thin gratings it is concluded that, although the convergence rate of the conventional formulation is faster than that of the new one, it is not as fast as that obtained with the new formulation for deep gratings.   In my opinion, this result reflects the conflicting situation that one encounters when using the conventional formulation to modelize thin gratings: The formulation is well conditioned in the small depth limit but does not uniformly satisfy the boundary conditions inside the grating region.   The conflicting situation results in the observation of an intermediate convergence rate.
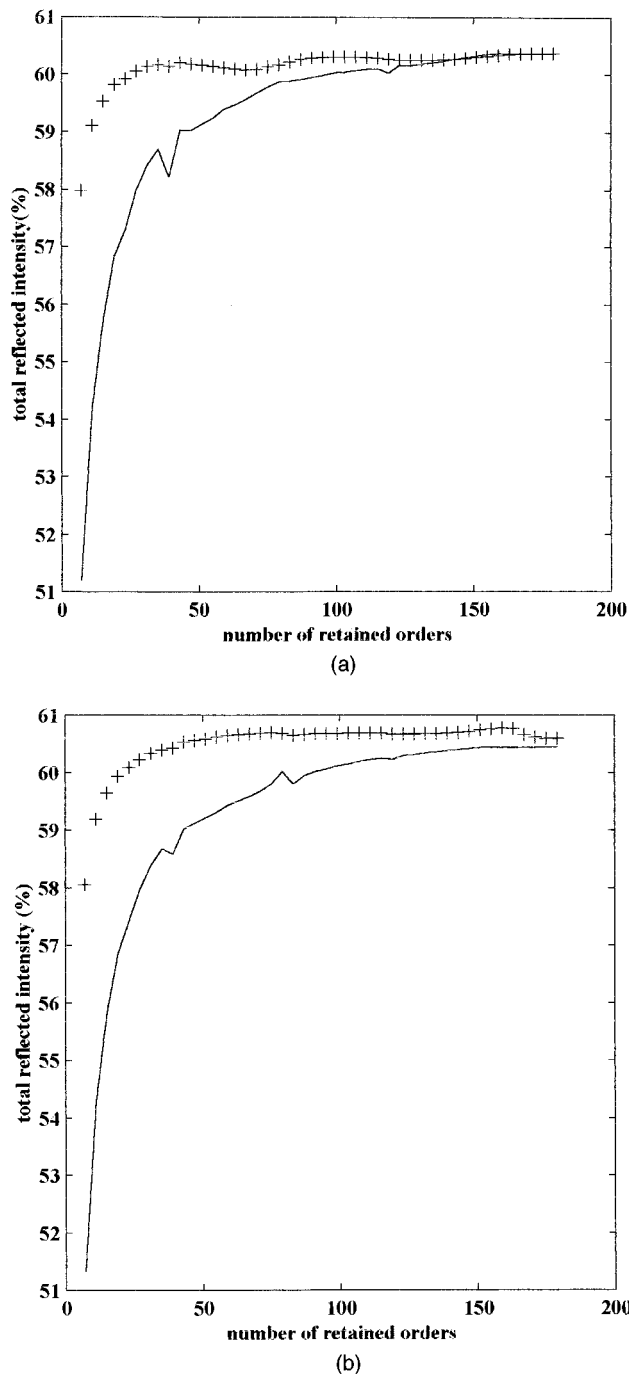
(a)

(b)

Fig. 8. Total reflected intensity as a function of the number of retained orders for the metallic ($n_3 = 3.18 + j4.41$) gratings shown in Fig. 5 and for $h = 0.1\lambda$. The solid curves and the plus signs correspond to the new and the conventional formulations, respectively. (a) Results obtained with the coupled-wave method and $N = 20$, (b) results obtained with the differential method and $N = 20$.

5. DISCUSSION AND CONCLUSIONS

In this paper, we have used an analytical derivation to show that the new formulation of the coupled-wave method recently proposed by several authors is badly conditioned in the small depth limit for TM polarization and

for normal incidence.   Numerous computational demonstrations provided by the differential and the coupled-wave methods comprise evidence that the bad conditioning has a negative effect on the convergence performance of the new formulation.   This degradation was observed in many different situations, including dielectric and metallic gratings, with lamellar, multilevel binary, and continuous-profile geometries.   It was even shown that, for small grating thicknesses, the conventional formulation outperforms the new one.   This surprising result has interesting practical consequences, since the use of the conventional formulation rather than the new one can be recommended for thin gratings.   However, it is noteworthy that, when the conventional formulation is used for thin gratings, boundary conditions are not uniformly matched in the grating region.   In fact, the study of thin gratings by computational methods relying on a Fourier expansion of the permittivity poses a conflicting situation: Neither the new nor the conventional formulation provides a fully satisfying modelization.   For instance, if one is interested in visualizing the electromagnetic field inside the grating region, the new formulation has to be preferred.   However, if one is interested in computing the diffracted intensities, the conventional formulation is instead recommended.   This distinction is particularly striking for metallic gratings, and it is due to the fact that the new formulation is badly conditioned in the small depth limit.   In fact, the situation is a little more complex, and two convergence regimes that reflect the conflicting situation can be distinguished.   When a moderate number of orders are retained in the computation, the convergence rate is imposed by the conditioning in the small depth limit rather than by the (uniform or nonuniform) matching of the boundary conditions.   When a large number of orders are retained in the computation, reverse conclusions hold, but the effect is rather small, since for metallic gratings it only affects the third or the fourth digit of the computed diffracted intensities.

One could be interested in the following question: Which characteristic value of the depth-to-wavelength ratio should determine the choice of using one formulation rather than the other?   Answering this question is a difficult task.   The answer probably depends on the grating profile and on the grating parameters and certainly depends on the accuracy required for the computed diffraction efficiencies.   An intuitive insight into the answer can be gained from the work reported in Ref. 12, where, by use of a derivation similar to that given in Section 2, it is shown that the effective index of subwavelength gratings is strongly dependent on the grating thickness for TM polarization.   From the examples of Ref. 12 and from those provided in this paper, it can be concluded that the characteristic value is approximately 0.1 and probably smaller for lamellar gratings.   However, considering complex situations such as that shown in Fig. 3, where the use of the formulation with the strongest oscillations can be recommended, it is preferable to perform empirical tests rather than to apply an approximate rule.

For TE polarization (electric-field vector perpendicular to the grating vector), the formulation given in Ref. 3 is

well conditioned in the quasi-static limit[1] and uniformly satisfies[5] the boundary conditions inside the grating region. Moreover, by use of a derivation similar to that given in Section 2, it can be shown that this formulation is also well conditioned in the small depth limit. So no problems occur for TE polarization. For conical mountings of one-dimensional gratings, the new formulation given in Ref. 1 is badly conditioned in the small depth limit; difficulties comparable with those encountered in this paper for TM polarization can be expected. For two-dimensional gratings, the differential equations used for conical mountings of one-dimensional gratings can be used with minor notation changes.[3] So the same terminology (conventional and new) as that for conical mountings[1] of one-dimensional gratings can be employed for two-dimensional gratings without any confusion. The conventional formulation was recently revisited by several authors,[3,13,14] but no change in the eigenvalue problem formulation was reported. From the work on the effective-medium theory of two-dimensional gratings reported in Ref. 15, it is clear that both the new and the conventional formulations are badly conditioned in the quasi-static limit. Moreover, from Li's work[5] it is easily shown that both formulations do not uniformly satisfy the boundary conditions inside the grating region. However, it is noteworthy that the conventional formulation is well conditioned in the small depth limit.

The fact that both the conventional and the new formulations do not provide a fully satisfying modelization of gratings with arbitrary thicknesses and profiles is clearly a weak point of the coupled-wave and the differential methods. It is not my intention to emphasize such a pessimistic conclusion. Rather, I expect that this work will contribute to the understanding of these two well-known methods and will make using them even easier. Also, I expect that the physical insights gained from the studies of the quasi-static and small depth limits for the coupled-wave method can successfully be applied to other rigorous computation methods to improve their performance.

ACKNOWLEDGMENTS

REFERENCES AND NOTES

1. Ph. Lalanne and G. M. Morris, "Highly improved convergence of the coupled-wave method for TM polarization," J. Opt. Soc. Am. A **13**, 779–784 (1996).
2. G. Granet and B. Guizal, "Efficient implementation of the coupled-wave method for metallic lamellar gratings in TM polarization," J. Opt. Soc. Am. A **13**, 1019–1023 (1996).
3. M. G. Moharam, E. B. Grann, D. A. Pommet, and T. K. Gaylord, "Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings," J. Opt. Soc. Am. A **12**, 1068–1076 (1995).
4. J. Turunen, "Form-birefringence limits of Fourier-expansion methods in grating theory," J. Opt. Soc. Am. A **13**, 1013–1018 (1996).
5. L. Li, "Use of Fourier series in the analysis of discontinuous periodic structures," J. Opt. Soc. Am. A **13**, 1870–1876 (1996).
6. R. Petit and G. Tayeb, "Theoretical and numerical study of gratings consisting of periodic arrays of thin and lossy strips," J. Opt. Soc. Am. A **7**, 1686–1692 (1990).
7. When lamellar gratings are considered, in general, matrices $\mathbf{E}$ and $\mathbf{A}^{-1}$ are different. However, $a_{0,0}^{(-1)}$ is equal to ε_0. For an enlightening discussion, see Ref. 5.
8. L. Li and C. W. Haggans, "Convergence of the coupled-wave method for metallic lamellar diffraction gratings," J. Opt. Soc. Am. A **10**, 1184–1189 (1993).
9. M. Nevière, P. Vincent, and R. Petit, "Sur la théorie du réseau conducteur et ses applications à l'optique," Nouv. Rev. Opt. **5**, 65–77 (1974).
10. F. Montiel and M. Nevière, "Differential theory of gratings: extension to deep gratings of arbitrary profile and permittivity through the R-matrix propagation algorithm," J. Opt. Soc. Am. A **11**, 3241–3250 (1994).
11. M. G. Moharam, D. A. Pommet, E. B. Grann, and T. K. Gaylord, "Stable implementation of the rigorous coupled-wave analysis for surface-relief gratings: enhanced transmittance matrix approach," J. Opt. Soc. Am. A **12**, 1077–1086 (1995).
12. Ph. Lalanne and D. Lemercier-Lalanne, "Depth dependence of the effective properties of subwavelength gratings," J. Opt. Soc. Am. A **14**, 450–458 (1997).
13. E. Noponen and J. Turunen, "Eigenmode method for electromagnetic synthesis of diffractive elements with three-dimensional profiles," J. Opt. Soc. Am. A **11**, 2494–2502 (1994).
14. S. Peng and G. M. Morris, "An efficient implementation of rigorous coupled-wave analysis for surface-relief gratings," J. Opt. Soc. Am. A **12**, 1087–1096 (1995).
15. Ph. Lalanne and D. Lemercier-Lalanne, "On the effective medium theory of subwavelength periodic structures," J. Mod. Opt. **43**, 2063–2085 (1996).