Progress In Electromagnetics Research, Vol. 133, 17–35, 2013

MODAL METHOD BASED ON SUBSECTIONAL GEGENBAUER POLYNOMIAL EXPANSION FOR LAMELLAR GRATINGS: WEIGHTING FUNCTION, CONVERGENCE AND STABILITY

K. Edee[1, *], I. Fenniche[1], G. Granet[1], and B. Guizal[2]

[1]Université Blaise Pascal, LASMEA, UMR-6602-CNRS, BP 10448, F-63000 Clermont-Ferrand, France

[2]Laboratoire Charles Coulomb, UMR 5221 of the CNRS, University of Montpellier 2, France

Abstract—The Modal Method by Gegenbauer polynomials Expansion (MMGE) has been recently introduced for lamellar gratings by Edee [8]. This method shows a promising potential of outstanding convergence but still suffers from instabilities when the number of polynomials is increased. In this work, we identify the origin of these instabilities and propose a way to remove them.

1. INTRODUCTION

Among the numerical methods developed for the analysis of lamellar diffraction gratings, modal methods play an important role because of their great versatility and relative effectiveness. In the classical modal method [1, 2], the eigenvalues are obtained by solving a transcendental equation. In other modal methods, the eigenmodes and propagation constants are generally obtained by searching the eigenvalues and eigenvectors [3–8] of a matrix which is derived from the Maxwell's equations by using the method of moments [9]. Mathematically speaking, for one-dimensional gratings with piecewise homogeneous media, and plane wave excitation, the eigenmodes are solutions of the Helmholtz equation subject to boundary conditions at the interfaces between two media and to the pseudo-periodicity condition. Numerically, the rate of convergence of the method depends on how the matrix from which eigenvalues are sought takes into account the continuity relations. Indeed, one of the main differences between

Received 13 June 2012, Accepted 20 July 2012, Scheduled 16 October 2012

* Corresponding author: M. Kofi Edee (kof_8@hotmail.fr).

the variants of modal methods is the choice of the expansion and test functions. In a previous paper [10], we have developed a modal method based on Gegenbaueur polynomials expansions (MMGE) and emphasized that the main advantage of such an approach is that continuity relations can be written in an exact manner. We have shown through various examples that this method outperforms other modal methods. However, we found that under its original form, the MMGE suffers from some instabilities for large values of polynomials degree. Even if it is not, for a certain class of grating problems, necessary to use a large number of polynomials in order to have very reliable results (because the MMGE converges very rapidly), it is of fundamental importance to identify the origin of instabilities and find a way to remove them. In [10], it has already been highlighted that the instabilities were linked with the way we calculated the inner product required by the Galerkin method. In the present work, we track precisely the origin of the numerical problems and study the influence of the weighting function appearing in the inner products. It is shown that introducing this latter makes the calculation of inner products analytical in one hand and ensures unconditional stability on the other hand; on the contrary to our first implementation. The Gegenbauer polynomials of degree m denoted by C_m^Λ differ from each other through a parameter Λ. Special cases where Λ is equal to 0.5 corresponds to the Legendre polynomial and as Λ approaches 0, these polynomials are Chebychev polynomials. In addition, we investigate the influence of Λ on the rate of convergence for the two fundamental cases of polarization.

2. STATEMENT OF THE PROBLEM AND THE FRAMEWORK OF MMGE

The MMGE consists in defining a partition Ω_i corresponding to the different homogeneous subintervals of Ω (which is nothing but the elementary period of the structure), that can be, possibly, divided into layers Ω_{ij}, $1 \leq j \leq N_i$. For example, in Figure 1, Ω is subdivided into two homogeneous subintervals Ω_1 and Ω_2 and each subinterval $\Omega_i, (i = 1, 2)$ contains N_i layers.

We will denote by N_Ω the number of subintervals. The eigenfunctions $X_p(x)$ of the $\mathcal{L}$-operator ($\mathcal{L}_{TE} = k^{-2}\partial_x^2 + \nu^2(x)$ and $\mathcal{L}_{TM} = k^{-2}\nu^2(x)\partial_x\nu^{-2}(x)\partial_x + \nu^2(x)$ with $k = 2\pi/\lambda$) are described in each homogeneous layer Ω_{ij} as follows:

$$\left|X_p^{i,j}\right\rangle = \sum_{n=1}^{N} a_{n,p}^{i,j}\left|b_n^{i,j}\right\rangle, \tag{1}$$

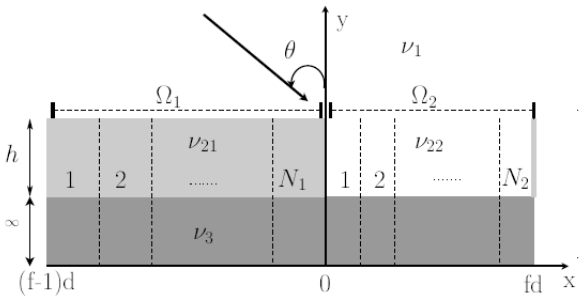

Figure 1. The grating configuration: one period is depicted.

where $|X_p^{i,j}\rangle$ is the restriction of the eigenfunction $|X_p\rangle$ to the homogeneous layer Ω_{ij}, characterized by its refractive index ν_i and N is the number of basis functions on each homogeneous layer. It is of fundamental importance to note that $|X_p^{i,j}\rangle$ satisfies:

(i) the Helmholtz equation identical for both TE and TM polarizations:

$$\mathcal{L}^{i,j}\big|X_p^{i,j}\big\rangle = \beta_p^2\big|X_p^{i,j}\big\rangle, \tag{2}$$

with

$$\mathcal{L}^{i,j} = \frac{1}{k^2}\frac{d^2}{dx^2} + \nu_i^2. \tag{3}$$

(ii) the following boundary conditions:

- for each layer Ω_{ij} of the same subinterval Ω_i, $(i,j) \in [1 \ : \ N_\Omega - 1] \times [1 \ : \ N_i]$, the boundary equations obtained by writing the continuity of the tangential components of the electromagnetic field at the interfaces $(x_{i,j})_{j\in[1\ :\ N_i-1]}$ separating two adjacent layers, Ω_{ij} and $\Omega_{i,j+1}$:

$$X_p^{i,j}(x_{i,j}) = X_p^{i,j+1}(x_{i,j}), \tag{4a}$$

$$\left[\frac{dX_p^{i,j}}{dx}\right]_{x=x_{i,j}} = \left[\frac{dX_p^{i,j+1}}{dx}\right]_{x=x_{i,j}}, \tag{4b}$$

- at the interfaces separating two adjacent subintervals Ω_i and Ω_{i+1}, i.e., for $j = N_i$, $i \in [1 \ : \ N_\Omega - 1]$,

$$X_p^{i,j}(x_{i,j}) = X_p^{i+1,1}(x_{i,j}), \tag{5a}$$

$$\frac{1}{\eta^i}\left[\frac{dX_p^{i,j}}{dx}\right]_{x=x_{i,j}} = \frac{1}{\eta^{i+1}}\left[\frac{dX_p^{i+1,1}}{dx}\right]_{x=x_{i,j}}, \tag{5b}$$

where $\eta^i = 1$ for TE polarization and $\eta^i = \nu_i^2$ for TM polarization.

- for the last subinterval Ω_i, i.e., when $i = N_\Omega$, Eq. (4) are written for the first layers $(\Omega_{ij})_{ij}$, $(i,j) \in \{N_\Omega\} \times [1 \,:\, N_i - 1]$ and the pseudo-periodic condition for $j = N_i$:

$$X_p^{i,N_i}(d) = e^{ik\sin\theta d} X_p^{1,1}(0), \tag{6a}$$

$$\frac{1}{\eta^i}\left[\frac{dX_p^{i,N_i}}{dx}\right]_{x=d} = \frac{e^{ik\sin\theta d}}{\eta^1}\left[\frac{dX_p^{1,1}}{dx}\right]_{x=0}, (i = N_\Omega). \tag{6b}$$

From Eq. (1), Eq. (2) may be written as follows:

$$\sum_{n=1}^{N} a_{n,p}^{i,j} \mathcal{L}^{i,j} \left|b_n^{i,j}\right\rangle = \beta_p^2 \sum_{n=1}^{N} a_{n,p}^{i,j} \left|b_n^{i,j}\right\rangle. \tag{7}$$

Finally, by projecting Eq. (7) on the basis functions $\left(|b_m^{i,j}\rangle\right)_{m\in[1\,:\,N-2]}$, we obtain:

$$\mathbf{L}_{[N-2]\times[N]}^{i,j} \mathbf{a}_{[1\,:\,N],p}^{i,j} = \beta_p^2 \mathbf{G}_{[N-2]\times[N]}^{i,j} \mathbf{a}_{[1\,:\,N],p}^{i,j}, \tag{8}$$

where

$$\mathbf{L}_{[N-2]\times[N]}^{i,j} = \frac{1}{k^2}\mathbf{D}_{[N-2]\times[N-1]}^{i,j}\left[\mathbf{G}_{[N-1]\times[N-1]}^{i,j}\right]^{-1}\mathbf{D}_{[N-1]\times[N]}^{i,j} + \nu_i^2 \mathbf{G}_{[N-2]\times[N]}^{i,j}, \tag{9}$$

$\mathbf{a}_{[1\,:\,N],p}^{i,j}$ is a column vector formed by the coefficients $a_{n,p}^{i,j}$, $n \in [1 \,:\, N]$:

$$\mathbf{a}_{[1\,:\,N],p}^{i,j} = \left[a_{1,p}^{i,j}, \ldots, a_{N,p}^{i,j}\right]^t, \tag{10}$$

and

$$\mathbf{G}_{[M]\times[Q]}^{i,j} = \left[\langle b_m^{i,j}, b_q^{i,j}\rangle\right], \tag{11a}$$

$$\mathbf{D}_{[M]\times[Q]}^{i,j} = \left[\langle b_m^{i,j}, \frac{db_q^{i,j}}{dx}\rangle\right]. \tag{11b}$$

The subscripts of the matrices denote their size; for example in Eq. (11), $(m,q) \in [1 \,:\, M] \times [1 \,:\, Q]$. We are, thus, led to the computation of the eigenvalues β_p^2 and their associated eigenvectors $\mathbf{a}_{[1\,:\,N-2],p}^{i,j}$ of a matrix with dimension $N_{\max} \times N_{\max}$, with

$$N_{\max} = (N-2)\sum_{i=1}^{N_\Omega} N_i. \tag{12}$$

We chose the basis functions formed by the Gegenbauer polynomials [11] $C_m^{\Lambda}(\xi)$ defined over the interval $[-1, 1]$ as follows:

$$C_m^{\Lambda}(\xi) = \frac{1}{\Gamma(\Lambda)} \sum_{q=0}^{[m/2]} (-1)^q \frac{\Gamma(\Lambda + m - q)}{(q+1)!(1+m-2q)!} (2\xi)^{m-2q}, \tag{13}$$

where $\Lambda > -1/2$ and m denoted the degree of the polynomials. The Gegenbauer polynomials C_m^{Λ} are m degree orthogonal polynomials on the interval $[-1, 1]$ satisfying:

$$\left\langle C_m^{\Lambda}, C_n^{\Lambda} \right\rangle = \int_{-1}^{1} \left(1 - \xi^2\right)^{\Lambda - \frac{1}{2}} C_m^{\Lambda}(\xi) C_n^{\Lambda}(\xi) d\xi = \delta_{nm} h_n^{\Lambda}, \tag{14}$$

where δ_{nm} denotes the Kronecker's symbol and

$$h_n^{\Lambda} = \pi^{\frac{1}{2}} C_n^{\Lambda}(1) \frac{\Gamma(\Lambda + \frac{1}{2})}{\Gamma(\Lambda)(n + \Lambda)}, \tag{15}$$

with

$$\begin{cases} C_n^{\Lambda}(1) = \dfrac{\Gamma(n + 2\Lambda)}{\Gamma(2\Lambda)\Gamma(n+1)}, \\ C_n^{\Lambda}(-1) = (-1)^n C_n^{\Lambda}(1). \end{cases} \tag{16}$$

These relationships will be essential for the boundary conditions associated with the transition points in the x direction. It is important to note that the parameter N is referred to the number of basis functions over each layer. Consequently, the highest degree of these polynomials is $N - 1$.

3. INNER PRODUCT CALCULATION

3.1. Computation without the Weighting Function: Numerical Instabilities and Gamma Function

In a first attempt and for sake of lightening the computations, the inner product described by Eq. (14) can be simplified by removing the weighting function $(1 - \xi^2)^{\Lambda - 1/2}$. This is the approach adopted in [10] and which led to very good convergence rates. However, and as we mentioned in the introduction, numerical instabilities arise when the number of polynomials is increased. Obviously, the first idea that comes to mind is that the origin of the instabilities might have something to do with the removed weighting function. In general, the instabilities of an algorithm based on a modal method may come either from: (i) the fact that the modal decomposition is not able to represent the actual fields (especially their discontinuities), (ii) the

numerical computation of the modes and especially at the level of the inner products computations, and (iii) the resolution of the algebraic system stemming from the boundary conditions. In the present case, we suspect the accuracy of evaluation of the inner products given in Eq. (11) that use the inner product of Eq. (14) (under its simplified form, i.e., by removing the weighting function). In order to clarify the situation, we return to the construction of the matrix $\mathbf{G}$ which elements are $G_{mn} =< C_m^\Lambda | C_n^\Lambda >$. Let's consider as an example the case of $\Lambda = 0.5$, where the elements G_{mn} computed by convolving and integrating the polynomials must be perfectly equal to the inner product defined by Eq. (14); i.e., if the polynomials are normalized by $\sqrt{h_n^\Lambda}$, G_{mn} must be equal to δ_{mn}. In Eq. (17) we give the numerical computation of $\mathbf{G}$ elements for $(m, n) \in [15 \ : \ 17] \times [15 \ : \ 17]$:

$$\mathbf{G}_{[15 \ : \ 17],[15 \ : \ 17]} = \begin{bmatrix} 1.0000 & 0 & 0.0000 \\ 0 & 1.0000 & 0 \\ 0.0000 & 0 & 1.0000 \end{bmatrix}. \tag{17}$$

It can be seen that the results are satisfactory for this range of integers. Nevertheless, when m or n increases, the results become highly unstable as is shown in matrix (18):

$$\mathbf{G}_{[29 \ : \ 31][29 \ : \ 31]} = 10^4 \begin{bmatrix} -0.0193 & 0 & 1.4470 \\ 0 & 0.1740 & 0 \\ 0.8932 & 0 & -6.4372 \end{bmatrix}. \tag{18}$$

This behavior suggests that the instabilities come from the manipulation of the coefficients of Gegenbauer polynomials. We verified and confirmed this fact through the numerical calculation of $C_n^\Lambda(1)$ by use of the expression (13). The numerical evaluation of the sum in Eq. (13) can be diverging. Indeed, this expression contains Gamma functions which numerical expression leads to very large values. The numerical calculation of the ratio between the numerator and denominator appearing in the expression of the coefficients of monomials (13) rapidly tends to infinity as a function of q and m, while in reality the fraction tends to a finite value. The same behavior is observed for $C_m^\Lambda(-1)$, $(dC_m^\Lambda/d\xi)_{\xi=-1,1}$, i.e., all values of $C_m^\Lambda(\xi)$ and $(dC_m^\Lambda/d\xi)_\xi$, which are essential for boundary conditions. One alternative to solve this problem, which was proposed in [10], consists in increasing the number of homogeneous layers Ω_{ij} for each subinterval Ω_i. Indeed, by doing so the number of Gegenbauer polynomials needed for the field description on each layer decreases. Another alternative, which is presented in Subsection 3.2, consists in analytically computing all the terms needed for the matrix of diffraction. This will be done by taking into account the weighting function in the inner products and by examining these ones case by case.

3.2. Analytical Computation of the Inner Products with the Weighting Function

The construction of the matrix of diffraction entailed the calculation of the terms:

- $\langle C_m^\Lambda, C_n^\Lambda \rangle = \int_{-1}^{1} (1-\xi^2)^{\Lambda-\frac{1}{2}} C_m^\Lambda(\xi) C_n^\Lambda(\xi) d\xi$,
- $\langle C_m^\Lambda, \frac{dC_n^\Lambda}{d\xi} \rangle = \int_{-1}^{1} (1-\xi^2)^{\Lambda-\frac{1}{2}} C_m^\Lambda(\xi) \left(\frac{dC_n^\Lambda}{d\xi}\right)(\xi) d\xi$.

The computation of $\langle C_m^\Lambda, C_n^\Lambda \rangle$ is simple and directly deduced, in closed form, from Eq. (14). For terms $\langle C_m^\Lambda, \frac{dC_n^\Lambda}{d\xi} \rangle$, we first treat the terms for $(m, n = 0)$ and $(m, n = 1)$ before dealing with the general case. It is easy to verify that $\langle C_m^\Lambda, \frac{d}{d\xi} C_0^\Lambda \rangle$ vanish for all m and, since, $\frac{dC_1^\Lambda}{d\xi}$ is a constant, we can write:

$$\left\langle C_m^\Lambda, \frac{dC_1^\Lambda}{d\xi} \right\rangle = \frac{dC_1^\Lambda}{d\xi} \int_{-1}^{1} \left(1-\xi^2\right)^{\Lambda-\frac{1}{2}} C_m^\Lambda(\xi) d\xi. \tag{19}$$

In the particular case of $m = 0$, the calculation of terms $\langle C_0^\Lambda, \frac{dC_1^\Lambda}{d\xi} \rangle$, leads to the following relationship:

$$\left\langle C_0^\Lambda, \frac{dC_1^\Lambda}{d\xi} \right\rangle = C_0^\Lambda \frac{dC_1^\Lambda}{d\xi} \int_0^\pi \sin^{2\Lambda}\theta d\theta. \tag{20}$$

The integral of the right-hand of Eq. (20) is computed by using the following expression, which involves Bessel and Gamma functions [11]:

$$\frac{J_\Lambda(\xi)}{\left(\frac{\xi}{2}\right)^\Lambda} = \frac{1}{\pi^{\frac{1}{2}}\Gamma(\Lambda+\frac{1}{2})} \int_0^\pi \cos(\xi\cos\theta)\sin^{2\Lambda}\theta d\theta. \tag{21}$$

For non-negative values of Λ and for small arguments ξ ($\xi \in \left[0 : \sqrt{\Lambda+1}\right]$), Bessel functions have the following asymptotic form:

$$\frac{J_\Lambda(\xi)}{\left(\frac{\xi}{2}\right)^\Lambda} \simeq \frac{1}{\Gamma(\Lambda+1)}. \tag{22}$$

By combining Eq. (22) and Eq. (21), when ξ is closed to 0, we obtain:

$$\int_0^\pi \sin^{2\Lambda}\theta d\theta \simeq \pi^{\frac{1}{2}} \frac{\Gamma(\Lambda+\frac{1}{2})}{\Gamma(\Lambda+1)}. \tag{23}$$

Thus $\langle C_0^\Lambda, \frac{dC_1^\Lambda}{d\xi}\rangle$ has the following analytical expression:

$$\left\langle C_0^\Lambda, \frac{dC_1^\Lambda}{d\xi} \right\rangle \simeq \pi^{\frac{1}{2}} C_0^\Lambda \frac{dC_1^\Lambda}{d\xi} \frac{\Gamma(\Lambda+\frac{1}{2})}{\Gamma(\Lambda+1)}. \tag{24}$$

For $m \geq 1$, by using the integral Eq. (25) which involves Gegenbauer polynomials:

$$\frac{m(2\Lambda+m)}{2\Lambda}\int_0^\xi \left(1-y^2\right)^{\Lambda-\frac{1}{2}} C_m^\Lambda(y)dy = C_{m-1}^{\Lambda+1}(0) - \left(1-\xi^2\right)^{\Lambda+\frac{1}{2}} C_{m-1}^{\Lambda+1}(\xi), \tag{25}$$

we easily demonstrate that $\langle C_m^\Lambda, \frac{dC_1^\Lambda}{d\xi}\rangle = 0$, for all values of $m \geq 1$.

At this stage, elements $\langle C_m^\Lambda, \frac{dC_0^\Lambda}{d\xi}\rangle$ and $\langle C_m^\Lambda, \frac{dC_1^\Lambda}{d\xi}\rangle$, are known for all values of m. In order to calculate terms $\langle C_m^\Lambda, \frac{dC_n^\Lambda}{d\xi}\rangle$, when $n \geq 2$ and for all values of m, we introduce the following recursive relation:

$$C_n^\Lambda(\xi) = \frac{1}{2(n+\Lambda)} \frac{d}{d\xi} \left[C_{n+1}^\Lambda(\xi) - C_{n-1}^\Lambda(\xi)\right], \tag{26}$$

which leads to

$$\frac{d}{d\xi} C_n^\Lambda(\xi) = 2(n-1+\Lambda) C_{n-1}^\Lambda(\xi) + \frac{d}{d\xi} C_{n-2}^\Lambda(\xi). \tag{27}$$

Consequently, terms $\langle C_m^\Lambda, \frac{d}{d\xi} C_n^\Lambda\rangle$ are obtained as:

$$\left\langle C_m^\Lambda, \frac{dC_n^\Lambda}{d\xi} \right\rangle = 2(n-1+\Lambda) \left\langle C_m^\Lambda, C_{n-1}^\Lambda \right\rangle + \left\langle C_m^\Lambda, \frac{dC_{n-2}^\Lambda}{d\xi} \right\rangle. \tag{28}$$

Finally, we will need to use the following formula

$$\left(\frac{dC_n^\Lambda}{d\xi}\right)(\xi) = 2\Lambda C_{n-1}^{\Lambda+1}(\xi), \tag{29}$$

which is essential for the boundary conditions;, i.e., to express the continuity of the field derivative at the interfaces $\xi = 1$ and $\xi = -1$. From Eq. (29), we obtain:

$$\begin{cases} \left(\frac{dC_n^\Lambda}{d\xi}\right)(\xi = 1) = 2\Lambda C_{n-1}^{\Lambda+1}(1), \\ \left(\frac{dC_n^\Lambda}{d\xi}\right)(\xi = -1) = 2\Lambda C_{n-1}^{\Lambda+1}(-1) = 2\Lambda(-1)^{n-1} C_{n-1}^{\Lambda+1}(1). \end{cases} \tag{30}$$

4. THE PLANE WAVE IN GEGENBAUER POLYNOMIALS BASIS AND BOUNDARY CONDITIONS IN THE Y DIRECTION: S-MATRIX ALGORITHM

Usually when solving problems of diffraction from lamellar gratings, with modal methods, one follows roughly the main steps consisting in (i) solving Maxwell's equations through an eigenvalue problem in the incidence, the transmittance and the grating regions, (ii) writing the appropriate boundary conditions (TE or TM) and (iii) solving the resulting algebraic system through the *S-matrix* algorithm for example. Except for the original Fourier Modal Method (FMM) where the solutions in the homogeneous media are given by the classical Rayleigh expansions, for the other modal approaches, it is necessary to solve *numerically* at least one eigenvalue problem [12]. This can lead to numerical difficulties if one is dealing with normal incidence or the Littrow configuration where some eigenvalues are degenerate which make it difficult to associate them with the appropriate orders. This problem has been first encountered with the C-method [6] and has been solved by simply replacing propagating plane waves by their expressions in the new modal basis. For the sake of completeness we give, in the following, the expressions of these waves in terms of Gegenbauer polynomials. Let us consider the x dependence of the function describing a plane wave:

$$X_m(x) = e^{ik\alpha_m x}. \tag{31}$$

Let $X_m^s(x)$ be the restriction of $X_m(x)$ to the interval $\Omega_s = [a, b]$ and ξ be the reduced variable in the interval $[-1, 1]$ defined as follow:

$$x = \frac{b-a}{2}\,\xi + \frac{b+a}{2}. \tag{32}$$

If we set

$$\epsilon = \frac{b-a}{2} \qquad \text{and} \qquad \delta = \frac{b+a}{2}, \tag{33}$$

then

$$X_m^s(\xi) = e^{ik\alpha_m\delta}e^{ik\alpha_m\epsilon\xi} = \sum_{n=0}^{M} B_{nm,\Lambda}^{\Omega_s} C_n^{\Lambda}(\xi), \tag{34}$$

with

$$B_{nm,\Lambda}^{\Omega_s} = \frac{e^{ik\alpha_m\delta}}{h_n^{\Lambda}} \int_{-1}^{1} (1-\xi^2)^{\Lambda-\frac{1}{2}} C_n^{\Lambda}(\xi) e^{ik\alpha_m\xi} d\xi. \tag{35}$$

Since Fourier integrals of Gegenbauer polynomials can be expressed in terms of Bessel functions, this integral can be finally expressed under

the form:

$$B^{\Omega_s}_{nm,\Lambda} = \Gamma(\Lambda)\left(\frac{2}{k\alpha_m\epsilon}\right)^{\Lambda} i^n(n+\Lambda)J_{n+\Lambda}(k\alpha_m\epsilon)e^{ik\alpha_m\delta}. \tag{36}$$

Remark that, in the case of $\alpha_m = 0$, the orthogonal properties of Gegenbauer polynomials can be used. The numerical resolution of the wave equation, in a medium l defined by a refractive index function $\nu_l(x)$, gives $[1 \ : \ N_{\max}]$ eigenvectors denoted by

$$\mathbf{\Psi^l_p} = \left[\mathbf{a}^{\mathbf{l},i}_{[1\ :\ N-2],p}\right]_{i\in[1\ :\ N_\Omega]}. \tag{37}$$

In this nomenclature, i is referred to the subinterval Ω_i; i.e.:

$$\mathbf{a}^{\mathbf{l},i}_{[1\ :\ N-2],p} = \left[\mathbf{a}^{\mathbf{l},i,j=1}_{[1\ :\ N-2],p} \quad \mathbf{a}^{\mathbf{l},i,j=2}_{[1\ :\ N-2],p} \quad \dots \quad \mathbf{a}^{\mathbf{l},i,j=N_i}_{[1\ :\ N-2],p}\right]^t. \tag{38}$$

The matrix of the eigenvectors $\mathbf{\Psi^l}$ has consequently the following form:

$$\mathbf{\Psi^l} = \left[\mathbf{\Psi^l}_p\right]_{p\in[1\ :\ N_{\max}]}. \tag{39}$$

The second components needed for boundary conditions in y direction, i.e., E_x in TM polarization and H_x in TE polarization are represented by the following vector $\mathbf{\Phi^l_p}$:

$$\mathbf{\Phi^l_p} = \left[\frac{\beta^l_p}{\eta^i}\,\mathbf{a}^{\mathbf{l},i}_{[1\ :\ N-2],\,p}\right]_{i\in[1\ :\ N_\Omega]}. \tag{40}$$

According to the $\exp(-ik\beta^l_p y)$ dependence and in order to satisfy the outgoing Summerfeld condition, the root of eigenvalues β^l_p are sorted such that:

$$\left\{\beta^l_p\right\} = U^+ \cup U^- \tag{41}$$

with

$$U^+ = \left\{\beta^l_p \ — \ \beta^l_p \in \mathbb{R}^+ \text{ or } \left(\beta^l_p \in \mathbb{C} \text{ and } \Im(\beta^l_p) < 0\right)\right\}, \tag{42a}$$

$$U^- = \left\{\beta^l_p \ — \ \beta^l_p \in \mathbb{R}^- \text{ or } \left(\beta^l_p \in \mathbb{C} \text{ and } \Im(\beta^l_p) > 0\right)\right\}. \tag{42b}$$

The eigenvalues belonging to U^+ (resp U^-) and their corresponding eigenvectors are affected by the subscript $+$ (resp $-$). According to this convention, the S matrix of the interface separating the media l and $l+1$ has the following form:

$$\mathbf{S^l} = \begin{bmatrix} \mathbf{\Psi^l_+} & -\mathbf{\Psi^{l+1}_-} \\ \mathbf{\Phi^l_+} & -\mathbf{\Phi^{l+1}_-} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{\Psi^{l+1}_+} & -\mathbf{\Psi^l_-} \\ \mathbf{\Phi^{l+1}_+} & -\mathbf{\Phi^l_-} \end{bmatrix}. \tag{43}$$

5. NUMERICAL RESULTS AND DISCUSSION

In order to discuss the issue of stability of the MMGE we chose a case known to be rather difficult for modal methods: a highly conducting lamellar grating with the following parameters: $\nu_1 = 1$, $\nu_{21} = 1 - 40i$, $\nu_{22} = 1$, $\nu_3 = \nu_{21}$, $h = 0.4\lambda$, $d = 1.2361\lambda$, $f = 0.57$, and $\theta = \arcsin(\lambda/2d)$. We give, as an example, the R_{-1} efficiency computed via the classical MMGE (that we will designate MMGE1 from now on) without subdividing the two homogeneous regions $\Omega_{1/2}$, i.e., $N_1 = N_2 = 1$. For a shake of fluidity these results are given for only one value of $\Lambda(\Lambda = 0.5)$. Nevertheless, the conclusions deduced from this study still valid for any value of Λ. Figure 2 summarizes the results for both TE and TM polarizations as the total number of polynomials N is varied.

As can be seen, the MMGE1 witnesses instabilities as the number of polynomials is increased regardless of the polarization. To overcome this problem, it has been proposed [10] to subdivide the subintervals Ω_i into layers. The upper part of the Table 1 contains the results when each homogeneous subinterval is subdivided into four layers $(N_1 = N_2 = 4)$.

The gain in stability is clearly established. Nevertheless, in the case of MMGE1 with subdivisions, one can notice that the size of the matrices increases too rapidly in comparison with the former case; i.e.,

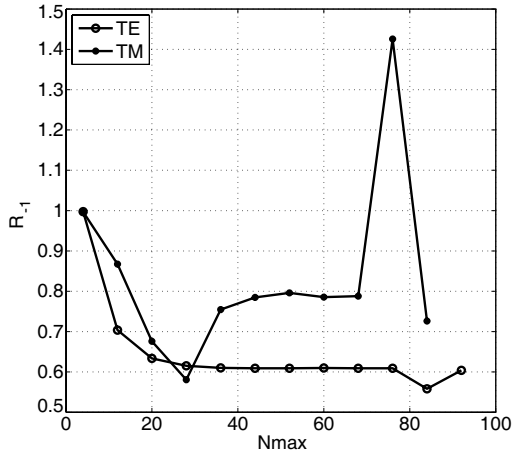

Figure 2. Minus-first order reflected efficiency R_{-1} of highly conductive metal in TE and TM polarizations. Instability of the results obtained by MMGE1 for $N_1 = N_2 = 1$.

Table 1. Minus-first order reflected efficiency R_{-1} of highly conductive metal in TE and TM polarizations. Comparison between the results obtained by MMGE1 for $N_1 = N_2 = 4$ and those obtained by MMGE2 for $N_1 = N_2 = 1$.

MMGE1			
N	$N_{\max}$	TE polarization	TM polarization
4	16	0.67443	0.95306
8	48	0.62024	0.52578
12	80	0.60925	0.78212
16	112	0.60875	0.79040
20	144	0.60875	0.79057
24	176	0.60875	0.79057
MMGE2			
N	$N_{\max}$	TE polarization	TM polarization
16	28	0.61503	0.58047
32	60	0.60873	0.79048
36	68	0.60875	0.79056
40	76	0.60875	0.79057
44	84	0.60875	0.79057
48	92	0.60875	0.79057

MMGE1 without subdivisions. This can be a serious drawback since the eigenvalue problem is the most time consuming step in the MMGE approach. The current version of the MMGE (MMGE2) removes this disadvantage. The results obtained by the MMGE2 are presented in the lower part of the Table 1 and show a remarkable stability. Under TE polarization, the forth digit of $R_{-1} = 0.6087$ is stabilized as soon as $N_{\max}$ reaches 60 while it is necessary to use $N_{\max} = 112$ in order to have the same result with the MMGE1 ($N_1 = N_2 = 4$). Under TM polarization, the same behavior is observed on $R_{-1} = 0.7905$ as soon as $N_{\max}$ reaches 76; while it is necessary to use $N_{\max} = 144$ with the MMGE1. Thus with the new implementation of the MMGE one not only gains in stability but lowers the computational cost by using smaller matrices.

In order to highlight how striking the improvement of convergence brought by the MMGE is, we present in Figures 3(a) and 3(b) a comparison between the results obtained by the Fourier Modal Method and MMGE2 for three arbitrary values of Λ: $\Lambda = 0.0005, 0.5$ and 1.

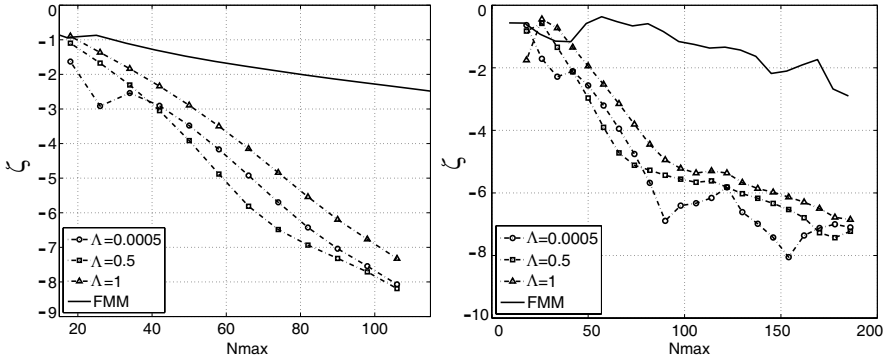

Figure 3. Minus-first order reflected efficiency of highly conductive metal obtained by MMGE2 in (a) TE and (b) TM polarizations. Comparison with Fourier modal method.

For that purpose, we introduced the error function:

$$\zeta(N_{\max}) = \log_{10} \mid 1 - \frac{R_{-1}(N_{\max} - \sum_{i=1}^{N_\Omega} N_i)}{R_{-1}(N_{\max})} \mid, \tag{44}$$

that measures the relative variation of the efficiency.

After this discussion on the stability, let us turn to the study of the influence of the parameter Λ in the MMGE2 implementation. It is of fundamental importance to notice that this parameter acts through the weighting function $w(\xi)$, also known as the density function, by introducing the measure $(1-\xi^2)^{\Lambda-1/2}d\xi$ in the integral of the inner product. The density increases or decreases in the vicinity of $\xi = \pm 1$ depending on the value of Λ. When Λ is close to zero, it is easy to see that the density increases around $\xi = \pm 1$; and in this case Gegenbauer polynomials are similar to Chebychev ones. The particular case of $\Lambda = 0.5$, which, corresponds to Legendre polynomials, involves an equipartition on the interval $[-1, 1]$. This density decreases around $\xi = \pm 1$ when Λ is greater than 0.5. All this, is from a theoretical point of view. In practice, it is reasonable to think, that for a given grating problem at least one optimal value of Λ may exist that allows describing the field at best. Indeed if we look to Tables 2 and 3, we find that the convergence is slightly improved in the case of TM polarization for $\Lambda = 0.45$ while $\Lambda = 0.5$ seems to be the best value in the case of TE polarization.

Other numerical investigations (not reported here) concerning the filling factor, the dielectric permittivity and the angle of incidence did not allow us to draw a general rule about the choice of Λ. All we can

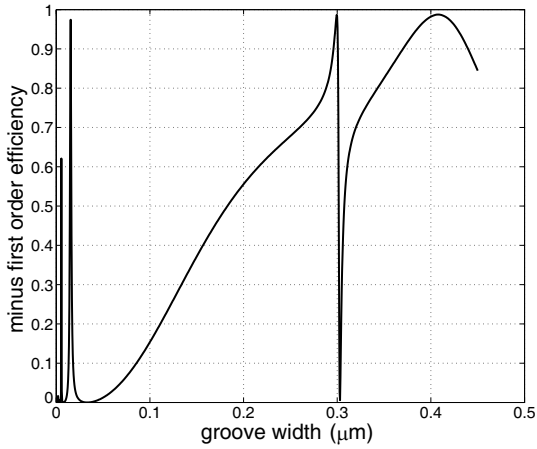

Figure 4. Minus-first order reflected efficiency of a metallic lamellar grating with respect to the groove width in TM polarization. Numerical parameters: $\theta = 30°$, $\nu_3 = \nu_{22} = 10i$, $\nu_1 = \nu_{21} = 1$, $d = h = 0.5\,\mu\text{m}$, $\lambda = 0.6328\,\mu\text{m}$, $N_{\max} = 20$.

Table 2. Minus-first order reflected efficiency obtained by MMGE2 of highly conductive metal in TE polarization. Influence of the parameter Λ in the convergence of the results.

TE polarization						
$N_{\max}$	N	$\Lambda = 5e-4$	$\Lambda = 0.15$	$\Lambda = 0.25$	$\Lambda = 0.45$	$\Lambda = 0.5$
50	27	0.60854	0.60858	0.60863	0.60877	0.60882
54	29	0.60865	0.60866	0.60868	0.60875	0.60877
58	31	0.60870	0.60871	0.60871	0.60874	0.60875
62	33	0.60873	0.60873	0.60873	0.60874	0.60875
66	35	0.60874	0.60874	0.60874	0.60874	0.60875
70	37	0.60874	0.60874	0.60874	0.60874	0.60875
74	39	0.60874	0.60874	0.60874	0.60874	0.60875
78	41	0.60874	0.60874	0.60874	0.60875	0.60875
82	43	0.60874	0.60874	0.60874	0.60875	0.60875
86	45	0.60874	0.60874	0.60874	0.60875	0.60875
90	47	0.60875	0.60875	0.60875	0.60875	0.60875
94	49	0.60875	0.60875	0.60875	0.60875	0.60875

Table 3. Minus-first order reflected efficiency obtained by MMGE2 of highly conductive metal in TM polarization. Influence of the parameter Λ in the convergence of the results.

TM polarization						
$N_{\max}$	N	$\Lambda = 5e-4$	$\Lambda = 0.15$	$\Lambda = 0.25$	$\Lambda = 0.45$	$\Lambda = 0.5$
50	27	0.79273	0.79229	7.9178	0.79023	0.78972
54	29	0.79164	0.79149	7.9128	0.79054	0.79028
58	31	0.79107	0.79102	7.9094	0.79060	0.79048
62	33	0.79079	0.79078	7.9075	0.79060	0.79054
66	35	0.79067	0.79067	7.9065	0.79059	0.79056
70	37	0.79061	0.79061	7.9061	0.79058	0.79057
74	39	0.79059	0.79059	7.9059	0.79058	0.79057
78	41	0.79058	0.79058	7.9058	0.79058	0.79057
82	43	0.79058	0.79058	7.9058	0.79058	0.79057
86	45	0.79058	0.79058	7.9058	0.79058	0.79057
90	47	0.79058	0.79058	7.9058	0.79058	0.79058
94	49	0.79058	0.79058	7.9058	0.79058	0.79058

assert is that the extreme values of the interval [0, 0.5] are far from being the optimal in certain cases.

Finally, we test the stability of our approach over a grating configuration that posed numerical problems to the FMM [13] and for which solutions have been proposed by some authors [14, 15]. It consists of a grating with a relative dielectric permittivity equal to -100 and all the other parameters are given in Figure 4.

The minus-first reflected order is drawn versus the groove width. It is remarkable to see that this curve is obtained by $N_{\max}$ as small as 20 without any instability.

6. CONCLUSION

In the present work, the modal method by Gegenbauer polynomials expansions has been improved by getting rid of undesirable instabilities. These ones have been shown to be directly linked to the computation of the Gegenbauer polynomials through a series sum. The introduction of a suitable weighting function in the inner product gives an additional degree of freedom that allows for stable and efficient evaluation of this latter. Thus the current version of the MMGE is very stable, accurate and converges very rapidly. It clearly extends the domain of action of modal methods with outstanding performances.

APPENDIX A. AN EXAMPLE OF THE MATRICES CONSTRUCTION IN THE CASE OF TWO MEDIA Ω_1 AND Ω_2

For the both media, the resolution of the wave equation leads to:

$$\begin{bmatrix} \mathbf{L}^{\Omega_1}_{[N-2]\times[N]} & 0 \\ 0 & \mathbf{L}^{\Omega_2}_{[N-2]\times[N]} \end{bmatrix} \begin{bmatrix} \mathbf{a}^{\Omega_1}_{[1\,:\,N],p} \\ \mathbf{a}^{\Omega_2}_{[1\,:\,N],p} \end{bmatrix} = \beta_p^2 \begin{bmatrix} \mathbf{G}^{\Omega_1}_{[N-2]\times[N]} & 0 \\ 0 & \mathbf{G}^{\Omega_2}_{[N-2]\times[N]} \end{bmatrix} \begin{bmatrix} \mathbf{a}^{\Omega_1}_{[1\,:\,N],p} \\ \mathbf{a}^{\Omega_2}_{[1\,:\,N],p} \end{bmatrix}. \tag{A1}$$

By using the boundary conditions of Eqs. (4), (5) and (6), the expansion coefficients of highest orders $a^{\Omega_1}_{(N-1),p}$, $a^{\Omega_1}_{N,p}$, $a^{\Omega_2}_{(N-1),p}$ and $a^{\Omega_2}_{N,p}$ are expressed in terms of the coefficients of lower orders. For that purpose let us set:

$$\mathbf{T}^{\Omega_i}_{[n]}(x_0,\kappa) = \kappa \begin{bmatrix} \left[b^{\Omega_i}_{[n]}(x_0)\right] \\ \left[\dfrac{db^{\Omega_i}_{[n]}}{dx}\right]_{x_0} \end{bmatrix}. \tag{A2}$$

Eqs. (4), (5) and (6) lead to

$$\begin{bmatrix} \mathbf{a}^{\Omega_1}_{[N-1,N],p} \\ \mathbf{a}^{\Omega_2}_{[N-1,N],p} \end{bmatrix} = \mathbf{T} \begin{bmatrix} \mathbf{a}^{\Omega_1}_{[1\,:\,N-2],p} \\ \mathbf{a}^{\Omega_2}_{[1\,:\,N-2],p} \end{bmatrix}, \tag{A3}$$

where the matrix $\mathbf{T}$ is defined as follows:

$$\mathbf{T} = -\begin{bmatrix} \mathbf{T}^{\Omega_1}_{[N-1,N]}(0,1) & -\mathbf{T}^{\Omega_2}_{[N-1,N]}(0,1) \\ \mathbf{T}^{\Omega_1}_{[N-1,N]}(fd-d,\tau) & -\mathbf{T}^{\Omega_2}_{[N-1,N]}(fd,1) \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{T}^{\Omega_1}_{[1\,:\,N-2]}(0,1) & -\mathbf{T}^{\Omega_2}_{[1\,:\,N-2]}(0,1) \\ \mathbf{T}^{\Omega_1}_{[1\,:\,N-2]}(fd-d,\tau) & -\mathbf{T}^{\Omega_2}_{[1\,:\,N-2]}(fd,1) \end{bmatrix}. \tag{A4}$$

$\tau = e^{ik\sin\theta d}$ is the pseudo-periodic factor and f denoted the filling factor of the lamellar grating. Therefore the vector $\left[\mathbf{a}^{\Omega_1}_{[1\,:\,N],p},\, \mathbf{a}^{\Omega_2}_{[1\,:\,N],p}\right]^t$ of Eq. (A1) can be expressed in term of the vector $\left[\mathbf{a}^{\Omega_1}_{[1\,:\,N-2],p},\, \mathbf{a}^{\Omega_2}_{[1\,:\,N-2],p}\right]^t$ with the following matrix relation:

$$\begin{bmatrix} \mathbf{a}^{\Omega_1}_{[1\,:\,N],p} \\ \mathbf{a}^{\Omega_2}_{[1\,:\,N],p} \end{bmatrix} = \mathbf{C} \begin{bmatrix} \mathbf{a}^{\Omega_1}_{[1\,:\,N-2],p} \\ \mathbf{a}^{\Omega_2}_{[1\,:\,N-2],p} \end{bmatrix}. \tag{A5}$$

The matrix $\mathbf{C}$ can be written in the following concise form:

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix}. \tag{A6}$$

The matrices $\mathbf{C}_{ii}$ of $N \times (N-2)$ size, describe the coupling between higher and lower coefficients of a same subinterval namely $\mathbf{C}_{11}$ in the subinterval Ω_1 and $\mathbf{C}_{22}$ in Ω_2:

$$\mathbf{C}_{11} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \ddots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ \mathbf{T}_{1,1} & \mathbf{T}_{1,2} & \dots & \mathbf{T}_{1,N-2} \\ \mathbf{T}_{2,1} & \mathbf{T}_{2,2} & \dots & \mathbf{T}_{2,N-2} \end{bmatrix}, \tag{A7}$$

$$\mathbf{C}_{22} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \ddots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ \mathbf{T}_{3,N-2+1} & \mathbf{T}_{3,N-2+2} & \dots & \mathbf{T}_{3,2(N-2)} \\ \mathbf{T}_{4,N-2+1} & \mathbf{T}_{4,N-2+2} & \dots & \mathbf{T}_{4,2(N-2)} \end{bmatrix}, \tag{A8}$$

whereas $\mathbf{C}_{ij}$, $i \neq j$ of $N \times (N-2)$ size take onto account the interconnection between the subintervals:

$$\mathbf{C}_{12} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \\ \mathbf{T}_{1,N-2+1} & \mathbf{T}_{1,N-2+2} & \dots & \mathbf{T}_{1,2(N-2)} \\ \mathbf{T}_{2,N-2+1} & \mathbf{T}_{2,N-2+2} & \dots & \mathbf{T}_{2,2(N-2)} \end{bmatrix}, \tag{A9}$$

$$\mathbf{C}_{21} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \\ \mathbf{T}_{3,1} & \mathbf{T}_{3,2} & \dots & \mathbf{T}_{3,N-2} \\ \mathbf{T}_{4,1} & \mathbf{T}_{4,2} & \dots & \mathbf{T}_{4,N-2} \end{bmatrix}. \tag{A10}$$

It is easy to show that

$$\begin{bmatrix} \mathbf{G}^{\Omega_1}_{[N-2]\times[N]} & 0 \\ 0 & \mathbf{G}^{\Omega_2}_{[N-2]\times[N]} \end{bmatrix} \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{G}^{\Omega_1}_{[N-2]\times[N-2]} & 0 \\ 0 & \mathbf{G}^{\Omega_2}_{[N-2]\times[N-2]} \end{bmatrix}, \tag{A11}$$

consequently, Eq. (A1) is written as follows:

$$\begin{bmatrix} \mathbf{L}^{\Omega_1}_{[N-2]\times[N]} & 0 \\ 0 & \mathbf{L}^{\Omega_2}_{[N-2]\times[N]} \end{bmatrix} \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{a}^{\Omega_1}_{[1\,:\,N-2],p} \\ \mathbf{a}^{\Omega_2}_{[1\,:\,N-2],p} \end{bmatrix} = \beta_p^2 \begin{bmatrix} \mathbf{G}^{\Omega_1}_{[N-2]\times[N-2]} & 0 \\ 0 & \mathbf{G}^{\Omega_2}_{[N-2]\times[N-2]} \end{bmatrix} \begin{bmatrix} \mathbf{a}^{\Omega_1}_{[1\,:\,N-2],p} \\ \mathbf{a}^{\Omega_2}_{[1\,:\,N-2],p} \end{bmatrix}. \tag{A12}$$

Finally in the case of two media, we are led to the computation of the eigenvalues β_p^2 and their associated eigenvectors $\left[\mathbf{a}^{\Omega_1}_{[1\,:\,N-2],p},\ \mathbf{a}^{\Omega_2}_{[1\,:\,N-2],p}\right]^t$ of a matrix with dimension $N_{\max} = (2(N-2)) \times (2(N-2))$.

REFERENCES

1. Botten, I. C., M. S. Craig, R. C. McPhedran, J. L. Adams, and J. R. Andrewartha, "The dielectric lamellar diffraction grating," *Optica Acta*, Vol. 28, 413–428, 1981.
2. Botten, I. C., M. S. Craig, R. C. McPhedran, J. L. Adams, and J. R. Andrewartha, "The finitely conducting lamellar diffraction grating," *Optica Acta*, Vol. 28, 1087–1102, 1981.
3. Knop, K., "Rigorous diffraction theory for transmission phase gratings with deep rectangular grooves," *J. Opt. Soc. Am.*, Vol. 68, 1206–1210, 1978.
4. Li, L., "New formulation of the fourier modal method for crossed surface-relief gratings," *J. Opt. Soc. Am.*, Vol. 14, 2758–2767, 1997.
5. Morf, R. H., "Exponentially convergent and numerically efficient solution of Maxwell's equations for lamellar gratings," *J. Opt. Soc. Am. A*, Vol. 12, 1043–1056, 1995.
6. Plumey, J. P., B. Guizal, and J. Chandezon, "Coordinate transformation method as applied to asymmetric gratings with vertical facets," *J. Opt. Soc. Am. A*, Vol. 14, 610–617, 1997.

7. Edee, K., P. Schiavone, and G. Granet, "Analysis of defect in E.U.V. lithography mask using a modal method by nodal B-spline expansion," *Japanese Journal of Applied Physics*, Vol. 44, No. 9A, 6458–6462, 2005.
8. Armeanu, A. M., M. K. Edee, G. Granet, and P. Schiavone, "Modal method based on spline expansion for the electromagnetic analysis of the lamellar grating," *Progress In Electromagnetics Research*, Vol. 106, 243–261, 2010.
9. Harrington, R. F., *Field Computation by Moment Methods*, The Macmillan Company, New York, 1968, reprinted by IEEE Press, New York, 1993.
10. Edee, K., "Modal method based on subsectional Gegenbauer polynomial expansion for lamellar gratings," *J. Opt. Soc. Am.*, Vol. 28, 2006–2013, 2011.
11. Hochstrasser, U. W., "Orthogonal polynomials," *Handbook of Mathematical Functions*, 771–802, M. Abramowitz and I. A. Stegun, eds., Dover, 1965.
12. Yala, H., B. Guizal, and D. Felbacq, "Fourier modal method with spatial adaptive resolution for structures comprising homogeneous layers," *J. Opt. Soc. Am. A*, Vol. 26, 2567–2570, 2009.
13. Popov, E., B. Chernov, M. Nevière, and N. Bonod, "Differential theory: Application to highly conducting gratings," *J. Opt. Soc. Am. A*, Vol. 21, 199–206, 2004.
14. Lyndin, N. M., O. Parriaux, and A. V. Tishchenko, "Modal analysis and suppression of the Fourier modal method instabilities in highly conductive gratings," *J. Opt. Soc. Am. A*, Vol. 24, 3781–3788, 2007.
15. Guizal, B., H. Yala, and D. Felbacq, "Reformulation of the eigenvalue problem in the Fourier modal method with spatial adaptive resolution," *Opt. Lett. A*, Vol. 34, 2790–2792, 2009.