Advances in Difference Equations
a SpringerOpen Journal

RESEARCH Open Access

Galerkin method for the scattering problem of strip gratings

Enxi Zheng[1*] and Yujie Wang[1]

*Correspondence:
enxizheng2003@dlmu.edu.cn
[1]School of Science, Dalian Maritime University, Dalian, P.R. China

Abstract

In this paper, the diffraction problem of periodic strip gratings is considered. The previous study of this problem usually concentrated on the numerical method; however, we try to analyze this problem and the convergence of the numerical solution from the mathematical point of view in this work. By use of the Dirichlet to Neumann operator on the slit between two strips, we reformulate the problem to an operator equation. The well-posedness of the solution to the operator equation is proved. The Galerkin method is applied to solve this operator equation and the convergence result of the numerical solution is also derived. Finally, some numerical experiments are presented to show the effectiveness of our method and verify the theoretical convergence result.

Keywords: Strip grating; Dirichlet to Neumann map; Helmholtz equation; Galerkin method

1 Introduction

The scattering theory of periodic structures has a wide variety of applications, for example, the micro-optics and the antenna engineering. In optics, the periodic structure is also called diffraction gratings. Introduction to the problem of electromagnetic diffraction through periodic structures and the corresponding numerical methods can be found in [1]. The reviews on the diffractive optics technology and the mathematical analysis of diffraction gratings are presented in [2] and [3], respectively.

In this paper, we focus on the diffraction problem of periodic perfectly conducting strip gratings. This is a classic model which has been investigated by many researchers. In [4] the method of moments (MoM) is employed to analyze the diffraction problem of strip gratings located in free space. In papers [5, 6] the same numerical method is used to solve the strip grating problem where the strips are printed on a dielectric substrate. As to the improved MoM for scattering problem of periodic strip gratings, we refer to paper [7] and the references therein. In addition to the MoM, there are also many other numerical methods for the strip gratings problem. For example, the singular integral equation approach is proposed to deal with the plane wave diffraction by an infinite strip grating at oblique incidence in [8]. Fourier modal method (FMM), also called combined boundary conditions method (CBCM), is introduced in [9–11]. CBCM is applied for solving finite strip grating problem in [11]. Because of the slow convergence and Gibbs phenomenon at the tips of the strips, CBCM has been substantially improved in [12] with the aid of adaptive spatial resolution in [11]. The parametric formulation of CBCM in [12] improves the convergence rate, the computational efficiency, and the numerical accuracy. In [13] the authors employ this improved method in multilayered structures of strip gratings. More references about CBCM can be found in [14]. In mathematics, because the grating diffraction problem is governed by differential equation, the finite difference method (FDM) and the finite element method (FEM) can also be applied in solving the problem. In the numerical experiment part of this paper, we compare our method with the FDM.

The studies above concentrate on the mathematical model and the numerical method of the strip gratings. In this paper, we will reformulate the diffraction problem of strip grating into an operator equation by use of the Dirichlet to Neumann (DtN) operator. Then we will give a rigorous mathematical analysis of the solution and the Galerkin method which is used to solve the operator equation. The convergence of the Galerkin method has also been verified by the numerical experiments.

The paper is organized as follows. In Sect. 2, we give some notations for describing the strip gratings problem and reformulate this problem into an operator equation by use of DtN operator on the slit between two strips. In Sect. 3, the well-posedness of the solution to the operator equation derived in Sect. 2 is proved. The well-posedness contains the existence, the uniqueness, and the stability of the solution. In Sect. 4, the Galerkin method is employed to solve the operator equation. The uniqueness and the convergence of the numerical solution are proved. We also obtain the error estimate of the numerical solution in Theorem 5. The concrete computing process is given at the end of this section. In Sect. 5, some numerical experiments are presented to show the effectiveness of the Galerkin method and convergence order proved in Sect. 4 is also verified by numerical Example 1 in this section.

2 Formulation of the problem

In this section, some notations are firstly presented to help us describe the diffraction problem of periodic strip gratings with period d. Then the definition of quasi-periodic, the famous Rayleigh expansion of diffraction field, and the DtN operator are introduced. Finally, we reformulate the diffraction problem into an operator equation through simple calculation and the knowledge prepared above.

Assume that the length of each slit is L ($L < d$), and the slits are separated by perfect conductive material strips $\tilde{\Gamma}$, see Figure 1. Denote the slit and the perfect conductive strip in one period by Γ_0 and Γ_1, respectively:

$$\Gamma_0 = \{(x_1, x_2) \in \mathbb{R}^2; 0 < x_1 < L, x_2 = 0\},$$

$$\Gamma_1 = \{(x_1, x_2) \in \mathbb{R}^2; L < x_1 < d, x_2 = 0\}.$$

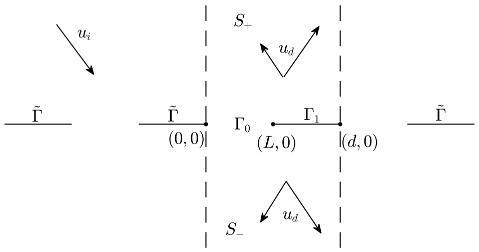

Figure 1 The diffraction problem of strip grating

Furthermore, define Γ as follows:

$$\Gamma = \{(x_1, x_2) \in \mathbb{R}^2; 0 < x_1 < d, x_2 = 0\}.$$

Suppose that the domain

$$S = \{(x_1, x_2) \in \mathbb{R}^2; 0 < x_1 < d, (x_1, x_2) \notin \Gamma_1\}$$

is filled with a homogeneous medium. Let S_+ and S_- denote the domain upon and below the strip gratings, respectively:

$$S_+ = \{(x_1, x_2) \in \mathbb{R}^2; 0 < x_1 < d, x_2 > 0\},$$
$$S_- = \{(x_1, x_2) \in \mathbb{R}^2; 0 < x_1 < d, x_2 < 0\}.$$

Assume that the plane wave $u_i = e^{i\alpha x_1 - i\beta x_2}$ is the incident wave upon the grating, where $\alpha = k\sin\theta$, $\beta = k\cos\theta$, $k > 0$ is the wave number, and $\theta \in (-\pi/2, \pi/2)$ is the angle of incidence. Denote the diffracted field by u_d. Then the total field u_t is

$$u_t = \begin{cases} u_i + u_d, & x_2 > 0, \\ u_d, & x_2 < 0, \end{cases} \tag{1}$$

and the diffraction problem of strip gratings reads as follows: when the incident field u_i is given, find the total field u_t such that

$$\Delta u_t + k^2 u_t = 0, \quad \text{in } S, \tag{2}$$
$$u_t = 0, \quad \text{on } \Gamma_1. \tag{3}$$

Among all the solutions of equations (2) and (3), we are interested in the quasi-periodic solution, i.e., $u_t e^{-i\alpha x}$ is a periodic function in x_1 with period d. Moreover, we require the diffracted field u_d to satisfy the bounded outgoing wave condition in S_+ and S_-.

In domains S_+ and S_-, the famous Rayleigh expansion of u_d is

$$u_d(x_1, x_2) = \sum_{n=-\infty}^{\infty} a_n e^{i(\alpha_n + \alpha)x_1 + i\beta_n x_2}, \quad (x_1, x_2) \in S_+, \tag{4}$$
$$u_d(x_1, x_2) = \sum_{n=-\infty}^{\infty} b_n e^{i(\alpha_n + \alpha)x_1 - i\beta_n x_2}, \quad (x_1, x_2) \in S_-, \tag{5}$$

where $\alpha_n = 2\pi n/d$,

$$\beta_n = \begin{cases} (k^2 - (\alpha_n + \alpha)^2)^{1/2}, & n \in A, \\ i((\alpha_n + \alpha)^2 - k^2)^{1/2}, & n \notin A, \end{cases}$$

$A = \{n \in \mathbb{Z}, k^2 - (\alpha_n + \alpha)^2 > 0\}$ and a_n, b_n are coefficients to be determined. We further assume that $k \neq |\alpha_n + \alpha|$ for every $n \in \mathbb{Z}$ in order to avoid resonances. Define

$$u_d^+(x_1) = \lim_{x_2 \to 0_+} u_d(x_1, x_2), \qquad u_d^-(x_1) = \lim_{x_2 \to 0_-} u_d(x_1, x_2),$$

$$\frac{\partial u_d^+}{\partial x_2}(x_1) = \lim_{x_2\to 0_+}\frac{\partial u_d}{\partial x_2}(x_1,x_2), \qquad \frac{\partial u_d^-}{\partial x_2}(x_1) = \lim_{x_2\to 0_-}\frac{\partial u_d}{\partial x_2}(x_1,x_2),$$

and $u_i^+(x_1)$, $\frac{\partial u_t^-}{\partial x_2}(x_1)$, *etc.* are defined in the same way. Then we have the following integral expression of a_n and b_n:

$$a_n = \frac{1}{d}\int_0^d u_d^+(x_1)e^{-i(\alpha_n+\alpha)x_1}\,\mathrm{d}x_1,$$
$$b_n = \frac{1}{d}\int_0^d u_d^-(x_1)e^{-i(\alpha_n+\alpha)x_1}\,\mathrm{d}x_1.$$

For any quasi-periodic function $f(x_1)$, i.e., $f(x_1)$ has the following expansion:

$$f(x_1) = \sum_{n=-\infty}^{+\infty} f_n e^{i(\alpha_n+\alpha)x_1}, \quad \text{where } f_n = \frac{1}{d}\int_0^d f(x_1)e^{-i(\alpha_n+\alpha)x_1}\,\mathrm{d}x_1,$$

define the Dirichlet to Neumann operator T_1 and T_2 as follows:

$$T_1 f(x_1) = \sum_{n=-\infty}^{+\infty} i\beta_n f_n e^{i(\alpha_n+\alpha)x_1},$$
$$T_2 f(x_1) = \sum_{n=-\infty}^{+\infty} -i\beta_n f_n e^{i(\alpha_n+\alpha)x_1}.$$

By simple calculation, we can get

$$\frac{\partial u_d^+}{\partial x_2}(x_1) = T_1\bigl(u_d^+\bigr)(x_1), \quad 0 < x_1 < d, \tag{6}$$
$$\frac{\partial u_d^-}{\partial x_2}(x_1) = T_2\bigl(u_d^-\bigr)(x_1), \quad 0 < x_1 < d. \tag{7}$$

Since the total field u_t and the normal derivative $\frac{\partial u_t}{\partial x_2}$ are continuous across Γ_1 and $u_t = 0$ on Γ_0,

$$u_t^+(x_1) = u_t^-(x_1), \quad 0 < x_1 < d, \tag{8}$$
$$\frac{\partial u_t^+}{\partial x_2}(x_1) = \frac{\partial u_t^-}{\partial x_2}(x_1), \quad 0 < x_1 < L. \tag{9}$$

Let $u(x_1) = u_d^-(x_1)$, $x_1 \in (0,L)$, then

$$u_d^-(x_1) = E_0(u), \quad x_1 \in (0,d), \tag{10}$$

where E_0 is the zero extension operator defined by the following:

$$E_0(u) = \begin{cases} u(x_1), & 0 < x_1 < L, \\ 0, & L \le x_1 \le d. \end{cases}$$

From (1), (6)–(9), and (10) we obtain

$$u_i^+(x_1) + u_d^+(x_1) = E_0(u)(x_1), \quad x_1 \in (0,d), \tag{11}$$

$$\frac{\partial u_i^+}{\partial x_2}(x_1) + T_1\big(u_d^+\big) = T_2\big(u_d^-\big), \quad x_1 \in (0, L). \tag{12}$$

Substitute (10) and (11) into (12)

$$\frac{\partial u_i^+}{\partial x_2}(x_1) + T_1\big(E_0(u) - u_i^+\big) = T_2\big(E_0(u)\big), \quad x_1 \in (0, L). \tag{13}$$

From the definitions of T_1 and T_2, we know that

$$T_1\big(u_i^+\big) = i\beta e^{i\alpha x_1}, \quad x_1 \in (0, L), \tag{14}$$

$$T_1\big(E_0(u)\big) = -T_2\big(E_0(u)\big), \quad x_1 \in (0, L). \tag{15}$$

Substitute (14) and (15) into (13)

$$T_2\big(E_0(u)\big) = -i\beta e^{i\alpha x_1}, \quad x_1 \in (0, L). \tag{16}$$

Denote $T = T_2 \circ E_0$ and $g(x_1) = -i\beta e^{i\alpha x_1}$, then equation (16) can be rewritten to

$$Tu = g, \quad x_1 \in (0, L). \tag{17}$$

Thus the diffraction problem of strip gratings can be reformulated into the operator equation (17). Once the solution of (17) is derived, the diffraction field u_d and the total field u_t can be obtained through (4), (5), and (1).

3 Well-posedness analysis

In this section, we will prove the well-posedness of the solution to (17). In order to give the uniqueness and existence of the solution to (17), we introduce the following definitions of spaces and norms.

For any real number s, define Sobolev spaces

$$H_\alpha^s(\Gamma) = \big\{u \in H^s(\Gamma); u(x_1)e^{-i\alpha x_1} \text{ is periodic in } x_1 \text{ with period } d\big\},$$

$$H_\alpha^s(\Gamma_0) = \big\{u \in \big(C_0^\infty(\Gamma_0)\big)', u = U|_{\Gamma_0}, \text{for some } U \in H_\alpha^s(\Gamma)\big\},$$

$$H_{\alpha,*}^s(\Gamma_0) = \big\{u \in \big(C_0^\infty(\Gamma_0)\big)', E_0(u) \in H_\alpha^s(\Gamma)\big\},$$

with norms

$$\|u\|_{s,\Gamma} = \left(d \sum_{n=-\infty}^{+\infty} \big(1 + (\alpha_n + \alpha)^2\big)^s |u_n|^2\right)^{\frac{1}{2}}, \quad u \in H_\alpha^s(\Gamma),$$

$$\|u\|_{s,\Gamma_0} = \inf_{U \in H_\alpha^s(\Gamma), U|_{\Gamma_0} = u} \|U\|_{s,\Gamma}, \quad u \in H_\alpha^s(\Gamma_0),$$

$$\|u\|_{s,*,\Gamma_0} = \big\|E_0(u)\big\|_{s,\Gamma}, \quad u \in H_{\alpha,*}^s(\Gamma_0),$$

where

$$u_n = \frac{1}{d}\int_0^d u(x_1)e^{-i(\alpha_n+\alpha)x_1}\,\mathrm{d}x_1.$$

Using the above notations, the operator equation problem (17) can be formulated as follows: given $g \in H_\alpha^{-\frac{1}{2}}(\Gamma_0)$, find $u \in H_{\alpha,*}^{\frac{1}{2}}(\Gamma)$ such that equation (17) is satisfied.

In the following, we will present two lemmas about the properties of spaces $H_*^{\frac{1}{2}}(\Gamma_0)$, $H_*^{-\frac{1}{2}}(\Gamma_1)$ and the operator T. With the aid of these two lemmas, we can obtain the well-posedness of the solution to equation (17).

Lemma 1 *The operator* $T : H_{\alpha,*}^{\frac{1}{2}}(\Gamma_0) \to H_\alpha^{-\frac{1}{2}}(\Gamma_0)$ *is a linear bounded operator.*

Proof Denote $u_n = \frac{1}{d}\int_0^d E_0(u)e^{-i(\alpha_n+\alpha)x_1}\,\mathrm{d}x_1$ for $u \in H_{\alpha,*}^{\frac{1}{2}}(\Gamma_0)$, and define

$$V(x_1) = \sum_{n=-\infty}^{+\infty} -i\beta_n u_n e^{i(\alpha_n+\alpha)x_1}, \quad x_1 \in (0,d).$$

Then $V|_{\Gamma_0} = Tu$. From the definition of $\|\cdot\|_{-\frac{1}{2},\Gamma_0}$, we have

$$\begin{aligned}\|Tu\|_{-\frac{1}{2},\Gamma_0} &= \inf_{U\in H_\alpha^s(\Gamma), U|_{\Gamma_0}=Tu} \|U\|_{-\frac{1}{2},\Gamma} \le \|V\|_{-\frac{1}{2},\Gamma} \\ &= \left(\sum_{n=-\infty}^{+\infty} \left[1+(\alpha_n+\alpha)\right]^{-\frac{1}{2}} |\beta_n|^2 |u_n|^2\right)^{\frac{1}{2}} \\ &\le \left(\sum_{n=-\infty}^{+\infty} \left[1+(\alpha_n+\alpha)\right]^{-\frac{1}{2}} \left[k^2+(\alpha_n+\alpha)^2\right] |u_n|^2\right)^{\frac{1}{2}} \\ &\le C_k\left(\sum_{n=-\infty}^{+\infty} \left[1+(\alpha_n+\alpha)\right]^{\frac{1}{2}} |u_n|^2\right)^{\frac{1}{2}} \\ &= C_k\|u\|_{\frac{1}{2},*,\Gamma_0},\end{aligned}$$

where $C_k = \max\{k,1\}^{\frac{1}{2}}$. □

Lemma 2 *The embedding operator* $I_* : H_{\alpha,*}^{\frac{1}{2}}(\Gamma_0) \to H_\alpha^{-\frac{1}{2}}(\Gamma_0)$ *is a compact operator.*

Proof The operator I_* can be decomposed into $I_* = R \circ I \circ E_0$, where E_0 is the zero extension operator from $H_{\alpha,*}^{\frac{1}{2}}(\Gamma_0)$ to $H_\alpha^{\frac{1}{2}}(\Gamma)$, I is the embedding operator from $H_\alpha^{\frac{1}{2}}(\Gamma)$ to $H_\alpha^{-\frac{1}{2}}(\Gamma)$, and R is the restriction operator from $H_\alpha^{-\frac{1}{2}}(\Gamma)$ to $H_\alpha^{-\frac{1}{2}}(\Gamma_0)$. Since E_0 and R are bounded and I is compact, we obtain that I_* is compact. □

Theorem 1 *Assume that* $\beta_n \neq 0$ *for all* $n \in \mathbb{Z}$, *then the homogeneous equation*

$$Tu = 0, \quad u \in H_{\alpha,*}^{\frac{1}{2}}(\Gamma_0)$$

has only one solution $u = 0$.

Proof From $Tu = 0$, we have $\langle Tu, u\rangle = 0$. Furthermore,

$$\langle Tu, u\rangle = \int_{\Gamma_0} Tu\bar{u}\,\mathrm{d}x_1 = \int_{\Gamma} Tu\overline{E_0(u)}\,\mathrm{d}x_1$$

$$\begin{aligned}
&= \int_{\Gamma} \sum_{n=-\infty}^{+\infty} -i\beta_n u_n e^{i(\alpha_n+\alpha)x_1} \overline{\sum_{n=-\infty}^{+\infty} u_n e^{i(\alpha_n+\alpha)x_1}} \,\mathrm{d}x_1 \\
&= -d \sum_{n=-\infty}^{+\infty} i\beta_n |u_n|^2 \\
&= -d \sum_{n\in A} i\beta_n |u_n|^2 - d \sum_{n\notin A} i\beta_n |u_n|^2.
\end{aligned}$$

From $\langle Tu, u\rangle = 0$, we can get

$$\begin{aligned}
&\operatorname{Re}\langle Tu, u\rangle = -d \sum_{n\notin A} i\beta_n |u_n|^2 = 0, \\
&\operatorname{Im}\langle Tu, u\rangle = -d \sum_{n\in A} \beta_n |u_n|^2 = 0.
\end{aligned}$$

Since $\beta_n \neq 0$ for all $n \in \mathbb{Z}$, from the above two equations, we can deduce that $u_n = 0$ for all $n \in \mathbb{Z}$. Thus $E_0(u) = 0$, i.e., $u = 0$. □

Theorem 2 *Assume that* $\beta_n \neq 0$ *for all* $n \in \mathbb{Z}$. *Then, for any* $g \in H_{\alpha}^{-\frac{1}{2}}(\Gamma_0)$, *the operator equation* $Tu = g$ *has a unique solution* $u \in H_{\alpha,*}^{\frac{1}{2}}(\Gamma_0)$, *and*

$$\|u\|_{\frac{1}{2},*,\Gamma_0} \leq C\|g\|_{-\frac{1}{2},\Gamma_0},$$

where $C > 0$ *is a constant independent of* g.

Proof Define operator B: $H_{\alpha,*}^{\frac{1}{2}}(\Gamma_0) \to H_{\alpha}^{-\frac{1}{2}}(\Gamma_0)$ and the corresponding bilinear form $b(\cdot,\cdot)$ as follows:

$$B = T + \sqrt{2}kI_*, \qquad b(u,v) = \langle Bu, v\rangle.$$

Since the operators T and I_* are bounded, $b(\cdot,\cdot)$ is a bounded bilinear form on $H_{\alpha,*}^{\frac{1}{2}}(\Gamma_0) \times H_{\alpha,*}^{\frac{1}{2}}(\Gamma_0)$. Next, we will show that $b(\cdot,\cdot)$ has a lower bound. By simple calculation, we can derive the following two inequalities:

$$\begin{aligned}
&\bigl|b(u,v)\bigr| \geq \frac{\sqrt{2}}{2}\bigl(\bigl|\operatorname{Re}\bigl\{b(u,v)\bigr\}\bigr| + \bigl|\operatorname{Im}\bigl\{b(u,v)\bigr\}\bigr|\bigr), \\
&\sqrt{\bigl|(\alpha_n+\alpha)^2 - k^2\bigr|} \geq \sqrt{\bigl|(\alpha_n+\alpha)^2 + k^2\bigr|} - \sqrt{2}k.
\end{aligned}$$

Thus

$$\begin{aligned}
\bigl|b(u,v)\bigr| &= \bigl|\langle Tu, v\rangle + \langle \sqrt{2}kI_* u, v\rangle\bigr| \\
&= \left|\int_{\Gamma} Tu\overline{E_0(u)}\,\mathrm{d}x_1 + \sqrt{2}k\int_{\Gamma} \bigl|E_0(u)\bigr|^2\,\mathrm{d}x_1\right| \\
&= \left|d\sum_{n=-\infty}^{+\infty} -i\beta_n|u_n|^2 + \sqrt{2}kd\sum_{n=-\infty}^{+\infty} |u_n|^2\right|
\end{aligned}$$

$$
\begin{aligned}
&= \Bigg| d\sum_{n\in A} -i\sqrt{k^2-(\alpha_n+\alpha)^2}|u_n|^2 \\
&\quad + d\sum_{n\notin A}\sqrt{(\alpha_n+\alpha)^2-k^2}|u_n|^2 + \sqrt{2}kd\sum_{n=-\infty}^{+\infty}|u_n|^2 \Bigg| \\
&\geq \frac{\sqrt{2}}{2}\Bigg(\Bigg| d\sum_{n\in A}\sqrt{k^2-(\alpha_n+\alpha)^2}|u_n|^2 \Bigg| \\
&\quad + \Bigg| d\sum_{n\notin A}\sqrt{(\alpha_n+\alpha)^2-k^2}|u_n|^2 + \sqrt{2}kd\sum_{n=-\infty}^{+\infty}|u_n|^2 \Bigg| \Bigg) \\
&\geq \frac{\sqrt{2}}{2}\Bigg(d\sum_{n=-\infty}^{+\infty}\sqrt{(\alpha_n+\alpha)^2+k^2}|u_n|^2 \\
&\quad - d\sum_{n\in A}\sqrt{2}k|u_n|^2 - d\sum_{n\notin A}\sqrt{2}k|u_n|^2 + \sqrt{2}kd\sum_{n=-\infty}^{+\infty}|u_n|^2 \Bigg) \\
&= \frac{\sqrt{2}}{2}\Bigg(d\sum_{n=-\infty}^{+\infty}\sqrt{(\alpha_n+\alpha)^2+k^2}|u_n|^2 \Bigg) \\
&\geq C\sum_{n=-\infty}^{+\infty}\sqrt{(\alpha_n+\alpha)^2+1}|u_n|^2 = C\|u\|_{\frac{1}{2},*,\Gamma_0},
\end{aligned}
$$

where $C = \frac{\sqrt{2}}{2}\min\{1,k\}$. By Lax–Milgram theorem, the operator B has a bounded inverse $B^{-1} : H_\alpha^{-\frac{1}{2}}(\Gamma_0) \to H_{\alpha,*}^{\frac{1}{2}}(\Gamma_0)$. Then the operator equation $Tu = g$ can be rewritten as

$$(B - \sqrt{2}kI_*)u = g.$$

From Theorem 1, we know that $-\sqrt{2}k$ is not an eigenvalue of B. Because I_* is a compact operator, by use of Fredholm alternative theorem, the operator equation $Tu = g$ has a unique solution $u \in H_{\alpha,*}^{\frac{1}{2}}(\Gamma_0)$ and

$$\|u\|_{\frac{1}{2},*,\Gamma_0} \leq C\|g\|_{-\frac{1}{2},\Gamma_0},$$

where C is a constant independent of g. □

4 Galerkin method

In this section we introduce the Galerkin method to solve equation (17) numerically. Assume

$$V_N = \mathrm{span}\{\varphi_1, \varphi_2, \ldots, \varphi_N\},$$

where

$$\varphi_n = \sin\left(\frac{n\pi x_1}{L}\right), \quad n = 1, 2, \ldots, N.$$

The Galerkin method to solve equation (17) is to find $u_N \in V_N$ such that

$$\langle Tu_N, \varphi_m\rangle = \langle g, \varphi_m\rangle, \quad m = 1, 2, \ldots, N, \tag{18}$$

where

$$\begin{aligned}\langle Tu_N, \varphi_m\rangle &= \int_0^L Tu_N \overline{\varphi_m}\, \mathrm{d}x_1,\\ \langle g, \varphi_m\rangle &= \int_0^L g\overline{\varphi_m}\, \mathrm{d}x_1.\end{aligned}$$

Theorem 3 *The Galerkin equation* (18) *has a unique solution* $u_N \in V_N$.

Proof For the finite dimensional problem (18), we only need to show that the homogeneous equation

$$\langle Tu_N, \varphi_m\rangle = 0, \quad m = 1, 2, \dots, N, \tag{19}$$

has only one solution $u_N = 0$. From (19), we can get $\langle Tu_N, u_N\rangle = 0$. Thus the proof for $u_N = 0$ is similar to Theorem 1.

The next theorem follows from standard estimates for Galerkin method associated with compact operator equation, so we omit the proof here and refer to [15] for details. □

Theorem 4 *Assume that* $\beta_n \neq 0$ *for all* $n \in \mathbb{Z}$, u *is the solution of* (17), *and* u_N *is the solution of* (18). *When* N *is large enough,*

$$\|u - u_N\|_{\frac{1}{2},*,\Gamma_0} \leq C \inf_{v_N \in V_N} \|u - v_N\|_{\frac{1}{2},*,\Gamma_0},$$

where C *is a positive constant independent of* N.

For $s \in \mathbb{R}$, define space $H_p^s(\Gamma_0)$ as follows:

$$H_p^s(\Gamma_0) = \left\{ u = \sum_{n=1}^{+\infty} a_n \sin\left(\frac{n\pi x}{L}\right); \frac{L}{2}\sum_{n=1}^{+\infty}\left[1 + \left(\frac{n\pi}{L}\right)^2\right]^s |a_n|^2 < +\infty \right\}.$$

The norm in $H_p^s(\Gamma_0)$ is given by

$$\|u\|_{s,p,\Gamma_0} = \left(\frac{L}{2}\sum_{n=1}^{+\infty}\left[1 + \left(\frac{n\pi}{L}\right)^2\right]^s |a_n|^2\right)^{\frac{1}{2}}.$$

With a little extension of Lemma 4.11 and Example 4.15 in [16] to quasi-periodic functions, the space $H_{\alpha,*}^s(\Gamma_0)$ is equal to $H_p^s(\Gamma_0)$, and the norms $\|\cdot\|_{s,*,\Gamma_0}$ and $\|\cdot\|_{s,p,\Gamma_0}$ are equivalent when $0 < s < 1$. Then the conclusion of Theorem 4 can be rewritten as

$$\|u - u_N\|_{\frac{1}{2},*,\Gamma_0} \leq C \inf_{v_N \in V_N} \|u - v_N\|_{\frac{1}{2},p,\Gamma_0}, \tag{20}$$

or

$$\|u - u_N\|_{\frac{1}{2},p,\Gamma_0} \leq C \inf_{v_N \in V_N} \|u - v_N\|_{\frac{1}{2},p,\Gamma_0}. \tag{21}$$

Theorem 5 *Assume that $\beta_n \neq 0$ for all $n \in \mathbb{Z}$, $u \in H_p^s(\Gamma_0)$ $(s > \frac{1}{2})$ is the solution of* (17), *and u_N is the solution of* (18), *then when N is large enough,*

$$\|u - u_N\|_{\frac{1}{2},*,\Gamma_0} \leq \frac{C}{N^{s-\frac{1}{2}}}\|u\|_{s,p,\Gamma_0}, \tag{22}$$

or

$$\|u - u_N\|_{\frac{1}{2},p,\Gamma_0} \leq \frac{C}{N^{s-\frac{1}{2}}}\|u\|_{s,p,\Gamma_0}, \tag{23}$$

where C is a positive constant independent of N.

Proof Because

$$\begin{aligned}
\inf_{v_N \in V_N} \|u - v_N\|_{\frac{1}{2},p,\Gamma_0} &\leq \left\| \sum_{n=1}^{+\infty} a_n \sin\left(\frac{n\pi x}{L}\right) - \sum_{n=1}^{N} a_n \sin\left(\frac{n\pi x}{L}\right) \right\|_{\frac{1}{2},p,\Gamma_0} \\
&= \left(\frac{L}{2} \sum_{n=N+1}^{+\infty} \left(1 + \left(\frac{n\pi}{L}\right)^2\right)^{\frac{1}{2}} |a_n|^2 \right)^{\frac{1}{2}} \\
&\leq \left(\frac{L}{2} \sum_{n=N+1}^{+\infty} \left(1 + \left(\frac{n\pi}{L}\right)^2\right)^{\frac{1}{2}-s} \left(1 + \left(\frac{n\pi}{L}\right)^2\right)^{s} |a_n|^2 \right)^{\frac{1}{2}} \\
&\leq \left(\frac{L}{2} \sum_{n=N+1}^{+\infty} \left(\frac{N\pi}{L}\right)^{1-2s} \left(1 + \left(\frac{n\pi}{L}\right)^2\right)^{s} |a_n|^2 \right)^{\frac{1}{2}} \\
&\leq \frac{C}{N^{s-\frac{1}{2}}}\|u\|_{s,p,\Gamma_0},
\end{aligned}$$

then (22) and (23) are derived by combining the above inequality with (20) and (21). □

5 Numerical results

In this section we demonstrate the numerical results of our method. All computations are performed using MATLAB. In all the following examples, we set the period $d = 4$, the length of the slit $L = 2$, and $\Gamma_0 = \{(x_1, 0); 0 < x_1 < 2\}$.

Example 1 In this example we consider the convergence results of our Galerkin method. Let $u(x_1) = x_1(x_1 - 2)$, $x_1 \in (0, 2)$ be the exact solution of the operator equation (17) with $k = 1$, $\alpha = 0$, and the corresponding right-hand side function

$$g(x_1) = -2 \sum_{n=-\infty}^{+\infty} i\beta_n \frac{n\pi + (-1)^n n\pi + 2i - (-1)^n 2i}{n^3\pi^3} e^{i\frac{n\pi x_1}{2}}.$$

Moreover, $u(x_1)$ has the following expansion:

$$u(x_1) = x_1(x_1 - 2) = \sum_{n=1}^{+\infty} \frac{32}{(2n-1)^3\pi^3} \sin\left(\frac{(2n-1)\pi x}{2}\right), \quad x \in (0, 2),$$

Table 1 $\|u - u_N\|_{\frac{1}{2},p,\Gamma_0}$ and $\|u - u_N\|_{\frac{1}{2},*,\Gamma_0}$ with respect to N

N	2	4	8	16	32	64
$\|u - u_N\|_{\frac{1}{2},p,\Gamma_0}$	9.0460E–02	2.6825E–02	7.1011E–03	1.8020E–03	4.5188E–04	1.1302E–04
order		1.7537	1.9175	1.9785	1.9956	1.9995
$\|u - u_N\|_{\frac{1}{2},*,\Gamma_0}$	8.4367E–02	2.5019E–02	6.6220E–03	1.6799E–03	4.2110E–04	1.0522E–04
order		1.7537	1.9177	1.9788	1.9959	1.9997

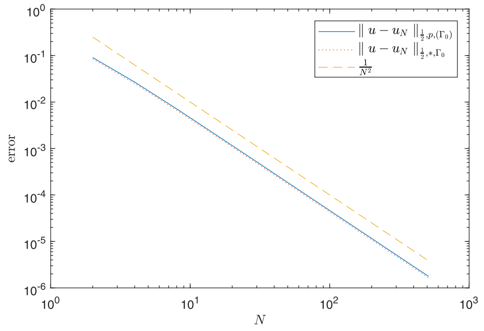

Figure 2 Numerical convergence results with respect to N in Example 1

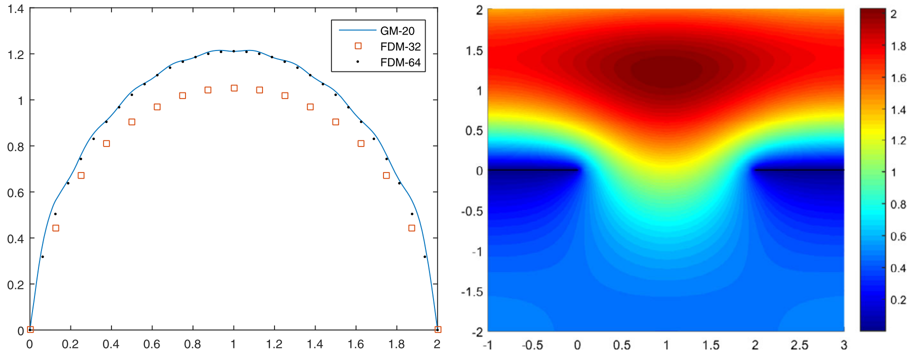

Figure 3 Example 2. (left) Absolute value of total field u_t on Γ_0, (right) Absolute value of total field u_t

so we have $u(x_1) \in H_p^{\frac{5}{2}-\varepsilon}(\Gamma_0)$ with $\varepsilon > 0$ arbitrarily small. Results are presented in Table 1 and Fig. 2. From these results, we can see that the errors $\|u - u_N\|_{\frac{1}{2},*,\Gamma_0}$ and $\|u - u_N\|_{\frac{1}{2},p,\Gamma_0}$ decay rapidly with respect to N, and the corresponding convergence orders are coincident with Theorem 5.

Example 2 We consider the situation of small wave number. Let the wave number $k = 1$, and the dimension of V_N be $N = 10$. The incident wave is plane wave with the incident angle $\theta = 0$. Results are shown in Figs. 3–4. In Fig. 3, we show the absolute value of the total field u_t on Γ_0 and the total field u_t in one period. In Fig. 4, we show the real part and the imaginary part of total field u_t. From these results we can see that our method is effective when the wave number k is small. Further, in the left one of Fig. 3, we also show the results given by FDM (finite difference method). We demonstrate two results given by FDM with 32 and 64 nodes in one period. Comparing with the FDM, we can see the

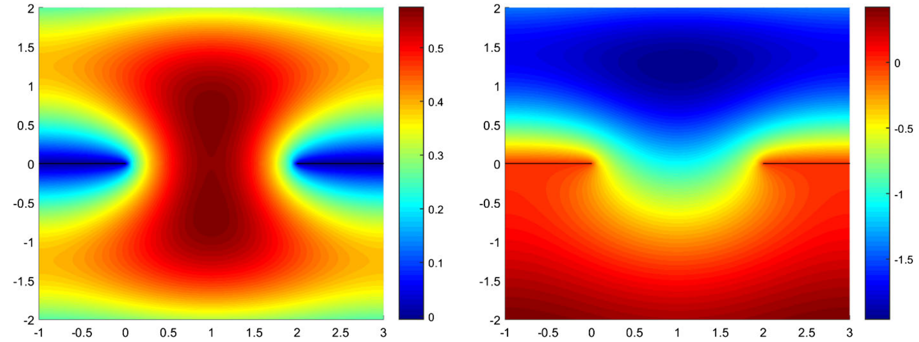

Figure 4 Example 2. (left) Real part of total field u_t, (right) Imaginary part of total field u_t

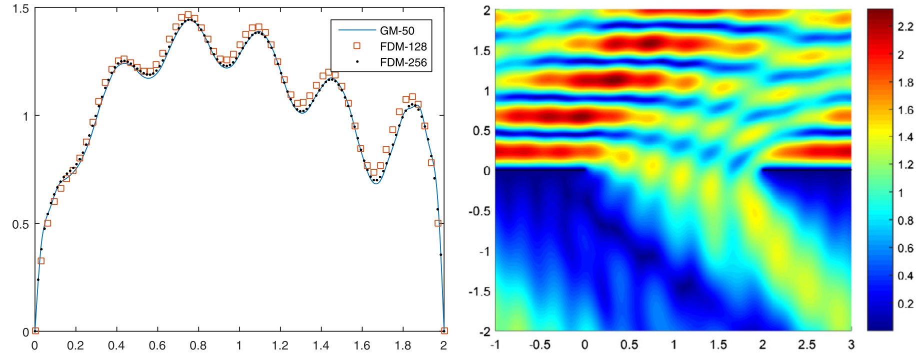

Figure 5 Example 3. (left) Absolute value of total field u_t on Γ_0, (right) Absolute value of total field u_t

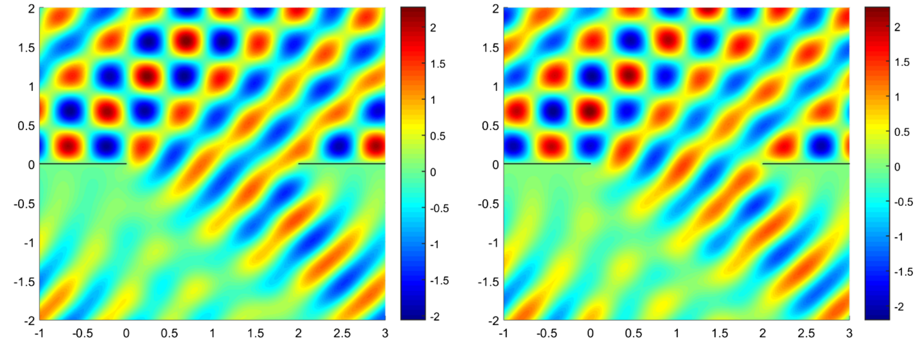

Figure 6 Example 3. (left) Real part of total field u_t, (right) Imaginary part of total field u_t

results of our Galerkin method with 20 dofs (degree of freedoms) are coincident with the results of FDM with 64 dofs, i.e., our Galerkin method needs fewer dofs than the FDM.

Example 3 We consider the case of wave number $k = 10$. The dimension of V_N is $N = 50$. The incident wave is plane wave with the incident angle $\theta = \pi/4$. Results are presented in Figs. 5–6. From these results, we can see that the Galerkin method works well and needs fewer dofs than the FDM to obtain reliable results.

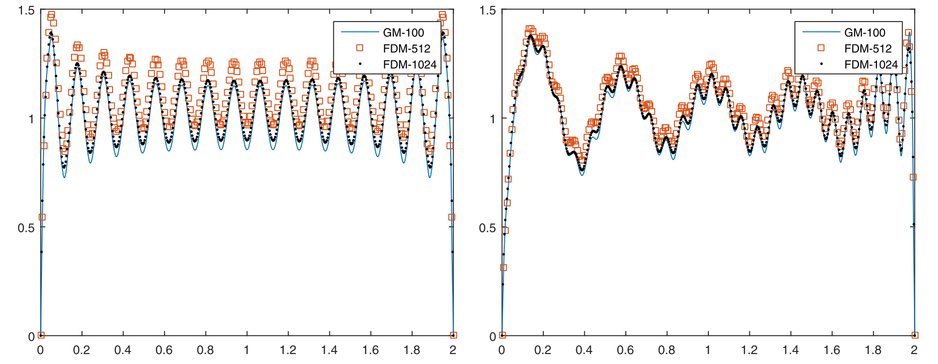

Figure 7 Example 4. (left) Absolute value of total field u_t on Γ_0 with $k = 50$ and $\theta = 0$, (right) Absolute value of total field u_t on Γ_0 with $k = 50$ and $\theta = \pi/4$

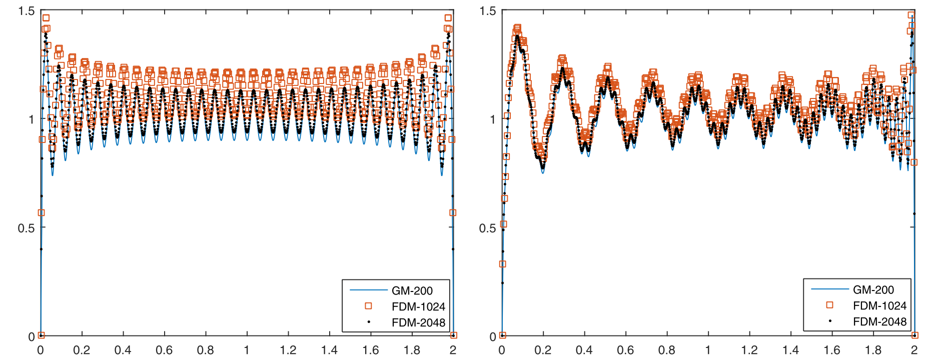

Figure 8 Example 4. (left) Absolute value of total field u_t on Γ_0 with $k = 100$ and $\theta = 0$, (right) Absolute value of total field u_t on Γ_0 with $k = 100$ and $\theta = \pi/4$

Example 4 In this example, we consider the situation of large wave number $k = 50$ and 100. For $k = 50$, we set $N = 100$. The absolute values of total fields u_t on Γ_0 are presented in Fig. 7 with incident angles $\theta = 0$ and $\theta = \pi/4$, respectively. For $k = 100$, we set $N = 200$. The absolute values of total fields u_t on Γ_0 are presented in Fig. 8 with incident angles $\theta = 0$ and $\theta = \pi/4$, respectively. Also the numerical solutions given by the FDM are shown in these figures. When $k = 50$, compared with the FDM which needs 1024 dofs, our Galerkin method needs only 100 dofs to get reliable computational results. When $k = 100$, the FDM needs 2048 dofs, while our Galerkin method needs only 200 dofs.

6 Conclusion

In this paper, we study the scattering problem of strip gratings. By use of the continuity of the total field across the slit in one period and the Dirichlet to Neumann map, this problem is reformulated to an operator equation on the slit. The well-posedness of the solution to the operator equation is proved and Galerkin method is employed to solve this operator equation. We also derive the error estimate for the Galerkin method and numerical examples show that our method is effective.

Acknowledgements
The authors would like to thank the referees for their valuable suggestions which helped to improve this work.

Funding
This work is supported by the Fundamental Research Funds for the Central Universities (Grant no. 3132017053, 3132018226), the Doctoral Scientific Research Foundation of Liaoning Province (Grant no. 20170520289), TianYuan Special Funds of the National Natural Science Foundation of China (Grant no. 11626054), and the National Natural Science Foundation of China (Grant no. 11601056).

Abbreviations
MoM, method of moments; FMM, Fourier modal method; CBCM, combined boundary conditions method; FDM, finite difference method; FEM, finite element method; DtN, Dirichlet to Neumann; dofs, degree of freedoms.

Availability of data and materials
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
The main idea of this paper was proposed by EZ, the theoretical analysis was carried out by the two authors together, and the numerical experiments were conducted by EZ. Both authors read and approved the final manuscript.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 9 August 2018 Accepted: 29 January 2019 Published online: 13 February 2019

References

1. Petit, R.: Electromagnetic Theory of Gratings, vol. 22. Springer, Heidelberg (1980)
2. Bao, G., Cowsar, L., Masters, W.: Mathematical Modeling in Optical Science, vol. 22. SIAM, Philadelphia (2001)
3. Wilcox, C.H.: Scattering Theory for Diffraction Gratings. Applied Mathematical Sciences Ser., vol. 46. Springer, New York (1984)
4. Uchida, K., Nora, T., Matsunaga, T.: Electromagnetic wave scattering by an infinite plane metallic grating in case of oblique incidence and arbitrary polarization. IEEE Trans. Antennas Propag. **36**(3), 415–422 (1988)
5. Wu, T.K.: Fast converging integral equation solution of strip gratings on dielectric substrate. IEEE Trans. Antennas Propag. **35**(2), 205–207 (1987)
6. Lee, C.W., Son, H.: Analysis of electromagnetic scattering by periodic strip grating on a grounded dielectric/magnetic slab for arbitrary plane wave incidence case. IEEE Trans. Antennas Propag. **47**(9), 1386–1392 (1999)
7. Florencio, R., Boix, R.R., Encinar, J.A.: Enhanced MoM analysis of the scattering by periodic strip gratings in multilayered substrates. IEEE Trans. Antennas Propag. **61**(10), 5088–5099 (2013)
8. Matsushima, A., Itakura, T.: Singular integral equation approach to plane wave diffraction by an infinite strip grating at oblique incidence. J. Electromagn. Waves Appl. **4**(6), 505–519 (1990)
9. Montiel, F., Neviere, M.: Electromagnetic study of the diffraction of light by a mask used in photolithography. Opt. Commun. **101**(3–4), 151–156 (1993)
10. Montiel, F., Neviere, M.: Perfectly conducting gratings: a new approach using infinitely thin strips. Opt. Commun. **144**(1–3), 82–88 (1997)
11. Guizal, B., Felbacq, D.: Electromagnetic beam diffraction by a finite strip grating. Opt. Commun. **165**(1–3), 1–6 (1999)
12. Granet, G., Guizal, B.: Analysis of strip gratings using a parametric modal method by Fourier expansions. Opt. Commun. **255**(1–3), 1–11 (2005)
13. Elamine, H., Guizal, B., Oueslati, M., Gharbi, T.: New reformulation of the Fourier modal method with spatial adaptive resolution for multilayered metallic strip grating. Opt. Commun. **283**(21), 4392–4396 (2010)
14. Li, L.: Fourier modal method. In: Popov, E. (ed.) Gratings: Theory and Numeric Applications, 2nd edn., pp. 1–40. Aix Marseille Université, Marseille (2014)
15. Wang, Y., Ma, F., Zheng, E.: Galerkin method for the scattering problem of a slit. J. Sci. Comput. **70**(1), 1–18 (2017)
16. Chandler-Wilde, S.N., Hewett, D.P., Moiola, A.: Interpolation of Hilbert and Sobolev spaces: quantitative estimates and counterexamples. Mathematika **61**(2), 414–443 (2015)