Received: 1 September 2022 | Revised: 23 January 2024 | Accepted: 27 September 2024

DOI: 10.1002/mma.10548

RESEARCH ARTICLE

WILEY

A nonuniform mesh method in the Floquet parameter domain for wave scattering by periodic surfaces

Tilo Arens[1] | **Ruming Zhang**[2]

[1]Institute for Applied and Numerical mathematics, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

[2]Institute for Mathematics, Technische Universität Berlin, Berlin, Germany

Correspondence
Ruming Zhang, Institute for Mathematics, Technische Universität Berlin, Berlin, Germany.
Email: ruming.zhang@tu-berlin.de

Communicated by: R. Rodriguez

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: 258734477

In this paper, we propose a new numerical method to simulate acoustic scattering problems in two-dimensional periodic structures with non-periodic incident fields. Applying the Floquet-Bloch transform to the scattering problem yields a family of quasi-periodic boundary value problems dependent on the Floquet-Bloch parameter. Consequently, the solution of the original scattering problem is written as the inverse Floquet-Bloch transform of the solutions to these boundary value problems. The key step in our method is the numerical approximation of this integral transform by a quadrature rule with a nonuniform choice of quadrature points adapted to the regularity of the family of quasi-periodic solutions. This achieved by a graded subdivision of the full interval for the Floquet-Bloch parameter and applying a Gauss-Legrendre quadrature rule on each subinterval. We prove that the numerical method converges exponentially with respect to both the number of subintervals and the number of Gaussian quadrature points. Some numerical experiments are provided to illustrate the results.

KEYWORDS

Dirichlet-to-Neumann map, nonuniform meshes, periodic surface, scattering theory, the Floquet-Bloch transform

MSC CLASSIFICATION

65D30, 65N30, 35P25, 78A45

1 | INTRODUCTION

Wave propagating in periodic structures has been a challenging and interesting topic in both theoretical analysis and numerical simulations for the past decades. For quasi-periodic incident fields, there is a well-established framework (see [1–16] for 2D and 3D acoustic, elastic, and electromagnetic waves) to reduce the problem to a bounded domain. However, when the incident field is non-periodic, this approach no longer works. It is required to develop much more sophisticated tools to solve the scattering problem in this case.

One possible approach is to apply techniques available for rough surface scattering. The basis is provided by an analysis of variational formulations for such problems in weighted Sobolev spaces [17, 18]. Possible numerical approaches include the integral equation method as studied in [19–24] for the Helmholtz equation in two dimensions and [25] for the full Maxwell system in three dimensions. Alternatively, the finite section method also provides a convergent algorithm to approximate the original problem by a bounded one. See [17, 20, 26] for its applications in both boundary integral equations and finite element methods.

The principal drawback of this approach is the loss of all information and structure relying on the periodic nature of the scatterer. A powerful tool to exploit such structure is provided by the Floquet-Bloch transform, and efficient numerical methods have been developed based on this transform.

For example, penetrable periodic media are considered in [27, 28]. For scattering by periodic surfaces, we refer to [29–31]. Note that the approach has also been extended to the three-dimensional bi-periodic surfaces in [32]. The method is proved to be convergent for 2D cases in [30, 31], but for 3D cases, convergence proofs are available only for some special situations [32].

After application of the Floquet-Bloch transform, the problem is reduced to solving a family of fully quasi-periodic problems for a range of the Floquet parameter. The solution of the original problem is obtained by inverting this transform. Although, numerically, this amounts to the approximate evaluation of an integral, it nevertheless proves to be a challenging task due to the presence of singularities.

Based on a detailed study of the regularity of the integrand, a high order method has been developed for 2D cases in [33], achieving any algebraic order of convergence. However, this scheme cannot easily be extended to the 3D case due to the complicated structure of the singularities. This has motivated the research presented in this paper, in which we design a new nonuniform mesh method for the 2D case, which is much more readily extendable to 3D cases.

Based on the singularities of the quasi-periodic solution with respect to the Floquet parameter, we first generate graded meshes in the integration interval and then apply a Gaussian quadrature rule in each subinterval. The convergence analysis is based on bounds for analytic extensions of the solutions of the quasi-periodic problems with respect to the Floquet parameter. We prove exponential convergence with respect to both the number of graded mesh points and the number of Gaussian quadrature points in each subinterval. These estimates are then coupled to error-estimates for the finite element method used to solve each quasi-periodic problem. Finally, we give some numerical examples to illustrate our theoretical results.

The paper is organized as follows. The mathematical model is introduced in Section 2, and the Floquet-Bloch transform is reviewed in Section 3. Then we discuss the analytic extension of quasi-periodic solutions with respect to the Floquet parameter in Section 4. In Section 5, nonuniform meshes are designed for definite integrals with square root singularities. With the results in Section 4, we prove exponential convergence of the numerical method, and also give some numerical experiments in the last section.

2 | MATHEMATICAL MODEL OF SCATTERING PROBLEMS

We consider the propagation of time-harmonic waves in a two-dimensional domain Ω bounded from below by a curve Γ given as the graph of a 2π-periodic function ζ, that is,

$$\Omega = \{(x_1, x_2) : x_1 \in \mathbb{R},\ x_2 > \zeta(x_1)\}, \qquad \Gamma = \{(x_1, \zeta(x_1)) : x_1 \in \mathbb{R}\}.$$

The total field u is assumed to satisfy the Helmholtz equation for some positive wave number k,

$$\Delta u + k^2 u = 0 \qquad \text{in } \Omega, \tag{1}$$

and to satisfy a Dirichlet boundary condition,

$$u = 0 \qquad \text{on } \Gamma. \tag{2}$$

As is usual, the total field is split into the given incident field and the unknown scattered field, $u = u^i + u^s$. A suitable radiation condition must be imposed on the scattered field to ensure uniqueness and existence of solution, and this requires some additional definitions. For detailed derivations and proofs of the statements made below, we refer to [17, 18].

Let $H > \max_{t\in\mathbb{R}}\{\zeta(t)\}$ and $\Gamma_H := \mathbb{R} \times \{H\}$ be a straight horizontal line above Γ. Let Ω_H be the periodic strip between Γ and Γ_H. For a visualization of the geometric setting, we refer to Figure 1.

The appropriate spaces for solutions of such a rough surface scattering problem are horizontally weighted Sobolev spaces [17]. Let $H^s(\Omega_H)$ and $H^s_{loc}(\Omega_H)$ denote the standard Sobolev spaces with real exponent s. For any $r \in \mathbb{R}$, let the weighted space $H^s_R(\Omega_H)$ be defined by

$$H^s_r(\Omega_H) := \left\{ \varphi \in H^s_{loc}(\Omega_H) : \left(1 + x_1^2\right)^{r/s} \varphi(x_1, x_2) \in H^s(\Omega_H) \right\}.$$

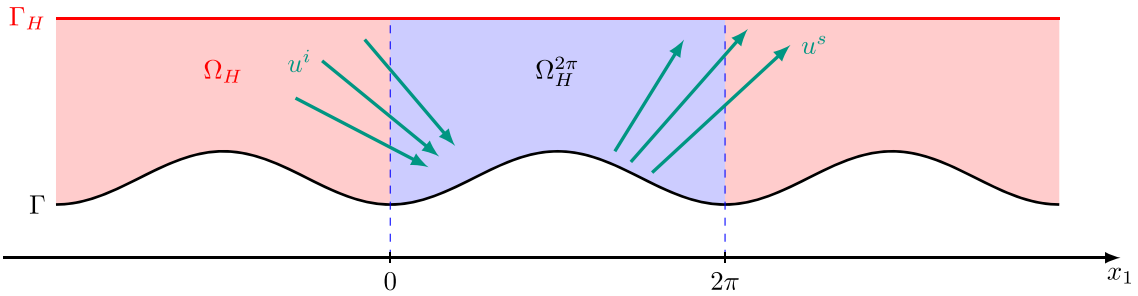

FIGURE 1 The full scattering problem is set in the periodic domain Ω_H. Using the Floquet-Bloch transform, the problem is reduced to the bounded domain $\Omega_H^{2\pi}$. [Colour figure can be viewed at wileyonlinelibrary.com]

To accommodate the Dirichlet boundary condition, we also define

$$\tilde{H}_r^1(\Omega_H) := \left\{ \varphi \in H_r^1(\Omega_H) : \varphi|_\Gamma = 0 \right\}.$$

To guarantee that the scattered field u^s propagates upwards, we require that u^s satisfies the following radiation condition:

$$u^s(x_1, x_2) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{ix_1\xi + i\sqrt{k^2-\xi^2}(x_2-H)} \hat{u}^s(\xi, H) \, d\xi, \quad x_2 > H, \tag{3}$$

where the square root takes non-negative real- and imaginary parts, $\hat{u}^s(\xi, H)$ is the Fourier transform of $u^s(\cdot, H)$. This condition is also known as the *spectral amplitude representation* and formally expresses the scattered field as a linear superposition of plane upward-propagating and evanescent waves. Detailed arguments on why and in what sense the expression on the right hand side of (3) makes sense for $u^s \in H_r^1(\Omega_H)$ for all $|r| < 1$ are given in [17]. Combining (3) with the Neumann trace operator on Γ_H gives rise to the Dirichlet-to-Neumann map $T^+ : H_r^{1/2}(\Gamma_H) \to H_r^{-1/2}(\Gamma_H)$,

$$(T^+\varphi)(x_1) = i \int_{\mathbb{R}} \sqrt{k^2 - \xi^2} \, e^{ix_1\xi} \, \hat{\varphi}(\xi) \, d\xi, \quad \text{where} \quad \varphi(x_1) = \int_{\mathbb{R}} e^{ix_1\xi} \, \hat{\varphi}(\xi) \, d\xi.$$

With these definitions, we are able to formulate the scattering problem under consideration: Given an incident field $u^i \in H_r^1(\Omega_H)$, where $|r| < 1$, find $u \in \tilde{H}_r^1(\Omega_H)$ such that u satisfies the Helmholtz equation (1) in Ω_H in the weak sense and the boundary condition

$$\frac{\partial u}{\partial x_2} - T^+ u = \frac{\partial u^i}{\partial x_2} - T^+ u^i =: f \quad \text{in} \quad H_r^{-1/2}(\Gamma_H).. \tag{4}$$

Explicitly, the weak form of this scattering problem is the following variational formulation: Given $f \in H_r^{-1/2}(\Gamma_H)$, where $|r| < 1$, find $u \in \tilde{H}_r^1(\Omega_H)$ such that

$$\int_{\Omega_H} \left[\nabla u \cdot \nabla \bar{\phi} - k^2 u \bar{\phi} \right] dx - \int_{\Gamma_H} T^+ \left(u|_{\Gamma_H} \right) \bar{\varphi} \, ds = \int_{\Gamma_H} f \bar{\varphi} \, ds \tag{5}$$

for all $\varphi \in \tilde{H}_{-r}^1(\Omega_H)$.

Theorem 1 ([17]). *For any $|r| < 1$, given $f \in H_r^{-1/2}(\Gamma_H)$, there is a unique solution $u \in \tilde{H}_r^1(\Omega_H)$ of the problem (5).*

Remark 2. In this paper, we only consider scattering problems with Dirichlet boundary condition on Γ. However, the method can also extended to other boundary conditions (e.g., impedance boundary conditions) or penetrable inhomogeneous media, when Theorem 1 still holds.

3 | FLOQUET-BLOCH TRANSFORMED FIELD AND THE REGULARITY

In this section, we recall the definition and some properties of the Floquet-Bloch transform and apply it to the solution of the original problem (5). We also present regularity results for the transformed field. For details, we refer to [33, 34].

3.1 | The Floquet-Bloch transform

The Floquet-Bloch transform of $\varphi \in C_0^\infty(\Omega_H)$ is defined as

$$(\mathcal{J}\varphi)(\alpha, x) = \sum_{j\in\mathbb{Z}} \varphi(x_1 + 2\pi j, x_2) e^{-\mathrm{i}\alpha(x_1+2\pi j)},$$

where $\alpha \in (-1/2, 1/2]$ (called Floquet-Bloch parameter) and $x \in \Omega_H^{2\pi} := \Omega_H \cap [-\pi, \pi]\times\mathbb{R}$ (see Figure 1). The domain $\Gamma_H^{2\pi}$ is defined in a similar way. As φ has a compact support, the transform is well-defined for $\alpha \in (-1/2, 1/2]$ and $x \in \Omega_H^{2\pi}$. It is also easy to check that when α is fixed, $(\mathcal{J}\varphi)(\alpha, \cdot)$ is 2π-periodic in x_1, that is,

$$(\mathcal{J}\varphi)\left(\alpha, \begin{pmatrix} x_1 + 2\pi \\ x_2 \end{pmatrix}\right) = (\mathcal{J}\varphi)(\alpha, x).$$

Moreover, $e^{\mathrm{i}\alpha x_1} w(\alpha, x)$ is 1-periodic in α for fixed x.

To introduce properties of the Floquet-Bloch transform, we define the space $H^m\left((-1/2, 1/2]; H^s\left(\Omega_H^{2\pi}\right)\right)$ $(m \in \mathbb{N})$ equipped with the norm:

$$\|\varphi\|_{H^m\left((-1/2,1/2];H^s(\Omega_H^{2\pi})\right)} := \left[\sum_{\ell=0}^{m} \int_{-1/2}^{1/2} \left\|\partial_\alpha^\ell \phi(\alpha, \cdot)\right\|^2_{H^s(\Omega_H^{2\pi})} \mathrm{d}\alpha\right]^2.$$

This definition is extended to any real number r by interpolation and duality arguments. We also define the subspace $H^r\left((-1/2, 1/2]; H^s_{\mathrm{per}}\left(\Omega_H^{2\pi}\right)\right)$ that contains functions which are 2π-periodic with respect to x_1 with fixed α. We conclude our overview of the properties of the Floquet-Bloch transform with the following theorem.

Theorem 3 (Theorem 8, [34]). *The transform $\mathcal{J}$ is extended to an isomorphism between $H^s_r(\Omega_H)$ and $H^r\left((-1/2, 1/2]; H^s_{per}(\Omega_H^\Lambda)\right)$ for any $s, r \in \mathbb{R}$.*

$$\left(\mathcal{J}^{-1} w\right)(x) = \int_{-1/2}^{1/2} w(\alpha, x) e^{\mathrm{i}\alpha x_1} \mathrm{d}\alpha, \ x \in \Omega_H.$$

When $s = r = 0$, $\mathcal{J}$ is an isometry with its inverse and $\mathcal{J}^{-1} = \mathcal{J}^*$.

3.2 | Floquet-Bloch transformed field and regularity

Following the process in [31, 34], we apply the Floquet-Bloch transform to the total field u, then $w(\alpha, x) := (\mathcal{J}u)(\alpha, x)$ satisfies the following variational equation with the test function $\psi(\alpha, x) = (\mathcal{J}\varphi)(\alpha, x)$ (see eq. (29) in [31]):

$$\int_{-1/2}^{1/2} a_\alpha(w(\alpha, \cdot), \psi(\alpha, \cdot)) \mathrm{d}\alpha = \int_{-1/2}^{1/2} \int_{\Gamma_H^{2\pi}} F(\alpha, \cdot)\bar{\psi}(\alpha, \cdot) \mathrm{d}s \mathrm{d}\alpha. \tag{6}$$

where

$$a_\alpha(\xi, \eta) = \int_{\Omega_H^{2\pi}} \left[\nabla\xi \cdot \nabla\bar{\eta} - 2\mathrm{i}\alpha \frac{\partial \xi}{\partial x_1}\bar{\eta} + \left(\alpha^2 - k^2\right)\xi\bar{\eta}\right] \mathrm{d}x - \int_{\Gamma_H^{2\pi}} T_\alpha^+ \left[\xi|_{\Gamma_H^{2\pi}}\right] \bar{\eta} \mathrm{d}s,$$
$$F(\alpha, x) = (\mathcal{J}f)(\alpha, x).$$

and T_α^+ is the periodic Dirichlet-to-Neumann map with index α from $H^{1/2}_{\text{per}}\left(\Gamma_H^{2\pi}\right)$ to $H^{-1/2}_{\text{per}}\left(\Gamma_H^{2\pi}\right)$. It has the following form:

$$(T_\alpha^+\varphi)(x_1) = \mathrm{i}\sum_{j\in\mathbb{Z}}\sqrt{k^2-(\alpha+j)^2}\hat{\varphi}(j)e^{\mathrm{i}jx_1} \qquad \text{where } \varphi(x_1)=\sum_{j\in\mathbb{Z}}\hat{\varphi}(j)e^{\mathrm{i}jx_1}.$$

From the equivalence between (5) and (6), we have the following results. For proofs, we refer to [30, 31, 34].

Theorem 4. *Given $f \in H_r^{-1/2}(\Gamma_H)$ for $|r| < 1$, the variational problem (6) is uniquely solvable in $H^r\left((-1/2,1/2];\tilde{H}^1_{per}\left(\Omega_H^{2\pi}\right)\right)$. Moreover,*

1. *When $f \in H_r^{1/2}(\Gamma_H)$ and $\zeta \in C^{2,1}$, $w \in H^r\left((-1/2,1/2];\tilde{H}^2_{per}\left(\Omega_H^{2\pi}\right)\right)$ and $u \in H_r^2(\Omega_H)$;*
2. *When $f \in H_r^{-1/2}(\Gamma_H)$ for $r \in (1/2,1)$, then $w \in L^2\left((-1/2,1/2];\tilde{H}^1_{per}\left(\Omega_H^{2\pi}\right)\right)$ equivalently satisfies*

$$a_\alpha(w(\alpha,\cdot),\varphi) = \int_{\Gamma_H^{2\pi}} F(\alpha,\cdot)\bar{\varphi}\,\mathrm{d}s \tag{7}$$

for any $\alpha \in (-1/2,1/2]$ and $\varphi \in \tilde{H}^1_{per}\left(\Omega_H^{2\pi}\right)$.

In particular for numerical applications, it is important to have a more detailed understanding of the regularity properties of $w(\alpha,x)$ with respect to the Floquet parameter α. Before the study of these properties, we first introduce the following notations and spaces. For any fixed positive wavenumber k, let

$$\underline{k} := \min\{|n-k| \,:\, n\in\mathbb{Z}\}.$$

From 1-periodicity of $e^{\mathrm{i}\alpha x_1}w(\alpha,x)$ with respect to α, the inverse Floquet-Bloch transform (see Theorem 3) applied to w can be written equivalently as

$$\left(\mathcal{J}^{-1}w\right)(x) = \int_{-\underline{k}}^{1-\underline{k}} w(\alpha,x)e^{\mathrm{i}\alpha x_1}\,\mathrm{d}\alpha,\ x\in\Omega_H.$$

Let

$$E := \left\{\alpha\in\left[-\underline{k},1-\underline{k}\right] \,:\, |n-\alpha| = k \ \text{ for some } \ n\in\mathbb{Z}\right\}.$$

Then. from direct calculation,

$$E = \begin{cases} \{-\underline{k},1-\underline{k}\}, & \text{when } \ k = n/2 \ \text{ for some } \ n\in\mathbb{N}_+;\\ \{-\underline{k},\underline{k},1-\underline{k}\}, & \text{otherwise.}\end{cases}$$

This implies that E contains the boundary of $\left[-\underline{k},1-\underline{k}\right]$ and may also contain a point in the interior this interval.

The formulation of the regularity properties of the Floquet-Bloch transformed field requires some appropriate function spaces. Let $\mathcal{I}\subset\mathbb{R}$ denote a bounded open interval, $D\subset\mathbb{R}^2$ a bounded domain, and $S(D)$ a Sobolev space independent of α of functions defined in D. First, define a space of functions that depend analytically on α:

$$C^\omega(\mathcal{I},S(D)) := \Bigg\{ \varphi\in\mathcal{D}'(\mathcal{I}\times D) \,:\, \forall\alpha_0\in\mathcal{I},\ \exists\delta>0,\ s.t.,\ \forall\alpha\in(\alpha_0-\delta,\alpha_0+\delta)\cap\mathcal{I},$$
$$\exists C>0,\ \varphi_n\in S(D),\ s.t.,\ \varphi(\alpha,x)=\sum_{n=0}^{\infty}(\alpha-\alpha_0)^n\varphi_n(x),\ \|\varphi_n\|_{S(D)}\le C^n\Bigg\},$$

where $\mathcal{D}'(\mathcal{I} \times D)$ is the space of distributions on $\mathcal{I} \times D$. Also, let the subspace of functions that are C^n-continuous with respect to α be defined as:

$$C^n(\mathcal{I}, S(D)) := \left\{ \varphi \in \mathcal{D}'(\mathcal{I} \times D) \,:\, \forall \alpha \in \mathcal{I},\, j = 0, 1, \ldots, n,\, \frac{\partial^j \varphi(\alpha, \cdot)}{\partial \alpha^j} \in S(D), \right.$$
$$\left. \text{moreover,} \quad \left\| \frac{\partial^j \varphi(\alpha, \cdot)}{\partial \alpha^j} \right\|_{S(D)} \text{ is uniformly bounded for } \alpha \in \mathcal{I} \right\}.$$

The regularity of the Floquet-Bloch transformed field can be characterized through two properties that we here formulate for a function $\varphi \in C^0(\mathcal{I}; S(D))$:

1. For any subinterval $\mathcal{I}_0 \subset \mathcal{I} \backslash E$, $\varphi \in C^\omega(\mathcal{I}_0; S(D))$.
2. For any $\alpha_0 \in \mathcal{I} \cap E$, there is a sufficiently small $\delta > 0$ and a pair $\varphi_1, \varphi_2 \in C^\omega(\mathcal{I}_0; S(D))$ such that

$$\varphi(\alpha, \cdot) = \varphi_1(\alpha, \cdot) + \sqrt{\alpha - \alpha_0}\, \varphi_2(\alpha, \cdot),$$

where $\mathcal{I}_0 = (\alpha_0 - \delta, \alpha_0 + \delta) \cap \mathcal{I}$.

The space of functions that satisfies both these properties will be denoted as

$$\mathcal{A}^\omega(\mathcal{I}; S(D); E) := \left\{ \varphi \in C^0(\mathcal{I}; S(D)) \,:\, \varphi \text{ satisfies condition 1 and 2} \right\}.$$

With the help of all these definitions, the regularity of the Floquet-Bloch transformed field $w(\alpha, x)$ can now be stated in the following theorem. For a proof, we refer to Theorem 16 in [33].

Theorem 5. *Given any $f \in H_r^{-1/2}(\Gamma_H)$ such that $\mathcal{J} f = F \in \mathcal{B}^\omega \left(\left(-\underline{k}, 1 - \underline{k}\right]; H_{per}^{-1/2}\left(\Gamma_H^{2\pi}\right); E \right)$ where $r \in (1/2, 1)$, then the transformed solution $\mathcal{J} u = w \in \mathcal{A}^\omega \left(\left(-\underline{k}, 1 - \underline{k}\right]; \tilde{H}_{per}^1\left(\Omega_H^{2\pi}\right); E \right)$. Moreover, if $f \in H_r^{1/2}(\Gamma_H)$ and $\zeta \in C^{2,1}(\mathbb{R})$, $w \in \mathcal{A}^\omega \left(\left(-\underline{k}, 1 - \underline{k}\right]; \tilde{H}_{per}^2\left(\Omega_H^{2\pi}\right); E \right)$.*

From Theorem 5, the transformed field $w(\alpha, \cdot)$ has only a finite number of square-root singularities. As $w(\alpha, \cdot)$ can be readily computed by well-established methods, the remaining difficulty is to approximate the inverse Bloch transform. Note that although in [33], one of the authors has proposed a highly efficient numerical method for the approximation, it is extremely difficult to extend this approach to scattering problems with bi-periodic structures in three dimensional space. With the ultimate goal of such an extension in mind, we introduce a nonuniform mesh method for the numerical approximation below. The estimation of quadrature errors arising in this method relies on bounds of analytic extensions of φ_1 and φ_2 in property 2 in the complex plane with respect to α. Thus, before the introduction to the adaptive method, we have to investigate the extension of the solutions to a neighborhood of $\left(-\underline{k}, 1 - \underline{k}\right]$ in $\mathbb{C}$.

4 | EXTENSION TO COMPLEX QUASI-PERIODICITIES

The convergence analysis of our method of inversion for the Floquet-Bloch transform relies on estimates for analytic extensions with respect to α of the solution $w(\alpha, \cdot)$ of the quasi-periodic problem. It is the goal of this section to characterize complex neighborhoods of the real axis to which $w(\alpha, \cdot)$ may be extended analytically and to provide estimates for these extensions. Our approach is, first, to precisely define analytic extensions of the variational formulation and of the corresponding operators and, second, to estimate the difference between these operators and their counterparts for real α for small deviations from the real axis. From these results, standard perturbation theory will yield analyticity of the solution $w(\alpha, \cdot)$ as well as the required bounds.

To simplify our estimates, let us slightly modify the spaces. Denote by V_H^{per} the space $\tilde{H}^1_{\text{per}}\left(\Omega_H^{2\pi}\right)$ with the norm replaced by

$$\|\varphi\|_{V_H^{\text{per}}} := \left[\int_{\Omega_H^{2\pi}} \left[|\nabla \phi|^2 + k^2 |\phi|^2 \right] \mathrm{d}x \right]^{1/2}.$$

Let the norm in $H^s_{\text{per}}\left(\Gamma_H^{2\pi}\right)$ be defined by

$$\|\varphi\|_{H^s_{\text{per}}(\Gamma_H^{2\pi})} = \left[\sum_{j\in\mathbb{Z}} (k^2+j^2)^s \left|\hat{\phi}_j\right|^2\right]^{1/2}.$$

As the sesquilinear form $a_\alpha(\cdot,\cdot)$ is well defined in $V_H^{\text{per}}\times V_H^{\text{per}}$, from Riesz's Lemma, there is a $\mathcal{B}_\alpha : V_H^{\text{per}} \to \left(V_H^{\text{per}}\right)^*$ such that

$$a_\alpha(\varphi,\psi) = \langle \mathcal{B}_\alpha\varphi, \psi\rangle,$$

where $\varphi,\psi \in V_H^{\text{per}}$ and $\langle\cdot,\cdot\rangle$ denotes the extension of the $L^2\left(\Omega_H^{2\pi}\right)$ inner product to the $\left(V_H^{\text{per}}\right)^*$-$V_H^{\text{per}}$ duality. Moreover, let $F(\alpha,\cdot) \in H^{-1/2}_{\text{per}}\left(\Gamma_H^{2\pi}\right)$.

Let δ denote a small complex number, and $\mathcal{D}_{\delta,\alpha}$ the perturbation of $\mathcal{B}_\alpha$ obtained from replacing α by $\alpha+\mathrm{i}\delta$, i.e.

$$\left\langle \left(\mathcal{B}_\alpha + \mathcal{D}_{\delta,\alpha}\right) v, \varphi\right\rangle = \int_{\Omega_H^{2\pi}} \left[\nabla v\cdot\nabla\bar{\phi} - 2\mathrm{i}\,(\alpha+\mathrm{i}\delta)\frac{\partial v}{\partial x_1}\bar{\phi} + ((\alpha+\mathrm{i}\delta)^2 - k^2) v\bar{\phi}\right] \mathrm{d}x - \int_{\Gamma_H^{2\pi}} T^+_{\alpha+\mathrm{i}\delta} v\bar{\varphi}\,\mathrm{d}x. \tag{8}$$

For the moment, we consider $T^+_{\alpha+\mathrm{i}\delta}$ as a formal symbol only, a precise definition of this operator will be given below. A direct calculation shows

$$\langle \mathcal{D}_{\delta,\alpha} v, \varphi\rangle = \int_{\Omega_H^{2\pi}} \left[2\delta\frac{\partial v}{\partial x_1}\bar{\phi} + (2\mathrm{i}\alpha\delta - \delta^2) v\bar{\phi}\right] \mathrm{d}x - \int_{\Gamma_H^{2\pi}} \left[T^+_{\alpha+\mathrm{i}\delta} - T^+_\alpha\right] v\bar{\varphi}\,\mathrm{d}s. \tag{9}$$

Obviously, the first integral depends analytically on δ in all of $\mathbb{C}$. Thus, we only need to investigate the analytic extension of the operator T^+_α which involves a countable number of functions with square-root singularities. As T^+_α is real analytic in $\left(-\underline{k}, 1-\underline{k}\right]$ except for the finite set E, we extend the operator also analytically in the neighbourhood of $\left(-\underline{k}, 1-\underline{k}\right]$ except for a finite number of vertical lines $\{\alpha_0\}\times\mathbb{R}$, where $\alpha_0 \in E$.

To this end, we redefine the square root operator "$\sqrt{}$" as follows.

Definition 6. For any $z \in \mathbb{C}\backslash\{0\}$, there is a unique representation such that

$$z = r\mathrm{e}^{\mathrm{i}\theta},\ r = |z| > 0,\ \theta \in \left(-\frac{\pi}{2}, \frac{3\pi}{2}\right].$$

Define $\sqrt{z} = \sqrt{r}\mathrm{e}^{\mathrm{i}\theta/2}$, where $\sqrt{r}$ denotes the usual square root for a positive real number. Moreover, when $z=0$, $\sqrt{z}=0$.

Via Definition 6, the square root function is analytically extended to the complex plane except for the negative imaginary axis. Thus for each term in the formal expression for $T^+_{\alpha+\mathrm{i}\delta}$, the map $\delta \mapsto \sqrt{k^2-(\alpha+j+\mathrm{i}\delta)^2}$ is real analytic in $\mathbb{R}$ when $|\alpha+j| \neq k$.

Consider now the analytic extension of the terms in the definition of the operator T^+_α. Let

$$A_1 = \left(-\underline{k}, \underline{k}\right), \qquad A_2 = \left(\underline{k}, 1-\underline{k}\right). \tag{10}$$

Note that if $\underline{k} = 0$, $A_1 = \varnothing$ while when $\underline{k} = 1/2$, $A_2 = \varnothing$, while neither is empty otherwise. The observations above show that for each $j\in\mathbb{Z}$, the function $\delta \mapsto \sqrt{k^2-(\alpha+j+\mathrm{i}\delta)^2}$ is analytic in the strips $A_m + \mathrm{i}\mathbb{R}$, $m = 1, 2$. We will show in Theorem 9 that the series over all these terms indeed converges and bound its difference from T^+_α. Before we can establish this result, however, we require two technical estimates for these square roots terms.

Lemma 7. *For any $\alpha \in A_m$, $m = 1, 2$, $\delta\in\mathbb{R}$ and $j\in\mathbb{Z}$,*

$$\left|\sqrt{(\alpha+j+\mathrm{i}\delta)^2-k^2} - \sqrt{(\alpha+j)^2-k^2}\right| \le \frac{|\delta|}{2\left|\sqrt{k^2-(\alpha+j)^2}\right|}\left(\max\{k, |\alpha+j|\} + \frac{|\delta|^2}{4\,|\alpha+j-k|}\right). \tag{11}$$

Proof. From Definition 6, we have for $x \in \mathbb{R}\backslash\{0\}$ and $y \in \mathbb{R}$ that

$$\left|\sqrt{x+\mathrm{i}y}+\sqrt{x}\right| = |\sqrt{x}|\left|1+\sqrt{1+\mathrm{i}\frac{y}{x}}\right| \geq 2|\sqrt{x}|.$$

Hence, by an elementary calculation, we obtain for $\alpha \in A_m$ that

$$\begin{aligned}
&\left|\sqrt{(\alpha+j+\mathrm{i}\delta)^2-k^2}-\sqrt{(\alpha+j)^2-k^2}\right| \\
\leq &\left|\sqrt{\alpha+j+\mathrm{i}\delta+k}-\sqrt{\alpha+j+k}\right|\left|\sqrt{\alpha+j+\mathrm{i}\delta-k}\right| \\
&+\left|\sqrt{\alpha+j+k}\right|\left|\sqrt{\alpha+j+\mathrm{i}\delta-k}-\sqrt{\alpha+j-k}\right| \\
= &\frac{|\delta|\left|\sqrt{\alpha+j+\mathrm{i}\delta-k}\right|}{\left|\sqrt{\alpha+j+\mathrm{i}\delta+k}+\sqrt{\alpha+j+k}\right|}+\frac{|\delta|\left|\sqrt{\alpha+j+k}\right|}{\left|\sqrt{\alpha+j+\mathrm{i}\delta-k}+\sqrt{\alpha+j-k}\right|} \\
\leq &\frac{|\delta|}{2}\left(\frac{\left|\sqrt{\alpha+j+\mathrm{i}\delta-k}\right|}{\left|\sqrt{\alpha+j+k}\right|}+\frac{\left|\sqrt{\alpha+j+k}\right|}{\left|\sqrt{\alpha+j-k}\right|}\right).x
\end{aligned}$$

We further estimate, again for $x \in \mathbb{R}\backslash\{0\}$ and $y \in \mathbb{R}$,

$$\left|\sqrt{x+\mathrm{i}y}\right| = (x^2+y^2)^{1/4} \leq |x|^{1/2}\left(1+\frac{y^2}{4x^2}\right)$$

Thus,

$$\left|\sqrt{\alpha+j+\mathrm{i}\delta-k}\right| \leq \left|\sqrt{\alpha+j-k}\right|+\frac{\delta^2}{4|\alpha+j-k|^{3/2}},$$

and the assertion follows. □

In the next step we further estimate the leading factor in (11) by estimating from below $\left|\sqrt{k^2-(\alpha+j)^2}\right|$ for $j \in \mathbb{Z}$, when α is fixed.

Lemma 8. *Let $\alpha \in A_m = (a_0, a_1)$ for $m = 1, 2$; for definitions, we refer to (10). Then*

$$\min_{j\in\mathbb{Z}}\left|\sqrt{k^2-(\alpha+j)^2}\right| \geq \sigma \min\left\{\sqrt{\alpha-a_0}, \sqrt{a_1-\alpha}\right\}$$

with $\sigma = 1$ if $k > 1/2$ and $\sigma = \sqrt{\underline{k}} > 0$ if $k \leq 1/2$.

Proof. Note first that all factors occurring on the right hand side of the asserted lower bound are less than or equal to 1.

Suppose that $k = \hat{j} + \underline{k}$, $\hat{j} \in \mathbb{Z}_{\geq 0}$ and write $j = \hat{j} + n$, $n \in \mathbb{Z}$. Then

$$\left|k^2-(\alpha+j)^2\right| = \left|(\hat{j}+\underline{k})^2-(\alpha+\hat{j}+n)^2\right| = \left|\underline{k}-\alpha-n\right|\left|\underline{k}+\alpha+n+2\hat{j}\right|.$$

First, let $\alpha \in A_1 = (-\underline{k}, \underline{k})$. Then,

$$\begin{aligned}
|\underline{k}-\alpha-n| &\geq \min\{|\underline{k}-\alpha|, |\underline{k}-\alpha-1| \geq \min\{|\underline{k}-\alpha|, |\underline{k}+\alpha|\}, \\
|\underline{k}+\alpha+n+2\hat{j}| &\geq \min\{|\underline{k}+\alpha|, |\underline{k}+\alpha-1|\} \geq \min\{|\underline{k}+\alpha|, |\underline{k}-\alpha|\},
\end{aligned}$$

as well as

$$|\underline{k}-\alpha-n|\geq 1,\ n\in\mathbb{Z}\backslash\{0,1\},\quad \text{and}\quad \left|(\underline{k}+\alpha+n+2\hat{j})\right|\geq 1,\ n+2\hat{j}\in\mathbb{Z}\backslash\{-1,0\}.$$

This proves the assertion for $\alpha\in A_1$ unless $n=\hat{j}=0$. In this case, we can obviously estimate

$$\left|\underline{k}-\alpha\right|\left|\underline{k}+\alpha\right|\geq \underline{k}\min\left\{|\underline{k}-\alpha|,|\underline{k}+\alpha|\right\}.$$

Taking the square root gives the estimate for $\alpha\in A_1$.

Now let $\alpha\in A_2=(\underline{k},1-\underline{k})$. In this case,

$$\begin{aligned}|\underline{k}-\alpha-n| &\geq \min\{|\underline{k}-\alpha|,|\underline{k}-\alpha+1|\}\geq\min\{|\underline{k}-\alpha|,|1-\underline{k}-\alpha|\},\\ |\underline{k}+\alpha+n+2\hat{j}| &\geq \min\{|\underline{k}+\alpha-1|,|\underline{k}+\alpha|\}\geq\min\{|\underline{k}+\alpha-1|,|\underline{k}-\alpha|\},\end{aligned}$$

as well as

$$|\underline{k}-\alpha-n|\geq 1,\ n\in\mathbb{Z}\backslash\{-1,0\},\qquad \text{and}\qquad |(\underline{k}+\alpha+n+2\hat{j})|\geq 1,\ n+2\hat{j}\in\mathbb{Z}\backslash\{-1,0\}.$$

The assertion is proven for $\alpha\in A_2$ unless $\hat{j}=0$ and $n\in\{-1,0\}$. In these exceptional cases, we have

$$\left|\underline{k}-\alpha\right|\left|\underline{k}+\alpha\right|\geq 2\underline{k}|\underline{k}-\alpha|,\qquad \left|\underline{k}-\alpha+1\right|\left|\underline{k}+\alpha-1\right|\geq 2\underline{k}|1-\underline{k}-\alpha|.$$

All arguments are repeated with obvious sign changes for $k=\hat{j}-\underline{k}$ □

With the previous two lemmas, we are now able to prove boundedness of $T^+_{\alpha+\mathrm{i}\delta}$ and bound the difference of the two DtN maps.

Theorem 9. *Let $\alpha\in A_m=(a_0,a_1)$, $m=1,2$; for definitions, we refer to (10). Let $\varrho=|\delta|/\min\left\{\sqrt{\alpha-a_0},\sqrt{a_1-\alpha}\right\}$. Then $T^+_{\alpha+\mathrm{i}\delta}:H^{1/2}_{per}\left(\Gamma^{2\pi}_H\right)\to H^{-1/2}_{per}\left(\Gamma^{2\pi}_H\right)$ is bounded for any $\delta\in\mathbb{R}$. Moreover, its difference from T^+_α is bounded by*

$$\left\|T^+_{\alpha+\mathrm{i}\delta}-T^+_\alpha\right\|\leq\frac{\varrho}{\sigma}+\frac{\varrho^3}{8\sigma k},$$

where σ is the constant defined in Lemma 8.

Proof. Consider $\varphi\in C^\infty_{\mathrm{per}}\left(\Gamma^{2\pi}_H\right)$ and its Fourier series, $\varphi(x_1)=\sum_{j\in\mathbb{Z}}\hat{\varphi}_j\mathrm{e}^{\mathrm{i}jx_1}$. Then

$$\left(T^+_{\alpha+\mathrm{i}\delta}-T^+_\alpha\right)\varphi=\mathrm{i}\sum_{j\in\mathbb{Z}}\left[\sqrt{k^2-(\alpha+j+1\delta)^2}-\sqrt{k^2-(\alpha+j)^2}\right]\hat{\varphi}_j\mathrm{e}^{\mathrm{i}(\alpha+j)x_1},$$

with convergence of the series ensured by the smoothness of φ. Hence,

$$\left\|\left(T^+_{\alpha+\mathrm{i}\delta}-T^+_\alpha\right)\varphi\right\|^2_{H^{-1/2}_{\mathrm{per}}(\Gamma^{2\pi}_H)}=\sum_{j\in\mathbb{Z}}\left(k^2+|j|^2\right)^{-1/2}\left|\sqrt{k^2-(\alpha+j+\mathrm{i}\delta)^2}-\sqrt{k^2-(\alpha+j)^2}\right|^2|\hat{\varphi}_j|^2.$$

Thus, by the previous two lemmas,

$$\begin{aligned}\left\|\left(T^+_{\alpha+\mathrm{i}\delta}-T^+_\alpha\right)\varphi\right\|_{H^{-1/2}_{\mathrm{per}}(\Gamma^{2\pi}_H)} &\le \sup_{j\in\mathbb{Z}}\frac{\left|\sqrt{k^2-(\alpha+j+\mathrm{i}\delta)^2}-\sqrt{k^2-(\alpha+j)^2}\right|}{\sqrt{k^2+|j|^2}}\,\|\varphi\|_{H^{1/2}_{\mathrm{per}}(\Gamma^{2\pi}_H)}\\ &\le \sup_{j\in\mathbb{Z}}\frac{|\delta|\left(\max\{k,|\alpha+j|\}+\frac{|\delta|^2}{4|\alpha+j-k|}\right)}{2\sigma\,\min\left\{\sqrt{|\alpha-a_0|},\sqrt{|\alpha-a_1|}\right\}}\,\frac{1}{\sqrt{k^2+|j|^2}}\,\|\varphi\|_{H^{1/2}_{\mathrm{per}}(\Gamma^{2\pi}_H)}\\ &\le \frac{|\delta|\left(1+\delta^2\sup\limits_{j\in\mathbb{Z}}\frac{1}{8\,|\alpha+j-k|\,\sqrt{k^2+|j|^2}}\right)}{\sigma\,\min\left\{\sqrt{|\alpha-a_0|},\sqrt{|\alpha-a_1|}\right\}}\,\|\varphi\|_{H^{1/2}_{\mathrm{per}}(\Gamma^{2\pi}_H)}.\end{aligned}$$

The proof is finished by observing $|\alpha+j-k|\ge\min\{\alpha-a_0,a_1-\alpha\}$. □

Our goal is to prove that the operator $\mathcal{B}_\alpha+\mathcal{D}_{\delta,\alpha}$ defined in (8) is boundedly invertible when δ is close to 0. To this end, for fixed α and δ *real,* we bound the operator $\mathcal{D}_{\delta,\alpha}$ with respect to δ. Theorem 9 provides an estimate for the second term in (9). It is also easily checked that

$$\left|\int_{\Omega^{2\pi}_H}\left[2\delta\frac{\partial v}{\partial x_1}\bar{\phi}+(2\mathrm{i}\alpha\delta-\delta^2)v\bar{\phi}\right]\mathrm{d}x\right|\le\left(4|\delta|+\delta^2/k\right)\|v\|_{V^{\mathrm{per}}_H}\|\varphi\|_{V^{\mathrm{per}}_H}. \tag{12}$$

As we are looking at small perturbations δ, we will assume that $|\delta|\le k$. Then, as a conclusion from (12) and Theorem 9, the norm of the operator $\mathcal{D}_{\delta,\alpha}$ is bounded by

$$\|\mathcal{D}_{\delta,\alpha}\|\le 5\,|\delta|+\frac{\varrho}{\sigma}+\frac{\varrho^3}{8\sigma k}.$$

In [18], explicit bounds for the inverse of the unperturbed operator $\mathcal{B}_\alpha$ are provided. It is shown in Theorem 4.1 of that reference that there exists a constant $M\ge 1$ with

$$\left\|\mathcal{B}_\alpha^{-1}\right\|\le M\qquad\text{for all }\ \alpha\in A_m.$$

Standard operator perturbation results hence show that $\mathcal{B}_\alpha+\mathcal{D}_{\delta,\alpha}$ is boundedly invertible when

$$5\,|\delta|+\frac{\varrho}{\sigma}+\frac{\varrho^3}{8\sigma k}<\frac{1}{M}. \tag{13}$$

In the following theorem, we provide sufficient conditions on δ to satisfy this inequality.

Theorem 10. *Let $\alpha\in A_m=(a_0,a_1)$, $m=1,2$; for definitions, we refer to (10). There exists a constant $C>0$ such that if*

$$|\delta|\le C\,\min\left\{\sqrt{\alpha-a_0},\sqrt{a_1-\alpha}\right\},$$

then $|\delta|\le k$ and the bound (13) is satisfied. Hence, the operator $\left(\mathcal{B}_\alpha+\mathcal{D}_{\delta,\alpha}\right)^{-1}$ and consequently also $w(\alpha+\mathrm{i}\delta,\cdot)$ depend analytically on $\alpha+\mathrm{i}\delta$ on the set

$$\mathbf{A}_m=\left\{\alpha+\mathrm{i}\delta\,:\,\alpha\in A_m,\ \delta\in\mathbb{R}\ \textit{with}\ |\delta|\le C\,\min\left\{\sqrt{\alpha-a_0},\sqrt{a_1-\alpha}\right\}\right\}.$$

Proof. Let $\mu=\max\limits_{\alpha\in A_m}\min\left\{\sqrt{\alpha-a_0},\sqrt{a_1-\alpha}\right\}$ and choose

$$C<\min\left\{\frac{k}{\mu},\frac{1}{M(5+\frac{1}{\sigma}+\frac{1}{8\sigma k})}\right\}.$$

Let $|\delta| \le C \min\left\{\sqrt{\alpha - a_0}, \sqrt{a_1 - \alpha}\right\}$. Then $|\delta| \le k$ and

$$\varrho = \frac{|\delta|}{\min\left\{\sqrt{\alpha - a_0}, \sqrt{a_1 - \alpha}\right\}} < \frac{1}{M\left(5 + \frac{1}{\sigma} + \frac{1}{8\sigma k}\right)} \le \frac{1}{M\left(5\min\left\{\sqrt{\alpha - a_0}, \sqrt{a_1 - \alpha}\right\} + \frac{1}{\sigma} + \frac{1}{8\sigma k}\right)}$$

as $\mu \le 1$. Note also $\varrho \le 1$ as $M \ge 1$ and $\sigma \le 1$. Now, we conclude

$$5|\delta| + \frac{\varrho}{\sigma} + \frac{\varrho^3}{8\sigma k} = \left(5\frac{|\delta|}{\varrho} + \frac{1}{\sigma} + \frac{\varrho^2}{8\sigma k}\right)\varrho < \frac{1}{M}.$$

Hence, (13) is satisfied. □

In this case, the set $\mathbf{A}_m$ is the intersection of the interior of two parabolas in the complex plane. For $A_m := (a_0, a_1)$, we can write

$$\mathbf{A}_m = \{\alpha + \mathrm{i}\delta : \alpha \ge a_0 + \delta^2/C^2 \text{ and } \alpha \le a_1 - \delta^2/C^2\}\backslash\{a_0, a_1\}$$

with the constant C from Theorem 10. In Section 5, we will require analytical extensions of $\alpha \mapsto w(\alpha, \cdot)$ to certain ellipses. Hence, we will now consider ellipses contained in $\mathbf{A}_m$.

Let us recall some basic definitions and properties of an ellipses. An ellipse with center at $(s, t) \in \mathbb{R}^2$ and half axes $a \ge b > 0$ parallel to the coordinate axes is defined as the set

$$\mathcal{E} = \left\{x \in \mathbb{R}^2 : \frac{(x_1 - s)^2}{a^2} + \frac{(x_2 - t)^2}{b^2} \le 1\right\},$$

The number $c := \sqrt{a^2 - b^2}$ is the called the linear eccentricity, $(-c + s, t)$ and $(c + s, t)$ are the foci and $(s - a, 0)$, $(s + a, 0)$ the vertices. By $\mathcal{E}^r_{c_1,c_2}$, we will denote the ellipse with foci at $(c_1, 0)$ and $(c_2, 0)$ and sum of the half-axes r; by $\widetilde{\mathcal{E}}^r_{a_1,a_2}$, we will denote the ellipse with vertices at $(a_1, 0)$ and $(a_2, 0)$ and sum of the half-axes r.

Lemma 11. *Let a and set $r = a + C\sqrt{a/2}$. Then*

$$\widetilde{\mathcal{E}}^r_{-a,a} \subseteq P = \left\{x \in \mathbb{R}^2 : x_1 \ge -a + x_2^2/C^2 \text{ and } x_1 \le a - x_2^2/C^2\right\}.$$

Proof. As both P and $\widetilde{\mathcal{E}}^r_{-a,a}$ are symmetric with respect to the x_2-axis, it is sufficient to consider case $x_1 \in [-a, 0]$ and the first inequality in the definition of P. For $x = (x_1, x_2) \in \widetilde{\mathcal{E}}^r_{-a,a}$ we have $x_2^2 \le C^2\frac{a^2 - x_1^2}{2a}$ and hence

$$-a + \frac{x_2^2}{C^2} \le -a + \frac{(a - x_1)(a + x_1)}{2a} \le -a + a + x_1 = x_1.$$

In the next lemma, we prove that a certain family of smaller ellipses are contained in $\widetilde{\mathcal{E}}^{a+b}_{-a,a}$. □

Lemma 12. *Let $a, b > 0$, $0 < \lambda \le a$, $0 < \mu \le \frac{b}{a}\lambda$ and $z \in [-a + \lambda, a - \lambda]$. Then $\widetilde{\mathcal{E}}^{\lambda+\mu}_{z-\lambda,z+\lambda} \subseteq \widetilde{\mathcal{E}}^{a+b}_{-a,a}$.*

Proof. Let $(z + \lambda\cos\theta, \mu\sin\theta) \in \partial\widetilde{\mathcal{E}}^{\lambda+\mu}_{z-\lambda,z+\lambda}$ where $\theta \in [0, 2\pi)$. As $z + \lambda\cos\theta \in [-a, a]$, there is a $\varphi \in [0, \pi)$ such that

$$z + \lambda\cos\theta = a\cos\varphi.$$

Thus $(a\cos\varphi, \pm b\sin\varphi) \in \partial\widetilde{\mathcal{E}}^{a+b}_{-a,a}$. We compare the squares of the x_2-coordinates.

$$\begin{aligned}\mu^2\sin^2\theta &\le \frac{b^2\lambda^2}{a^2}\sin^2\theta = \frac{b^2}{a^2}\left(\lambda^2 - \lambda^2\cos^2\theta\right)\\ &= \frac{b^2}{a^2}\left(\lambda^2 - (a\cos\varphi - z)^2\right) = b^2\sin^2\varphi + \frac{b^2}{a^2}\left(\lambda^2 - a^2 + 2az\cos\varphi - z^2\right)\\ &= b^2\sin^2\varphi + \frac{b^2}{a^2}\left(\lambda^2 - (a-z)^2 + 2az(\cos\varphi - 1)\right).\end{aligned}$$

As $\lambda \le a - z$, the second term is negative and we conclude $\mu^2\sin^2\theta \le b^2\sin^2\varphi$. We have shown that the boundary of $\widetilde{\mathcal{E}}^{\lambda+\mu}_{z-\lambda,z+\lambda}$ is in the interior of the ellipse $\widetilde{\mathcal{E}}^{a+b}_{-a,a}$ which proves the assertion. □

With the following corollary, we apply the results of Lemmas 11 and 12 to ellipses contained in the set $\mathbf{A}_m$.

Corollary 13. *Let $A_m = (a_0, a_1)$, $m = 1, 2$, and set $\hat{z} = (a_0 + a_1)/2$. Let $a \le |A_m|/2$ and set $b = C\sqrt{a/2}$, $r = a + b$. Choose λ and μ such that all assumptions of Lemma 12 are satisfied. Then $\widetilde{\mathcal{E}}^{\lambda+\mu}_{\hat{z}+z-\lambda,\hat{z}+z+\lambda} \subseteq \widetilde{\mathcal{E}}^{r}_{\hat{z}-a,\hat{z}+a} \subseteq \overline{\mathbf{A}_m}$.*

5 | NONUNIFORM MESHES FOR THE INVERSE FLOQUET-BLOCH TRANSFORM

In this section, we introduce a method based on nonuniform meshes for numerical integration of functions of one variable with a square root singularity. Later on, we extend the method to approximate the inverse Floquet-Bloch transform.

Let us start by considering the numerical approximation of the integral

$$I(\xi) := \int_0^h \xi(t)\,\mathrm{d}t, \tag{14}$$

where $\xi(t) = \xi_1(t) + \sqrt{t}\,\xi_2(t)$, $t \in [0, h]$ and both ξ_1 and ξ_2 are analytic in $[0, h]$.

Given a positive parameter $p \in (0, 1)$ and $N \in \mathbb{N}$, the method is described as follows. Let the nodal points be defined as

$$t_0 = 0, \qquad t_n = p^{n-1}h,\ n = 1, \dots, N+1.$$

These points are the end points of the subintervals:

$$J_0 = [t_0, t_{N+1}], \qquad J_n = [t_{n+1}, t_n],\ n = 1, \dots, N.$$

The integrand ξ is analytic in any J_n, $n = 1, \dots, N$, and we use an M-point Gauss-Legendre quadrature to approximate the integral on any such interval. On the interval J_0, the trapezoidal rule is used. Let the points and weights for the M-point Gauss-Legendre quadrature in $[-1, 1]$ be denoted by

$$\left\{(\tau_j, w_j) :\ j = 1, \dots, M\right\}.$$

For $n = 1, \dots, N$, the integral $I_n(\xi) := \int_{t_{n+1}}^{t_n} \xi(t)\,\mathrm{d}t$ is approximated by

$$I_n^M(\xi) = \frac{p^{n-1}h - p^n h}{2}\sum_{j=1}^{M} \xi\left(\frac{p^{n-1}h - p^n h}{2}\tau_j + \frac{p^{n-1}h + p^n h}{2}\right) w_j.$$

For $n = 0$, the integral $I_0(\xi) := \int_0^{p^N h} \xi(t)\,\mathrm{d}t$ is approximated by

$$I_0^N(\xi) = \frac{p^N h}{2}\xi(0) + \frac{p^N h}{2}\xi\left(p^N h\right).$$

Thus, the complete composite quadrature formula is

$$I_{N,M}(\xi) = \sum_{n=1}^{N} I_n^M(\xi) + I_N^0(\xi) = \frac{p^N h}{2}\xi(0) + \frac{p^N h}{2}\xi(p^N h) + \sum_{n=1}^{N}\left[\frac{p^{n-1}h - p^n h}{2}\sum_{j=1}^{M}\xi\left(\frac{p^{n-1}h - p^n h}{2}\tau_j + \frac{p^{n-1}h + p^n h}{2}\right)w_j\right]. \tag{15}$$

Our goal is to estimate the error of the approximation of $I(\xi)$ by $I_{N,M}(\xi)$. We first quote a derivative free error estimate for Gaussian quadrature; for details, we refer to [35]. We also recall our notation for ellipses from before Lemma 11.

Theorem 14 (Theorem 5.3.13, [35]). *Let $\varrho > 1/2$ and consider the ellipse $\mathcal{E}_{0,1}^{\varrho}$ as a subset of the complex plane. Let $\zeta : [0,1] \to \mathbb{C}$ be real analytic with complex analytic extension to $\mathcal{E}_{0,1}^{\varrho}$. Denote by I the integral over $(0,1)$ with integrand ζ and by Q_M its approximation by the M-point Gauss-Legendre quadrature. Then*

$$|I - Q_M| \le C(2\varrho)^{-2M} \max_{z\in\partial\mathcal{E}_{0,1}^{\varrho}} |\zeta(z)|.$$

The result can be extended to more general cases. Suppose ζ is an analytic function in $[\alpha,\beta]$ and let $\varrho > \frac{\beta-\alpha}{2}$. If ζ can be analytically extended to $\mathcal{E}_{\alpha,\beta}^{\varrho}$, then with analogous notation,

$$|I - Q_M| \le C\left(\frac{2\varrho}{\beta-\alpha}\right)^{-2M} \max_{z\in\partial\mathcal{E}_{\alpha,\beta}^{\varrho}} |\zeta(z)|. \tag{16}$$

We apply these results to estimating the error in our quadrature rule for each interval J_n, $n = 1, \dots, N$.

Lemma 15. *Suppose ξ can be analytically extended to an ellipse $\widetilde{\mathcal{E}}_{0,2h}^{\varrho h}$ with vertices 0, $2h$ and sum of semi-axis ϱh where $1 < \varrho < 2$. For any $n = 1, \dots, N$, there is a constant $C > 0$ such that*

$$\left|I_n(\xi) - I_n^M(\xi)\right| \le C\left(\min\left\{\frac{1+\sqrt{p}}{1-\sqrt{p}}, \sqrt{\frac{\varrho}{2-\varrho}}\right\}\right)^{-2M} \max_{z\in\partial\widetilde{\mathcal{E}}_{0,2h}^{\varrho h}} |\xi(z)|. \tag{17}$$

Proof. We wish to apply (16) and thus need to find the largest ellipse with foci at $p^n h$ and $p^{n-1}h$ that lies inside of $\widetilde{\mathcal{E}}_{0,2h}^{\varrho h}$. Denote the semi-axis by $\lambda > \mu$, respectively and note that the linear eccentricity is $c = \left[\frac{p^{n-1}-p^n}{2}\right]h$. Hence, we have the necessary condition:

$$\mu^2 + \left[\frac{p^{n-1}-p^n}{2}\right]^2 h^2 = \lambda^2.$$

We wish to apply Corollary 13, and in the notation there we have $\hat{z} = h$, $z = \left[\frac{p^{n-1}+p^n}{2}\right]h$, $a = h$ and $b = (\varrho-1)h$. The necessary conditions to apply the corollary hence are

$$0 < \lambda \le \left[\frac{p^{n-1}+p^n}{2}\right]h, \ \frac{\mu}{\lambda} \le \varrho - 1.$$

Our goal is to maximize $\lambda + \mu$ within these constraints. Note that for $\lambda > 0$, the line $\mu = (\varrho-1)\lambda$ intersects the hyperbola $\mu^2 + \left[\frac{p^{n-1}-p^n}{2}\right]^2 h^2 = \lambda^2$ in exactly one point $(\hat{\lambda}, \hat{\mu})$, where

$$\hat{\lambda} = \frac{1-p}{2\sqrt{2\varrho-\varrho^2}}p^{n-1}h, \qquad \hat{\lambda} + \hat{\mu} = \sqrt{\frac{\varrho}{2-\varrho}}\frac{1-p}{2}p^{n-1}h.$$

If $(1+p)/2p^{n-1}h < \hat{\lambda}$, we obtain the maximal value:

$$\lambda + \mu = \left(\frac{1+p}{2} + \sqrt{p}\right) p^{n-1}h = \frac{(1+\sqrt{p})^2}{2} p^{n-1}h.$$

Thus,

$$\lambda = \min\left\{\frac{1+p}{2},\ \frac{1-p}{2\sqrt{2\varrho - \varrho^2}}\right\} p^{n-1}h$$

and

$$\frac{\lambda+\mu}{c} = \min\left\{\frac{1+\sqrt{p}}{1-\sqrt{p}},\ \sqrt{\frac{\varrho}{2-\varrho}}\right\}.$$

Using (16) and the maximum modulus principle in complex analysis, we obtain

$$|I_n(\xi) - I_M^n(\xi)| \le C\left(\min\left\{\frac{1+\sqrt{p}}{1-\sqrt{p}},\ \sqrt{\frac{\varrho}{2-\varrho}}\right\}\right)^{-2M} \max_{z\in\partial\widetilde{\mathcal{E}}_{0,2h}^{\varrho h}} |\xi(z)|.$$

The proof is finished. □

We also estimate the error of the trapezoidal rule on J_0.

Lemma 16. *The error of the trapezoidal rule on J_0 is bounded by*

$$\left|I_0(\xi) - I_0^N(\xi)\right| \le C(p^N h)^{3/2}. \tag{18}$$

Proof. Recall the representation $\xi(t) = \xi_1(t) + \sqrt{t}\xi_2(t)$. It is a standard result that the error in approximating the integral over the analytic function ξ_1 by the trapezoidal rule is of order $\mathrm{O}\left((p^N h)^2\right)$. Thus, we only consider the second term, which we approximate by linear interpolation

$$\xi_{\mathrm{lin}}(t) = \frac{1}{\sqrt{p^N h}}\xi_2(p^N h)t.$$

For any $t \in J_0 = [0, p^N h]$,

$$\left|\sqrt{t}\xi_2(t) - \xi_{\mathrm{lin}}(t)\right| = \left|\sqrt{t}\xi_2(t) - \frac{\xi_2(p^N h)}{\sqrt{p^N h}}t\right| \le 2\sqrt{p^N h}\sup_{t\in J_0}|\xi_2(t)| \le C(p^N h)^{1/2}.$$

and the application of the trapezoidal rule can be estimated by

$$\left|\int_0^{p^N h}\left[\sqrt{t}\xi_2(t) - \xi_{\mathrm{lin}}(t)\right]\mathrm{d}t\right| \le C(p^N h)^{3/2}.$$

□

With Lemmas 16 and 15, we are now prepared to state an error estimate for the complete composite quadrature rule:

Theorem 17. *When N and M are two positive integers, there is a constant $C > 0$ such that*

$$\left|I(\xi) - I_{N,M}(\xi)\right| \le CN\left(\min\left\{\frac{1+\sqrt{p}}{1-\sqrt{p}},\ \sqrt{\frac{\varrho}{2-\varrho}}\right\}\right)^{-2M} + C(p^N h)^{3/2}. \tag{19}$$

Proof. Combine Lemmas 15 and 16. □

We conclude by applying the method introduced above to the approximation of the inverse Bloch transform,

$$(\mathcal{J}^{-1}w)(x) = \int_{-\underline{k}}^{1-\underline{k}} w(\alpha,x)\mathrm{e}^{\mathrm{i}\alpha x_1}\,\mathrm{d}\alpha,\ x\in\Omega_H^{2\pi}. \tag{20}$$

Depending on the different cases in the definition of E, this equation is

$$\left(\mathcal{J}^{-1}w\right)(x) = \begin{cases} \int_{-\underline{k}}^{1/2-\underline{k}} w(\alpha,x)\mathrm{e}^{\mathrm{i}\alpha x_1}\,\mathrm{d}\alpha + \int_{1/2-\underline{k}}^{1-\underline{k}} w(\alpha,x)\mathrm{e}^{\mathrm{i}\alpha x_1}\,\mathrm{d}\alpha\,, & \text{when } \underline{k}=0,\ \frac{1}{2}; \\ \int_{-\underline{k}}^{0} w(\alpha,x)\mathrm{e}^{\mathrm{i}\alpha x_1}\,\mathrm{d}\alpha + \int_{0}^{\underline{k}} w(\alpha,x)\mathrm{e}^{\mathrm{i}\alpha x_1}\,\mathrm{d}\alpha \\ \quad + \int_{\underline{k}}^{1/2} w(\alpha,x)\mathrm{e}^{\mathrm{i}\alpha x_1}\,\mathrm{d}\alpha + \int_{1/2}^{1-\underline{k}} w(\alpha,x)\mathrm{e}^{\mathrm{i}\alpha x_1}\,\mathrm{d}\alpha\,, & \text{otherwise.} \end{cases} \tag{21}$$

Note that in each interval, $w(\alpha,x)$ depends analytically on α except for a square root singularity at one edge point. From the definition of $\underline{k}$, the length of each interval is not larger than 1/2. With a change of variables, we can rewrite any integral in the form

$$\int_0^h \varphi(\alpha,x)\,\mathrm{d}\alpha\,,$$

where φ has the form

$$\varphi(\alpha,x) = \varphi_1(\alpha,x) + \sqrt{\alpha}\,\varphi_2(\alpha,x)$$

with $\varphi_1, \varphi_2 \in C^{\omega}([0,2h];S(D))$, where $S(D)$ can be any Sobolev space defined on $\Omega_H^{2\pi}$. From Theorem 10 and Corollary 13, we know that φ can be extended analytically to $\widetilde{\mathcal{E}}^r_{0,2h}$ with $r = h + C\sqrt{h/2}$. Thus, ϱ in Lemma 15 and Theorem 17 can be chosen as

$$\varrho = \min\left\{\frac{r}{h}, \frac{3}{2}\right\} = \min\left\{1 + \frac{C}{\sqrt{2h}}, \frac{3}{2}\right\}.$$

We redefine the integrals with new integrand as

$$I_0(\varphi)(x) = \int_0^{p^N h} \varphi(\alpha,x)\,\mathrm{d}\alpha, \quad I_n(\varphi)(x) = \int_{p^n h}^{p^{n-1}h} \varphi(\alpha,x)\,\mathrm{d}\alpha,\ n=1,2,\ \dots\ ,N.$$

The numerical approximations are

$$\begin{aligned} I_N^0(\varphi)(x) &= \frac{p^N h}{2}\,\varphi(0,x) + \frac{p^N h}{2}\,\varphi\left(p^N h, x\right); \\ I_M^n(\varphi)(x) &= \frac{p^{n-1}h - p^n h}{2} \sum_{j=1}^{M} \varphi\left(\frac{p^{n-1}h - p^n h}{2}\tau_j + \frac{p^{n-1}h + p^n h}{2}, x\right) w_j\,. \end{aligned}$$

We can now apply Theorem 17 to the approximation of any of the integrals in (21).

Theorem 18. *There exists constants $C>0$ and $\Theta>1$ such that*

$$\|I(\varphi) - I_{N,M}(\varphi)\|_{S(D)} \le C\left(N\Theta^{-2M} + (p^N h)^{3/2}\right). \tag{22}$$

Proof. Set

$$\Theta = \min\left\{\frac{1+\sqrt{p}}{1-\sqrt{p}},\ \sqrt{\frac{\sqrt{2h}+C}{\sqrt{2h}-C}},\ \sqrt{3}\right\} > 1.$$

From our choice of ρ and Theorem 17, the result follows. □

6 | NUMERICAL APPROXIMATION OF SCATTERING PROBLEMS

6.1 | Error estimation

In this section, we conclude our analysis by providing error estimates for the numerical solution of the original scattering problem (1)–(4). The algorithm can be divided into three steps:

Algorithm 19.

1. *Depending on k, find all the nodal points α_j and weights σ_j where $j = 1, 2, \dots, L$.*
2. *For any α_j, compute the numerical approximation of $w_\varepsilon(\alpha_j, x)$ of $u = \mathcal{I}(w)$, where $\varepsilon > 0$ is a parameter corresponding to the discretization (see below).*
3. *Compute $u_{N,M,\varepsilon}$ by the inverse Floquet-Bloch transform:*

$$u_{N,M,\varepsilon}(x) := \sum_{j=1}^{L} w_\varepsilon\left(\alpha_j, x\right) e^{\mathrm{i}\alpha_j x_1}\sigma_j.$$

In this algorithm, it remains to discuss the second step, that is, how to approximate $w(\alpha_j, x)$ numerically for any fixed α_j. In principle, this may be carried out by any preferred numerical method for solving a boundary value problem in a periodic domain such as the integral equation method or the finite element method. In the present work, we have chosen the latter approach.

Assume that $\mathcal{M}_\varepsilon$ is a family of regular, quasi-uniform triangular meshes in the finite domain $\Omega_H^{2\pi}$ with mesh width $0 < \varepsilon \le \varepsilon_0$, for some sufficiently small ε_0. For simplicity, we assume that the nodal points on the left and right boundaries have got identical x_2-coordinates. We omit all the nodal points on the left boundary by imposing periodic boundary conditions at $-\pi$ and π, as well as those on $\Gamma^{2\pi}$ due to the Dirichlet boundary conditions, and number the remaining nodes from 1 to M_0. Let $\psi^j_{M_0}$, $j = 1, 2, \dots, M_0$, denote the globally continuous function that is 2π-periodic with respect to x_1, linear on each mesh triangle and equals to 1 at nodal point j as well as to 0 at all other nodal points. We define the space spanned by these functions by

$$V_{\mathrm{per},\varepsilon} := \operatorname{span}\left\{\psi^j_{M_0}(x) : j = 1, 2, \dots, M_0\right\} \subset \tilde{H}^1_0\left(\Omega_H^{2\pi}\right).$$

For any fixed α, we have the following error estimate for the Galerkin approximation to the solution of (7). For details, we refer to [30, Theorem 14].

Theorem 20. *Suppose that $\zeta \in C^{1,1}(\mathbb{R})$. For any $\alpha \in W$, let $F(\alpha, \cdot) \in H^{1/2}_{per}\left(\Gamma_H^{2\pi}\right)$ and denote by $w(\alpha, \cdot)$ the solution of the variational equation (7). Then $w(\alpha, \cdot) \in H^2\left(\Omega_H^{2\pi}\right)$. Moreover, when $\varepsilon_0 > 0$ is sufficiently small and $w_\varepsilon(\alpha, \cdot) \in V_{per,\varepsilon}$ solves*

$$a_\alpha(w_\varepsilon(\alpha, \cdot), \varphi_\varepsilon) = \int_{\Gamma_H^{2\pi}} F(\alpha, \cdot)\overline{\varphi_\varepsilon}\, \mathrm{d}s \quad \text{for all } \varphi_\varepsilon \in V_{per,\varepsilon}, \tag{23}$$

then

$$\|w_\varepsilon(\alpha, \cdot) - w(\alpha, \cdot)\|_{H^\ell(\Omega_H^{2\pi})} \le C\varepsilon^{2-\ell}\|F(\alpha, \cdot)\|_{H^{1/2}_{per}(\Gamma_H^{2\pi})},\ \ell = 0, 1,$$

where C is independent of $\alpha \in W$.

From Theorems 18 and 20, we can immediately derive an error estimate for the solution computed using Algorithm 19.

Theorem 21. *Suppose that* $f \in H_r^{1/2}(\Omega_H)$ *with* $r \in (1/2, 1)$ *such that* $F(\alpha, x) := (\mathcal{J}f)(\alpha, x) \in \mathcal{A}^{\omega}\left(\left(-\underline{k}, 1-\underline{k}\right]; H_{per}^{1/2}\left(\Gamma_H^{2\pi}\right); E\right)$. *Then the error between numerical approximation* $u_{N,M,\varepsilon}$ *from Algorithm 19 and the exact solution u is bounded by*

$$\|u_{N,M,\varepsilon} - u\|_{H^{\ell}(\Omega_H^{2\pi})} \leq C\left[\epsilon^{2-\ell} + N\Theta^{-2M} + p^{3N/2}\right],\ \ell = 0, 1,$$

where $\Theta > 1$ *is defined as in Theorem 18 and C depends on* $\|f\|_{H_r^{1/2}(\Omega_H)}$, *k and p.*

Proof. We only present detailed arguments for the case that $\underline{k} = 0, 0.5$. In this case, the numerical integration is treated separately in two intervals, that is, $\left[1/2 - \underline{k}, -\underline{k}\right]$ and $\left[1 - \underline{k}, 1/2 - \underline{k}\right]$ (see (21)). For the other cases, the proof is carried out similarly. Let the nodes and weights in the ℓ-th interval ($\ell = 1, 2$) be denoted by α_m^{ℓ} and σ_m^{ℓ} where $m = 1, 2, \ldots, NM + 2$. Using the estimate from Theorem 18, we obtain

$$\begin{aligned}
\|u_{N,M,\varepsilon} - u\|_{H^{\ell}(\Omega_H^{2\pi})} \leq & \left\| \sum_{\ell=1}^{2} \sum_{m=1}^{NM+2} w_{\epsilon}\left(\alpha_m^{\ell}, \cdot\right) e^{\mathrm{i}\alpha_m^{\ell}()_1} \sigma_m^{\ell} - \sum_{\ell=1}^{2} \sum_{m=1}^{NM+2} w\left(\alpha_m^{\ell}, \cdot\right) e^{\mathrm{i}\alpha_m^{\ell}()_1} \sigma_m^{\ell} \right\|_{H^{\ell}(\Omega_H^{2\pi})} \\
& + \left\| \sum_{\ell=1}^{2} \sum_{m=1}^{NM+2} w\left(\alpha_m^{\ell}, \cdot\right) e^{\mathrm{i}\alpha_m^{\ell}()_1} \sigma_m^{\ell} - u \right\|_{H^{\ell}(\Omega_H^{2\pi})} \\
\leq & \sum_{\ell=1}^{2} \sum_{m=1}^{NM+2} \sigma_m^{\ell} \left\| w_{\epsilon}\left(\alpha_m^{\ell}, \cdot\right) - w\left(\alpha_m^{\ell}, \cdot\right) \right\|_{H^{\ell}(\Omega_H^{2\pi})} + CN\Theta^{-2M} + Cp^{3N/2} \\
\leq & C\epsilon^{2-\ell} \sum_{\ell=1}^{2} \sum_{m=1}^{NM+2} \sigma_m^{\ell} \|F(\alpha_m^{\ell}, \cdot)\|_{H_{\mathrm{per}}^{1/2}(\Gamma_H^{2\pi})} + CN\Theta^{-2M} + Cp^{3N/2},
\end{aligned}$$

where $()_1$ denotes the first coordinate of a two dimensional argument vector. As $f \in H_r^{1/2}\left(\Gamma_H^{2\pi}\right)$ for $r > 1/2$, by Theorem 5, we have $F = \mathcal{J}f \in H^r\left(\left(-\underline{k}, 1-\underline{k}\right]; H_{\mathrm{per}}^{1/2}\left(\Gamma_H^{2\pi}\right)\right)$, and

$$\|F\|_{H_0^r\left(\left(-\underline{k}, 1-\underline{k}\right]; H_{\mathrm{per}}^{1/2}\left(\Gamma_H^{2\pi}\right)\right)} = \|f\|_{H_r^{1/2}(\Gamma_H^{2\pi})}.$$

From Sobolev's embedding theorem, $F \in C^0\left(\left(-\underline{k}, 1-\underline{k}\right]; H_{\mathrm{per}}^{1/2}\left(\Gamma_H^{2\pi}\right)\right)$ and

$$\|F\|_{C^0\left(\left(-\underline{k}, 1-\underline{k}\right]; H_{\mathrm{per}}^{1/2}\left(\Gamma_H^{2\pi}\right)\right)} \leq C\|F\|_{H_0^r\left(\left(-\underline{k}, 1-\underline{k}\right]; H_{\mathrm{per}}^{1/2}\left(\Gamma_H^{2\pi}\right)\right)} = \|f\|_{H_r^{1/2}(\Gamma_H^{2\pi})}.$$

From the fact that $\sum_{\ell=1}^{2} \sum_{m=1}^{NM+2} \sigma_m^{\ell} = h$,

$$\|u_{N,M,\varepsilon} - u\|_{H^{\ell}(\Omega_H^{2\pi})} \leq C\epsilon^{2-\ell}\|f\|_{H_r^{1/2}(\Gamma_H^{2\pi})} + CN\Theta^{-2M} + Cp^{3N/2} \leq C\left[\epsilon^{2-\ell} + N\Theta^{-2M} + p^{3N/2}\right].$$

The proof is finished. □

6.2 | Numerical experiments

We present six numerical examples that demonstrate the convergence properties of Algorithm 19. In all of the examples, we use the same periodic surface given by

$$\zeta(t) = \frac{3}{2} + \frac{\sin t}{3} - \frac{\cos 2t}{4}.$$

The following parameters are also fixed:

$$H = 3,\ p = 0.5,\ h = 0.5.$$

We use two different wave numbers, $k = \sqrt{2}$ and $k = 10$, respectively. Note that when $k = \sqrt{2}, E = \left\{1-\sqrt{2}, \sqrt{2}-1, 2-\sqrt{2}\right\}$ whereas when $k = 10$, $E = \{0, 1\}$.

We also use three different incident fields,

$$u_1^i(k,x) = \Phi(x,a_1) - \Phi\left(x,a_1'\right);\ u_2^i(k,x) = \Phi(x,a_2) - \Phi\left(x,a_2'\right);\ u_3^i(k,x) = \int_{-\pi/2}^{\pi/2} e^{\mathrm{i}kx_1\sin t - \mathrm{i}kx_2\cos t} g(t)\,\mathrm{d}t.$$

Here, $\Phi(x,y) = \frac{\mathrm{i}}{4}H_0^{(1)}(k|x-y|)$ denotes the free space fundamental solution of the Helmholtz equation and $H_0^{(1)}(\cdot)$ is the Hankel function of the first kind of order 0. As source points, we use $a_1 = (0.4, 0.2)^\perp$ and $a_1' = (0.4, -0.2)^\perp$; $a_2 = (0.4, 3)^\perp$ and $a_2' = (0.4, -3)^\perp$. u_3^i is a downward propagating Herglotz wave function with the density function g defined as

$$g(t) = \begin{cases} \frac{(x-a)^6(x-b)^6}{((b-a)/2)^{12}}, & a < x < b; \\ 0, & \text{otherwise}; \end{cases}$$

with $a = 0.4$, $b = 0.5$.

With the definitions of the three incident fields, we apply Algorithm 19 to the following examples.

- **Example 1.** $k = \sqrt{2}$, $u^i(x) = u_1^i\left(\sqrt{2}, x\right)$.
- **Example 2.** $k = 10$, $u^i(x) = u_1^i(10, x)$.
- **Example 3.** $k = \sqrt{2}$, $u^i(x) = u_2^i\left(\sqrt{2}, x\right)$.
- **Example 4.** $k = 10$, $u^i(x) = u_2^i(10, x)$.
- **Example 5.** $k = \sqrt{2}$, $u^i = u_3^i\left(\sqrt{2}, x\right)$.
- **Example 6.** $k = 10$, $u^i(x) = u_3^i(10, x)$.

For all the examples, we collect the value of $u_{N,M,\varepsilon}$ on the line segment $\Gamma_h := [-\pi, \pi] \times \{2.9\}$ and study the dependence of errors on the parameters N, M and ε. Supposing we know the exact solution u_{exa} on Γ_h, we can compute the relative error defined by

$$err_{N,M,\varepsilon} := \frac{\left\|u_{N,M,\varepsilon} - u_{exa}\right\|_{L^2(\Gamma_h)}}{\|u_{exa}\|_{L^2(\Gamma_h)}}.$$

In Examples 1 and 2, since u^i is the half-space Green's function with source (0.4, 0.2) which lies below the periodic surface, u^i satisfies the radiation condition (3), that is, $f = 0$ in (4). Thus, we have to modify the problem (5). In this case, we are looking for a solution $u^s \in H_r^1(\Omega_H)$ such that

$$\int_{\Omega_H} \left[\nabla u^s \cdot \nabla \bar{\phi} - k^2 u^s \bar{\phi}\right] \mathrm{d}x - \int_{\Gamma_H} T^+\left(u|_{\Gamma_H}\right) \bar{\varphi}\,\mathrm{d}s = 0$$

with the boundary condition $u^s = -u^i$ on Γ. It is well known that the exact solution $u_{exa} = -u^i$ in Ω. For each example, we first fix sufficiently large M and N ($M = 10$, $N = 20$) and check the dependence of error on the finite element discretization; that is, for $\varepsilon = 0.04$, 0.02, 0.01, and 0.005, we compute the relative errors. The results are shown in Table 1 and are plotted on both logarithm scales in Figure 2a. Since the slopes of both curves are approximately 2, the convergence rates coincide with that shown in Theorem 21.

Next, we study the convergence of Algorithm 19 with respect to the parameters M and N. Fix $\varepsilon = 0.005$, and compute the relative errors with $M = 2, 3, 4, 5$ and $N = 4, 8, 12, 16, 20$ (also 24 when $k = \sqrt{2}$). The results are presented in Tables 2 and 3. In both tables, showing results for Examples 1 and 2, respectively, we clearly observe convergence of the numerical solution when N and M increase. When both these numbers are sufficiently large, the errors no longer decay, since the discretization error of the finite element method plays the more important role. The errors appear to be more sensitive with respect to the parameter N, since for $M \geq 3$ we observe that the errors almost only depend on N.

TABLE 1 Relative errors of Examples 1 and 2, with respect to ε.

k	$\varepsilon = 0.04$	$\varepsilon = 0.02$	$\varepsilon = 0.01$	$\varepsilon = 0.005$
$\sqrt{2}$	3.5×10^{-4}	8.8×10^{-5}	2.2×10^{-5}	6.2×10^{-6}
10	8.5×10^{-2}	2.2×10^{-2}	5.5×10^{-3}	1.4×10^{-3}

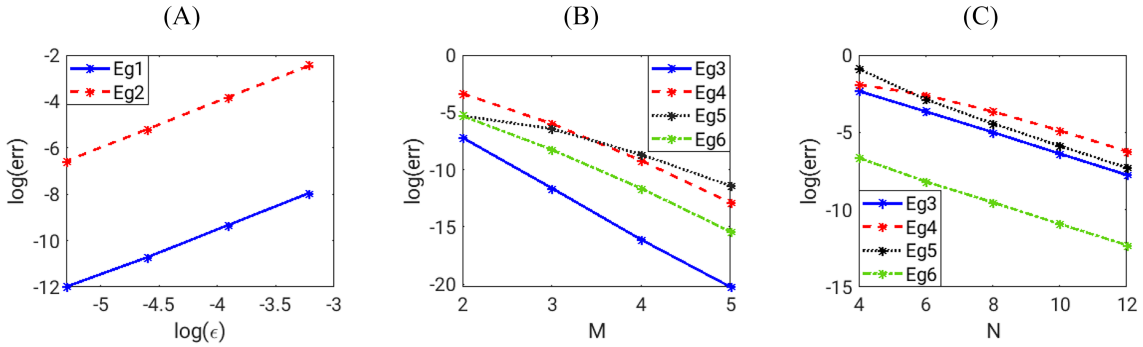

FIGURE 2 Dependence of relative errors on parameters (A) ε, (B) M, and (C) N. [Colour figure can be viewed at wileyonlinelibrary.com]

	$N=4$	$N=8$	$N=12$	$N=16$	$N=20$	$N=24$
$M=2$	6.7×10^{-2}	4.3×10^{-3}	3.2×10^{-4}	1.3×10^{-4}	1.3×10^{-4}	1.3×10^{-4}
$M=3$	6.7×10^{-2}	4.3×10^{-3}	2.7×10^{-4}	2.0×10^{-5}	6.2×10^{-6}	5.7×10^{-6}
$M=4$	6.7×10^{-2}	4.3×10^{-3}	2.7×10^{-4}	2.0×10^{-5}	6.2×10^{-6}	5.6×10^{-6}
$M=5$	6.7×10^{-2}	4.3×10^{-3}	2.7×10^{-4}	2.0×10^{-5}	6.2×10^{-6}	5.6×10^{-6}

TABLE 2 Relative errors of Example 1, with respect to M and N.

	$N=4$	$N=8$	$N=12$	$N=16$	$N=20$
$M=2$	8.3×10^{-2}	8.3×10^{-2}	8.3×10^{-2}	8.3×10^{-2}	8.3×10^{-2}
$M=3$	5.6×10^{-3}	5.9×10^{-3}	5.9×10^{-3}	5.9×10^{-3}	5.9×10^{-3}
$M=4$	1.9×10^{-3}	1.5×10^{-3}	1.5×10^{-3}	1.5×10^{-3}	1.5×10^{-3}
$M=5$	1.9×10^{-3}	1.4×10^{-3}	1.4×10^{-3}	1.4×10^{-3}	1.4×10^{-3}

TABLE 3 Relative errors of Example 2, with respect to M and N.

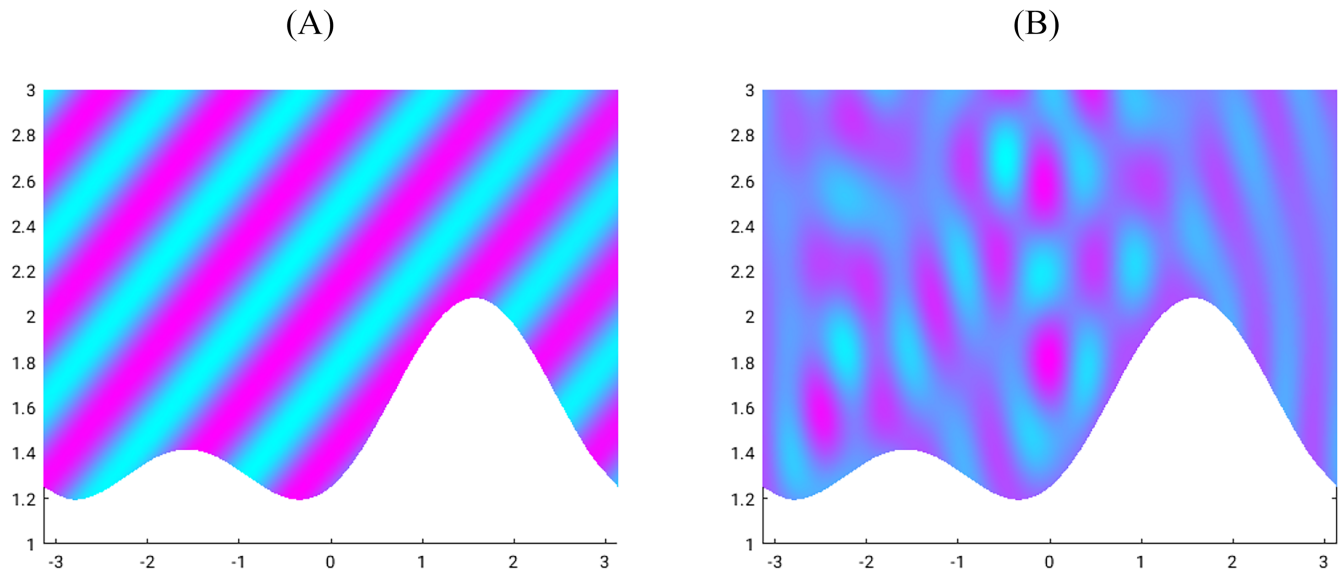

FIGURE 3 Real part of waves in Example 6: (A) incident field, (B) scattered field. [Colour figure can be viewed at wileyonlinelibrary.com]

In Examples 3 and 4, the incident fields are point sources located above the periodic surface, and in Examples 5 and 6, the incident fields are Herglotz wave functions propagating downwards. For all these examples, we no longer know the exact solutions. Instead, we choose parameters ($M = 10$, $N = 20$ and $\varepsilon = 0.005$) for which we expect the result to be sufficiently accurate and use the corresponding numerical solutions $u_{N,M,\varepsilon}$ as a reference solution instead of an exact solution. For a plot of the wave field in Example 6, we refer to Figure 3. For all examples, we first fix $N = 20$ and compute relative errors with $M = 2, 3, 4, 5$; in a second set of computations, we fix $M = 10$ and compute relative errors with $N = 4$, 6, 8, 10, and 12. The relative errors with respect to M are given in Table 4 and plotted in Figure 2b, while relative errors with respect to N are given in Table 5 and plotted in Figure 2c. From both graphs, we clearly observe the exponential convergence we expected from Theorem 21.

TABLE 4 Relative errors of Examples 3–6 with respect to M.

	$M = 2$	$M = 3$	$M = 4$	$M = 5$
Example 3	7.4×10^{-4}	9.0×10^{-6}	1.0×10^{-7}	1.6×10^{-9}
Example 4	3.5×10^{-2}	2.5×10^{-3}	1.0×10^{-4}	2.5×10^{-6}
Example 5	2.8×10^{-4}	2.6×10^{-5}	1.4×10^{-6}	2.2×10^{-7}
Example 6	1.6×10^{-5}	1.6×10^{-7}	2.8×10^{-9}	5.9×10^{-11}

TABLE 5 Relative errors of Examples 3–6 with respect to N.

	$N = 4$	$N = 6$	$N = 8$	$N = 10$	$N = 12$
Example 3	9.8×10^{-2}	2.6×10^{-2}	6.8×10^{-3}	1.7×10^{-3}	4.3×10^{-4}
Example 4	1.5×10^{-1}	7.5×10^{-2}	2.6×10^{-2}	7.6×10^{-3}	2.0×10^{-3}
Example 5	9.1×10^{-2}	2.3×10^{-2}	5.7×10^{-3}	1.4×10^{-3}	3.8×10^{-4}
Example 6	6.3×10^{-2}	1.6×10^{-2}	4.0×10^{-3}	9.9×10^{-4}	2.5×10^{-4}

At the end, we also discuss the convergence rates with respect to the parameters M and N. First, let us focus on the dependence of M. The slopes of the curves Figure 2b are approximately -4.4 (Example 3), -3.2 (Example 4), -2.1 (Example 5), and -3.4 (Example 6). Thus, exponential convergence is observed with respect to the parameter M. Similarly, slopes of the curves in Figure 2c are approximately -0.68 (Example 3), -0.55 (Example 4), -0.79 (Example 5), and -0.70 (Example 6), which show the exponential convergence with respect to the parameter N. The convergence results coincide with the theoretical results in Theorem 21.

7 | CONCLUSION

In this paper, a novel numerical method based nonuniform meshes in the Floquet-Bloch parameter domain is proposed to simulate scattering problems by periodic surfaces in two dimensional spaces. From the direct and inverse Floquet-Bloch transform, the problem can be written as an integral of a family of cell problems with respect to the Floquet-Bloch parameters. Based on a deeper understand of the analyticity of quasi-periodic solutions with respect to the Floquet-Bloch parameter, we design nonuniform meshes which are finer when they are close to the singular points, and coarser elsewhere. The Gaussian quadrature rule is adopted in coarser meshes and the trapezoidal rule is used in the fine meshes. The method is proved to converge exponentially with respect to the number of nodal points and the number of Gaussian quadrature rule.

As is mentioned, this method is expected to be extended to 3D cases. In a 3D problem, where an incident field is scattered by a bi-periodic surface, the singularities of the Floquet-Bloch transformed field lie on restrictions to a square of a finite number of circles (called singular curves), which are determined by the wave number. Thus, there is a potentially very complex structure of intersecting, touching or closely running curves of singularities. In future research, with similar ideas to the ones presented in this paper, we intend to generate non-uniform meshes which are finer near the singular curves, and coarser elsewhere. We then expect to be able to prove similar convergence for the discretization of the inverse transform as are presented here for the 2D case.

AUTHOR CONTRIBUTIONS

Tilo Arens: Conceptualization; formal analysis; methodology; writing—original draft. **Ruming Zhang:** Conceptualization; data curation; formal analysis; methodology; project administration; visualization.

ACKNOWLEDGEMENTS

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project-ID 258734477 – SFB 1173.

ORCID

Ruming Zhang https://orcid.org/0000-0003-2336-1020

REFERENCES

1. T. Abboud, *Electromagnetic waves in periodic media*, In Second International Conference on Mathematical and Numerical Aspects of Wave Propagation. SIAM, Philadelphia, Newark, DE, 1993, pp. 1–9.

2. T. Arens, *The scattering of plane elastic waves by a one-dimensional periodic surface*, Math. Meth. Appl. Sci. **22** (1999), 55–72.
3. T. Arens. (2010). *Scattering by biperiodic layered media: the integral equation approach*, Habilitation Thesis, Universität Karlsruhe.
4. G. Bao, *Diffractive optics in periodic structures: the TM polarization. Technical report, Institute for Mathematics and Its Applications*, University of Minnesota, Minneapolis, 1994.
5. G. Bao, *Finite element approximation of time harmonic waves in periodic structures*, SIAM J. Numer. Anal. **32** (1995), no. 4, 1155–1169.
6. G. Bao, *Numerical analysis of diffraction by periodic structures: TM polarization*, Numer. Math. **75** (1996), 1–16.
7. G. Bao, *Variational approximation of Maxwell's equations in biperiodic structures*, SIAM J. Appl. Math. **57** (1997), 364–381.
8. G. Bao and D. C. Dobson, *On the scattering by a biperiodic structure*, Proc. Amer. Math. Soc. **128** (2000), 2715–2723.
9. O. Bruno and F. Reitich, *Numerical solution of diffraction problems: a method of variation of boundaries III. Doubly-periodic gratings*, J. Opt. Soc. Amer. **10** (1993), 2551–2562.
10. D. C. Dobson, *A variational method for electromagnetic diffraction in biperiodic structures*, Math. Model. Numer. Anal. **28** (1994), 419–439.
11. D. C. Dobson and A. Friedman, *The time-harmonic Maxwell equations in biperiodic structures*, Math. Anal. Appl. **166** (1992), 507–528.
12. J. Elschner and G. Schmidt, *Diffraction of periodic structures and optimal design problems of binary gratings. Part I: direct problems and gradient formulas*, Math. Meth. Appl. Sci. **21** (1998), 1297–1342.
13. A. Kirsch, *Diffraction by periodic structures*, Proc. Lapland Conf. on Inverse Problems *Edited by* L. Pävarinta and E. Somersalo. Springer, 1993, pp. 87–102.
14. J.-C. Nédélec and F. Starling, *Integral equation methods in a quasi-periodic diffraction problem for the time-harmonic Maxwell's equations*, SIAM J. Math. Anal. **22** (1991), no. 6, 1679–1701.
15. G. Schmidt, *On the diffraction by biperiodic anisotropic structures*, Appl. Anal. **82** (2003), 75–92.
16. B. Strycharz, *An acoustic scattering problem for periodic, inhomogeneous media*, Math. Meth. Appl. Sci. **21** (1998), no. 10, 969–983.
17. S. N. Chandler-Wilde and J. Elschner, *Variational approach in weighted Sobolev spaces to scattering by unbounded rough surfaces*, SIAM J. Math. Anal. **42** (2010), 2554–2580.
18. S. N. Chandler-Wilde and P. Monk, *Existence, uniqueness, and variational methods for scattering by unbounded rough surfaces*, SIAM J. Math. Anal. **37** (2005), 598–618.
19. T. Arens, S. N. Chandler-Wilde, and J. A. DeSanto, *On integral equation and least squares methods for scattering by diffraction gratings*, Commun. Comput. Phys. **1** (2006), 1010–1042.
20. S. N. Chandler-Wilde, M. Rahman, and C. R. Ross, *A fast two-grid finite section method for a class of integral equations on the real line with application to an acoustic scattering problem in the half-plane*, Numer. Math. **93** (2002), 1–51.
21. K. Haseloh. (2004). *Second kind integral equations on the real line: solvability and numerical analysis in weighted spaces*, PhD thesis, Universität Hannover.
22. A. Lechleiter. (2008). *Factorization methods for photonics and rough surface scattering*, PhD thesis, Karlsruhe, Germany.
23. J. Li, G. Sun, and R. Zhang, *The numerical solution of scattering by infinite rough surfaces based on the integral equation method*, Comput. Math. Appl. **71** (2016), no. 7, 1491–1502.
24. A. Meier, T. Arens, S. N. Chandler-Wilde, and A. Kirsch, *A Nyström method for a class of integral equations on the real line with applications to scattering by diffraction gratings and rough surfaces*, J. Int. Equ. Appl. **12** (2000), 281–321.
25. P. Li, H. Wu, and W. Zheng, *Electromagnetic scattering by unbounded rough surfaces*, SIAM J. Math. Anal. **43** (2011), no. 3, 1205–1231.
26. A. Meier and S. N. Chandler-Wilde, *On the stability and convergence of the finite section method for integral equation formulations of rough surface scattering*, Math. Meth. Appl. Sci. **24** (2001), 209–232.
27. J. Coatléven, *Helmholtz equation in periodic media with a line defect*, J. Comp. Phys. **231** (2012), 1675–1704.
28. H. Haddar and T. P. Nguyen, *Volume integral method for solving scattering problems from locally perturbed periodic layers*, In WAVES 2015 Proceed, KIT, Karlsruhe, pp. 2015.
29. A. Lechleiter and D.-L. Nguyen, *Scattering of Herglotz waves from periodic structures and mapping properties of the Bloch transform*, Proc. Roy. Soc. Edinburgh Sect. A. **231** (2015), 1283–1311.
30. A. Lechleiter and R. Zhang, *A convergent numerical scheme for scattering of aperiodic waves from periodic surfaces based on the Floquet-Bloch transform*, SIAM J. Numer. Anal **55** (2017), no. 2, 713–736.
31. A. Lechleiter and R. Zhang, *A Floquet-Bloch transform based numerical method for scattering from locally perturbed periodic surfaces*, SIAM. J. Sci. Comput. **39** (2017), no. 5, B819–B839.
32. A. Lechleiter and R. Zhang, *Non-periodic acoustic and electromagnetic scattering from periodic structures in 3d*, Comput. Math. Appl. **74** (2017), no. 11, 2723–2738.
33. R. Zhang, *A high order numerical method for scattering from locally perturbed periodic surfaces*, SIAM J. Sci. Comput. **40** (2018), no. 4, A2286–A2314.
34. A. Lechleiter, *The Floquet-Bloch transform and scattering from locally perturbed periodic surfaces*, J. Math. Anal. Appl. **446** (2017), no. 1, 605–627.
35. S. Sauter and C. Schwab, *Boundary element methods*, Springer, Berlin-New York, 2007.

How to cite this article: T. Arens and R. Zhang, *A nonuniform mesh method in the Floquet parameter domain for wave scattering by periodic surfaces*, Math. Meth. Appl. Sci. (2024), 1–21, DOI 10.1002/mma.10548.