Convergent adaptive finite element methods for photonic crystal applications

STEFANO GIANI

School of Mathematical Sciences, University of Nottingham, University Park, Nottingham, NG7 2RD, UK

stefano.giani@nottingham.ac.uk

Abstract

We prove the convergence of an adaptive finite element method for computing the band structure of 2D periodic photonic crystals with or without compact defects in both the TM and TE polarization cases. These eigenvalue problems involve non-coercive elliptic operators with discontinuous coefficients. The error analysis extends the theory of convergence of adaptive methods for elliptic eigenvalue problems to photonic crystal problems, and in particular deals with various complications which arise essentially from the lack of coercivity of the elliptic operator with discontinuous coefficients. We prove the convergence of the adaptive method in an oscillation-free way and with no extra assumptions on the initial mesh, beside the conformity and shape regularity. Also we present and prove the convergence of an adaptive method to compute efficiently an entire band in the spectrum. This method is guaranteed to converge to the correct global maximum and minimum of the band, which is a very useful piece of information in practice. Our numerical results cover both the cases of periodic structures with and without compact defects.

1 Introduction

In the last years the question of convergence of adaptive methods for eigenvalue problems has received intensive interest and a number of convergence results have appeared. The first proofs to appear had some extra assumptions on the initial mesh and some extra marking strategies to control the oscillations [19, 20]. Then newer proofs appeared with no extra assumptions or oscillations strategies [7, 17, 18, 9]. The proof in this paper has been inspired by [18], which is only for elliptic eigenproblems based on coercive bilinear forms and which makes use of the green refinement strategy to adapt the mesh. As we shall see in Section 3 the sesquilinear form of the photonic crystals (PCs) eigenvalue problem (1.1) is not coercive for all values of the quasimomentum κ, so an extension of the analysis is required. The first convergence proof for PCs eigenvalue problems is in [19], however that proof could be considered quite dated now, since it possesses an extra assumption on the initial mesh and it is not oscillation-free. With this work we want to present a more up to date proof of convergence for PCs which is suitable for the red-green refinement strategy, which is probably the most widely used in practice.

PCs are constructed by assembling portions of periodic media composed of dielectric materials and they are designed to exhibit interesting properties in the propagation of electromagnetic waves, such as spectral band gaps. Media with band gaps have many potential applications, for example, in optical communications, filters, lasers, switches and optical transistors; e.g. see [25, 34, 1] for an introduction. In this paper we consider only 2D PCs, whose structure is periodic in a plane determined by two orthogonal directions and constant in the normal direction to that plane.

The propagation of light in any kind of PCs is governed by Maxwell's equations [28]. In 2D PCs the 3D Maxwell's equations reduce to a two-dimensional one-component wave equation, which

determines either the electric field (TH mode) or the magnetic field (TE mode). Because the problem is periodic, the Floquet transform [28, 27] can be applied to split each mode into a family of eigenvalue problems on the primitive cell Ω [2] of the periodic medium with periodic boundary conditions. This family is parameterized by the quasimomentum κ, which varies in the first Brillouin zone, see Section 2. All eigenvalue problems in the family have the weak form: seek eigenpairs of the form $(\lambda, u) \in \mathbb{C} \times \mathcal{H}$ such that

$$\int_\Omega ((\nabla + i\kappa)v)^* A(\nabla + i\kappa)u \, dx \;=\; \lambda \int_\Omega B u \bar{v} \, dx \quad \text{in } \Omega, \text{ for all } v \in \mathcal{H}, \tag{1.1}$$

where $\mathcal{H}$ is a Hilbert space that will be specified in Section 3, Ω is the primitive cell of the photonic crystal and u, v are required to satisfy periodic boundary conditions. Here, the matrix-valued function A is real symmetric and uniformly positive definite, i.e.,

$$\forall x \in \Omega, \quad 0 < \underline{a} \;\le\; \xi^* A(x) \xi \;\le\; \overline{a} \quad \text{for all} \quad \xi \in \mathbb{C}^2 \quad \text{with} \quad |\xi| = 1 \;, \tag{1.2}$$

where * denotes Hermitian transpose. The scalar function B is real and bounded above and below by positive constants for all $x \in \Omega$, i.e.,

$$0 \;<\; \underline{b} \;\le\; B(x) \;\le\; \overline{b} \quad \text{for all} \quad x \in \Omega. \tag{1.3}$$

In this work we will assume (as is generally the case in applications), that A and B are both piecewise constant on Ω and we will also assume that any jumps in A and B are aligned with the meshes used in this work. Due to the jumps of the coefficients, the eigenfunctions of (1.1) could have localized singularities in the gradient, which could diminish the rate of convergence of finite element methods on uniformly refined meshes.

A very popular practical numerical method for PCs is the Fourier spectral method (also called the "plane-wave expansion method") [33, 24, 6, 32, 31]. The overall rate of convergence for this method of the approximate spectra to the true spectra is slow because the jumps in the dielectric destroy the exponential accuracy which is achieved by Fourier spectral methods for smooth problems. Other spectral methods include expansions in terms of eigenfunctions for the crystal without any defects [13]. Semi-analytical methods which impose considerable limitations on the geometry of the crystal are also considered [14].

We use adaptive finite element methods because they provide flexible solvers for partial differential equation (PDE) problems and they are able to deal optimally with heterogeneous media problems. There are already some works about finite element methods for PCs, sometimes with adaptivity [3, 5, 10, 11, 23, 26, 36, 35, 12, 22]. However, until [22] a proper a posteriori analysis for PC problems was missing, as far as we know.

The outline of the paper is as follows. A brief §2 describes how problem (1.1) is derived form Maxwell's equations and an equally brief §3 contains some basic properties of the sesquilinear form in (1.1) and presents the a priori convergence estimates for finite element approximation of PC eigenvalue problems. These results have been already presented in [22], they are reported here only for clarity. In §4 some a posteriori results regarding the estimators are presented and in §5 the convergence proof of the adaptive finite element method is presented. In §6 an efficient and convergent method to compute an entire band is presented. This method is useful in practice since the maximum and the minimum of the computed approximation of the band converge to the true global maximum and minimum of the band, which are useful quantities to assess the performances of a PCs. Finally, numerical experiments illustrating the results with our method are collected in §7. These include both results on infinite periodic structures and on periodic structures with defect. We believe that the present paper is the first contribution to the topic of oscillation-free convergence of adaptive finite element methods for PC applications.

2 Photonic Crystals (PCs)

Two-dimensional PCs are interesting because they may have spectral band gaps, in other words, monochromatic electromagnetic waves of certain frequencies may not propagate inside them. Such crystals are much easier to fabricate than general 3D photonic crystals, while still allowing for many important applications. Theoretical analysis for 2D PCs is significantly simpler than for 3D photonic crystals, because a 2D PC dielectric system has two fundamental types of modes, E-polarised (TH mode) and H-polarised (TE mode). The propagation of a monochromatic beam of light of frequency ω inside a periodic medium of dielectric material is governed by Maxwell's equations (in the absence of free charges and currents):

$$\begin{aligned} \nabla \times \mathbf{E}_\omega &= -\tfrac{i\omega}{c}\mu \mathbf{H}_\omega, \quad \nabla \cdot \varepsilon \mathbf{E}_\omega = 0 \ , \\ \nabla \times \mathbf{H}_\omega &= \tfrac{i\omega}{c}\varepsilon \mathbf{E}_\omega \ , \quad \nabla \cdot \mu \mathbf{H}_\omega = 0. \end{aligned} \tag{2.1}$$

where $\mathbf{E}_\omega$ is the electric field, $\mathbf{H}_\omega$ is the magnetic field, ε and μ are the dielectric permittivity and magnetic permeability tensors and c is the speed of light in the vacuum. We will assume that the medium is "orthotropic", i.e., it has a periodic structure in a certain plane (here taken to be $x-y$) and is constant in the third (z) dimension. The tensor $\varepsilon = \varepsilon(x,y)$ then has the form

$$\varepsilon = \begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} & 0 \\ \varepsilon_{21} & \varepsilon_{22} & 0 \\ 0 & 0 & \varepsilon_{33} \end{pmatrix},$$

with $\varepsilon_{12} = \varepsilon_{21}$ and with $\det(\varepsilon) > 0$. In the rest of the work we assume the magnetic permeability μ constant and equal to 1, as done by other authors, e.g. see [10, 5, 28].

A 2D periodic medium can be described using lattices. Any basis of vectors $\{\mathbf{r}_1, \mathbf{r}_2\}$ for $\mathbb{R}^2$ generates a lattice $\ell := \{\mathbf{R} \in \mathbb{R}^2 : \ \mathbf{R} = n_1\mathbf{r}_1 + n_2\mathbf{r}_2 \ , \ n_1, n_2 \in \mathbb{Z}\}$. We may think of elements in ℓ equivalently as either vectors in $\mathbb{R}^2$ or as points in the 2D plane. Clearly ℓ is a group under vector addition, with the neutral element $\mathbf{0}$. The primitive cell (more precisely the Wigner-Seitz primitive cell [2]) for ℓ is defined to be the set Ω of all points in $\mathbb{R}^2$ which are closer to $\mathbf{0}$ than to any other point in ℓ. When $\overline{\Omega}$ is translated through all $\mathbf{R} \in \ell$, we obtain a covering of $\mathbb{R}^2$ with overlapping of measure 0.

The reciprocal lattice [2] for ℓ is the lattice $\hat{\ell}$ generated by a basis $\{\mathbf{k}_1, \mathbf{k}_2\}$, chosen to have the property

$$\mathbf{r}_i.\mathbf{k}_j = 2\pi\delta_{i,j} \ , \quad i,j = 1,2 \ . \tag{2.2}$$

Suitable formulae for $\{\mathbf{k}_1, \mathbf{k}_2\}$ are

$$\mathbf{k}_1 = \frac{2\pi S\mathbf{r}_2}{\mathbf{r}_1.(S\mathbf{r}_2)} \ , \quad \mathbf{k}_2 = \frac{2\pi S\mathbf{r}_1}{\mathbf{r}_2.(S\mathbf{r}_1)} \ , \quad \text{where} \quad S = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \ .$$

Clearly (2.2) implies that $\exp(i\mathbf{K}\cdot\mathbf{R}) = 1$ for all $\mathbf{R} \in \ell$ and all $\mathbf{K} \in \hat{\ell}$. The primitive cell for the reciprocal lattice is called the first Brillouin zone [2] which we denote here by $\mathcal{K}$.

For example, if ℓ is generated by $\{\mathbf{e}_1, \mathbf{e_2}\}$ (where $\mathbf{e}_i$ are the standard basis vectors in $\mathbb{R}^2$), then $\Omega = [-0.5, 0.5]^2$, $\hat{\ell}$ is generated by $\{2\pi\mathbf{e}_1, 2\pi\mathbf{e_2}\}$ and the first Brillouin zone is $\mathcal{K} = [-\pi, +\pi]^2$.

When ε is orthotropic the application of the Floquet transform splits each one of the polarized modes TE and TH into a direct sum of a family of scalar eigenvalue problems in the parameter $\kappa \in \mathcal{K}$ on the primitive cell Ω with periodic boundary conditions [28]:

$$(\nabla + i\kappa)\cdot(\nabla + i\kappa)u + \frac{\omega_\kappa^2}{c^2}\mathcal{B}u \ = \ 0 \ , \quad \text{where} \quad \mathcal{B} := \varepsilon_{33} \ , \tag{2.3}$$

for the so called TH-polarization and

$$(\nabla + i\kappa) \cdot \mathcal{A}(\nabla + i\kappa)u + \frac{\omega_\kappa^2}{c^2}u \;=\; 0 \;, \quad \text{where} \quad \mathcal{A} := \frac{1}{\varepsilon_{11}\varepsilon_{22} - \varepsilon_{12}^2} \begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} \\ \varepsilon_{21} & \varepsilon_{22} \end{pmatrix}, \tag{2.4}$$

for the so called TE-polarization. Both (2.3) and (2.4) may be written in the abstract form

$$(\nabla + i\kappa) \cdot A(\nabla + i\kappa)u + \lambda_\kappa B u \;=\; 0 \quad \text{on } \Omega, \quad \kappa \in \mathcal{K} \;, \tag{2.5}$$

which should be understood in the weak form.

3 Eigenvalue problems and a priori convergence results

The results in this section have already been presented in [22], but for sake of completeness they are presented again below. Throughout $L^2_B(\Omega)$ denotes the usual space of square integrable complex valued functions which we shall equip with the weighted norm

$$\|f\|_{0,B} \;=\; b(f,f)^{1/2} \;, \quad b(f,g) := \int_\Omega B f \bar{g}\, dx \;. \tag{3.1}$$

$H^1(\Omega)$ denotes the usual space of complex valued functions in $L^2(\Omega)$ with square integrable gradient and $H^1_\pi(\Omega)$ denotes the functions $f \in H^1(\Omega)$ which satisfy periodic boundary conditions on $\partial\Omega$, e.g. for $\Omega = [-0.5, 0.5]^2$ then

$$f((-0.5, y)) \;=\; f((0.5, y)) \;, \quad f((x, -0.5)) \;=\; f((x, 0.5)) \;, \quad \text{with } x, y \in [-0.5, 0.5] \;.$$

$H^1_\pi(\Omega)$ is equipped with the usual H^1-norm $\|f\|_1$. We will also use fraction spaces $H^{1+s}(\Omega)$, $s \in [0,1]$. Finally we denote by $\|\cdot\|_\infty$ the standard norm of $L^\infty(\Omega)$. On occasions we may want to restrict these norms to a measurable subset $\mathcal{S} \subseteq \Omega$, in which case we write $\|f\|_{0,B,\mathcal{S}}$, $\|f\|_{1,\mathcal{S}}$, etc.

Since $a_\kappa(\cdot,\cdot)$ is Hermitian, the eigenvalues are real and problem (1.1) can be rewritten as: *seek eigenpairs of the form* $(\lambda_j, u_j) \in \mathbb{R} \times H^1_\pi(\Omega)$ *such that*

$$\left.\begin{array}{rcll} a_\kappa(u_j, v) & = & \lambda_j \; b(u_j, v) \;, & \text{for all} \quad v \in H^1_\pi(\Omega) \\ \|u_j\|_{0,B} & = & 1 & \end{array}\right\} \tag{3.2}$$

where

$$a_\kappa(u,v) := \int_\Omega ((\nabla + i\kappa)v)^* A((\nabla + i\kappa)u)\, dx \;, \tag{3.3}$$

and

$$b(u,v) := \int_\Omega u \; v \, dx \;.$$

The sesquilinear for $a_\kappa(\cdot,\cdot)$ is bounded in $H^1_\pi(\Omega)$ and non-negative, [22, Lemma 3.1] i.e.,

$$a_\kappa(u,v) \;\le\; C_a \|u\|_1 \|v\|_1, \quad \text{for all } u, v \in H^1_\pi(\Omega) \quad \text{and} \quad a_\kappa(u,u) \ge 0 \;. \tag{3.4}$$

Moreover, since $a_\kappa(\cdot,\cdot)$ is Hermitian, the spectrum of (3.2) is real and it is also non-negative since $0 \;\le\; a_\kappa(u,u) \;=\; \lambda \; b(u,u) \;=\; \lambda$, for any solution of (3.2).

However, the analysis of (3.2) is complicated by the fact that the problem is not coercive for all values of κ and to deal with that, we introduce a new sesquilinear form related to $a_\kappa(\cdot,\cdot)$ which is coercive for all $\kappa \in \mathcal{K}$. The same approach has already been used in [22, 19].

Definition 3.1: We define the shifted sesquilinear form

$$(u,v)_{\kappa,A,B} \;:=\; a_\kappa(u,v) + \; \sigma \; b(u,v) \;, \quad \text{for all } u, v \in H^1_\pi(\Omega) \;,$$

where $\sigma = (\max_{\kappa\in\mathcal{K}} |\kappa|^2 \underline{a}/\underline{b}) + 1$. We also define $\|f\|_{\kappa,A,B} = (f,f)^{1/2}_{\kappa,A,B}$ on $H^1_\pi(\Omega)$. Consequently, we define the norm $\|u\|^2_{\kappa,A,B} \;=\; (u,u)_{\kappa,A,B}$.

In [22, Theorem 3.3] the coercivity of the shifted sesquilinear form is proved, i.e.,

$$\|u\|^2_{\kappa,A,B} \;=\; (u,u)_{\kappa,A,B} \;\geq\; c_a\|u\|^2_1, \quad \text{for all } \kappa \in \mathcal{K}, \quad u \in H^1_\pi(\Omega)\ , \tag{3.5}$$

where $c_a = \min\{\underline{a}/2, \underline{b}\}$ and also there it is proved that for any value of the quasimomentum $\kappa \in \mathcal{K}$, the sesquilinear form $(\cdot,\cdot)_{\kappa,A,B}$ is continuous with continuity constant $C_{a,b}$, which depends on $\overline{a}$, $\overline{b}$ and on the diameter of $\mathcal{K}$:

$$(u,v)_{\kappa,A,B} \;\leq\; C_{a,b}\|u\|_1\|v\|_1, \quad \text{for all } u,v \in H^1_\pi(\Omega). \tag{3.6}$$

Now we introduce the discrete version of (3.2). Let $\mathcal{T}_n\ , n = 1,2,\ldots$ denote a family of conforming and periodic triangular meshes on Ω. These meshes may be computed adaptively. We also assume that the meshes $\mathcal{T}_n$ are shape regular, i.e., there exists a constant $C_{\rm reg}$ independent of n such that

$$H_\tau \leq C_{\rm reg}\ \rho_\tau, \quad \text{for all} \quad \tau \in \mathcal{T}_n, \tag{3.7}$$

where H_τ is the diameter of element τ and ρ_τ is the diameter of its largest inscribed ball in the same element τ. We define

$$H_n^{\max} := \max_{\tau\in\mathcal{T}_n}\{H_\tau\}.$$

On any mesh $\mathcal{T}_n$ we denote by V_n^p the finite dimensional space of complex continuous functions which are affine polynomials of order less or equal p on each element $\tau \in \mathcal{T}_n$. For problem (3.2) the space $V_n^p \subset H^1_\pi(\Omega)$. The discrete formulation of problem (3.2) is: seek eigenpairs of the form $(\lambda_{j,n}, u_{j,n}) \in \mathbb{R} \times V_n^p$ such that

$$\left.\begin{array}{rcl} a_\kappa(u_{j,n}, v_n) &=& \lambda_{j,n}\ b(u_{j,n}, v_n)\ , \quad \text{for all} \quad v_n \in V_n^p \\ \|u_{j,n}\|_{0,B} &=& 1 \end{array}\right\} \tag{3.8}$$

We also introduce shifted versions of problems (3.2) and (3.8): seek eigenpairs of the form $(\zeta_j, u_j) \in \mathbb{R} \times H^1_\pi(\Omega)$ such that

$$\left.\begin{array}{rcl} a_\kappa(u_j, v) \;+\; \sigma\ b(u_j, v) &=& \zeta_j\ b(u_j, v)\ , \quad \text{for all} \quad v \in H^1_\pi(\Omega) \\ \|u_j\|_{0,B} &=& 1\ , \end{array}\right\} \tag{3.9}$$

Seek eigenpairs of the form $(\zeta_{j,n}, u_{j,n}) \in \mathbb{R} \times V_n^p$ *such that*

$$\left.\begin{array}{rcl} a_\kappa(u_{j,n}, v_n) \;+\; \sigma\ b(u_{j,n}, v_n) &=& \zeta_{j,n}\ b(u_{j,n}, v_n)\ , \quad \text{for all} \quad v_n \in V_n^p \\ \|u_{j,n}\|_{0,B} &=& 1\ . \end{array}\right\} \tag{3.10}$$

The following is self-evident:

Proposition 3.2: The eigenpairs of (3.2) and (3.9) are in one-one correspondence. In fact, (u_j, λ_j) is an eigenpair of (3.2) if and only if (u_j, ζ_j), with $\zeta_j = \lambda_j + \sigma$, is an eigenpair of (3.9). Similarly $(u_{j,n}, \lambda_{j,n})$ is an eigenpair of (3.8) if and only if $(u_{j,n}, \zeta_{j,n})$, with $\zeta_{j,n} = \lambda_{j,n} + \sigma$, is an eigenpair of (3.10).

It follows from (3.5) that all eigenvalues of (3.9) and all $N = \dim V_n^p$ eigenvalues of (3.10) are positive. We can order them as $0 < \zeta_1 \leq \zeta_2 \ldots$ and $0 < \zeta_{1,n} \leq \zeta_{2,n} \ldots \leq \zeta_{N,n}$. Moreover, we know (see [40, §6.3]) $\zeta_{j,n} \to \zeta_j$, for any j, as $H_n^{\max} \to 0$ and (by the minimum-maximum principle - see e.g. [40, §6.1]) that $\zeta_{j,n}$ is monotone decreasing, i.e.,

$$\zeta_{j,n} \;\geq\; \zeta_{j,m} \;\geq\; \zeta_j\ , \quad \text{for all} \quad j = 1,\ldots,N, \quad \text{and all} \quad m \geq n\ . \tag{3.11}$$

Now, by Proposition 3.2, it follows that $\lambda_{j,n} \to \lambda_j$, for any j, as $H_n^{\max} \to 0$ and $\lambda_{j,n}$ is monotone decreasing i.e.,

$$\lambda_{j,n} \;\geq\; \lambda_{j,m} \;\geq\; \lambda_j\ , \quad \text{for all} \quad j = 1,\ldots,N, \quad \text{and all} \quad m \geq n\ . \tag{3.12}$$

The distance of an approximate eigenfunction from the true eigenspace is a crucial quantity in the convergence analysis for eigenvalue problems especially in the case of non-simple eigenvalues.

Definition 3.3: Given a function $v \in L^2_B(\Omega)$ and a finite dimensional subspace $\mathcal{P} \subset L^2_B(\Omega)$, we define:

$$\operatorname{dist}(v,\mathcal{P})_{0,B} \;:=\; \min_{w\in\mathcal{P}} \|v-w\|_{0,B} \ .$$

Similarly, given a function $v \in H^1_\pi(\Omega)$ and a finite dimensional subspace $\mathcal{P} \subset H^1_\pi(\Omega)$, we define:

$$\operatorname{dist}(v,\mathcal{P})_{\kappa,A,B} \;:=\; \min_{w\in\mathcal{P}} \|v-w\|_{\kappa,A,B} \ ,$$

where $\|\cdot\|_{\kappa,A,B}$ is defined in Definition 3.1.

Now let λ_j be any eigenvalue of problem (3.2) for some value $\kappa \in \mathcal{K}$ and let $E(\lambda_j)$ denote the span of all corresponding eigenfunctions. In [22, Lemma 4.3] it is proved that for any computed eigenpair $(\lambda_{j,n}, u_{j,n})$ of (3.8) the two distances are minimized by the same eigenfunction in the continuous eigenspace or in other words:

$$\|u_{j,n} - u_j\|_{0,B} = \operatorname{dist}(u_{j,n}, E_1(\lambda_j))_{0,B} \ , \tag{3.13}$$

if and only if

$$\|u_{j,n} - u_j\|_{\kappa,A,B} = \operatorname{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B} \ , \tag{3.14}$$

where $E_1(\lambda_j) = \{u \in E(\lambda_j) : \|u\|_{0,B} = 1\}$.

To conclude this section we state the a priori convergence results proved in [22]: Suppose $1 \le j \le \dim V_n^p$. Let λ_j be an eigenvalue of (3.2) for some value of $\kappa \in \mathcal{K}$ and with corresponding eigenspace $E(\lambda_j)$ of dimension $R+1 > 0$ and let $(\lambda_{j,n}, u_{j,n})$ be an eigenpair of (3.8) for the same value of κ. Then

(i) For all n,

$$|\lambda_j - \lambda_{j,n}| \;\le\; (\operatorname{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B})^2; \tag{3.15}$$

(ii) For sufficiently small $H_n^{\max}$,

$$\operatorname{dist}(u_{j,n}, E_1(\lambda_j))_{0,B} \quad\le\quad C_1 (H_n^{\max})^s \operatorname{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B} \ ; \tag{3.16}$$

(iii) For sufficiently small $H_n^{\max}$,

$$\operatorname{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B} \;\le C_2 (H_n^{\max})^s \ . \tag{3.17}$$

Where s depends on the regularity of the continuous eigenfunctions, i.e., $E(\lambda_j) \subset H^{1+s}(\Omega)$ and the constants C_1, C_2 depend on the distances between the eigenvalues λ_ℓ, with $\ell = 1, \ldots, j$, the constant C_{reg} in (3.7), on the bounds $\overline{a}, \underline{a}, \overline{b}, \underline{b}$ in (1.2), (1.3) and the ellipticity of the problem.

4 A posteriori error estimator

In this section we are going to introduce two a posteriori error estimators, both based on residuals. The difference between the two is the presence of some weights depending on A. It has been shown numerically in [22] that the presence of these weights improve sensibly the rate of convergence of the method especially for quasimomentum $\kappa \neq \mathbf{0}$. The same idea has been already explored in [4] for a more simple class of problems.

The main results in this section are the upper bound for the residual and the stability of the error estimator, i.e., Theorem 4.6, Corollary 4.7, Theorem 4.9 and Corollary 4.10. These results are used in Section 5 to prove the convergence of the adaptive method.

From the applications point of view, it is very important that the error estimators are reliable and efficient. The reliability ensures that the actual error is always smaller than the error estimators

multiplied by a constant (ignoring higher order terms). On the other hand the efficiency ensures that the a posteriori error estimators are proportional to the actual error (plus higher order terms). So together these results shows that the actual error, which is unknown, and the error estimators, which are computable, are linked together in a linear way casting some confidence on the numerics. The proofs of reliability and efficiency for both error estimators are in [22] for $p = 1$, but the analysis can be easily extended to higher p.

Notation 4.1: From now on, we write $A \lesssim B$ when A/B is bounded by a constant independent of H_n. The notation $A \cong B$ means $A \lesssim B$ and $A \gtrsim B$.

The a posteriori error estimators $\eta_{j,n}$ and $\tilde{\eta}_{j,n}$ are defined as a sum of element residuals and edge residuals, which are all computable quantities. We denote by $\mathcal{F}_n$ the set of all the edges of the elements of the mesh $\mathcal{T}_n$, and we assume to have already chosen an ordering and a preorientated unit normal vector $\vec{n}_f$ for each $f \in \mathcal{F}_n$. Moreover we denote by H_f the length of the face f. Also we denote by $\mathcal{F}_n(\tau)$ the faces of the element $\tau \in \mathcal{T}_n$ and furthermore, we denote by $\tau_1(f)$ and $\tau_2(f)$ the elements sharing $f \in \mathcal{F}_n$. To simplify the notation, we define the functional $[\cdot]_f$ as follow:

Definition 4.2: We can define for any function $g : \Omega \to \mathbb{C}$ and for any $f \in \mathcal{F}_n$

$$[g]_f(x) := \left(\lim_{\substack{\tilde{x}\in\tau_1(f)\\ \tilde{x}\to x}} g(\tilde{x}) \; - \lim_{\substack{\tilde{x}\in\tau_2(f)\\ \tilde{x}\to x}} g(\tilde{x}) \right), \quad \text{with } x \in f \; .$$

Definition 4.3 (Standard a posteriori residual): The definition of the residual estimator $\eta_{j,n}$ involves two functionals: the functional $R_I(\cdot,\cdot)$, which expresses the contributions from the elements in the mesh:

$$R_I(u,\lambda)(x) := \big((\nabla + i\kappa)\cdot A(\nabla + i\kappa)u \; + \; \lambda B u\big)(x), \quad \text{with } x \in \text{int}(\tau), \quad \tau \in \mathcal{T}_n,$$

and the functional $R_F(\cdot)$, which expresses the contributions from the edges of the elements:

$$R_F(u)(x) := \big[\vec{n}_f \cdot A(\nabla + i\kappa)u\big]_f(x), \quad \text{with } x \in \text{int}(f), \quad f \in \mathcal{F}_n \; .$$

(Recall that the jumps of the coefficients are assumed to be aligned with the meshes.) Then the residual estimator $\eta_{j,n}$ for the computed eigenpair $(\lambda_{j,n}, u_{j,n})$ is defined as:

$$\eta_{j,n} := \left\{ \sum_{\tau\in\mathcal{T}_n} \eta_{j,n,\tau}^2 \right\}^{1/2}, \tag{4.1}$$

where

$$\eta_{j,n,\tau} := \left\{ H_\tau^2 \|R_I(u_{j,n},\lambda_{j,n})\|_{0,\tau}^2 \; + \sum_{f\in\mathcal{F}_n(\tau)} H_f \|R_F(u_{j,n})\|_{0,f}^2 \right\}^{1/2}.$$

Definition 4.4 (Modified a posteriori residual): The residual estimator $\tilde{\eta}_{j,n}$ for the computed eigenpair $(\lambda_{j,n}, u_{j,n})$ is defined as:

$$\tilde{\eta}_{j,n} := \left\{ \sum_{\tau\in\mathcal{T}_n} \tilde{\eta}_{j,n,\tau}^2 \right\}^{1/2}, \tag{4.2}$$

where

$$\tilde{\eta}_{j,n,\tau} := \left\{ H_\tau^2 \alpha_\tau^{-1} \|R_I(u_{j,n},\lambda_{j,n})\|_{0,\tau}^2 \; + \sum_{f\in\mathcal{F}_n(\tau)} H_f \alpha_f^{-1} \|R_F(u_{j,n})\|_{0,f}^2 \right\}^{1/2},$$

and where $\alpha_\tau := A_{\max}|_\tau$, $\alpha_f := \max\{A_{\max}|_{\tau_1(f)}, A_{\max}|_{\tau_2(f)}\}$, and $A_{\max}$ denotes the maximum eigenvalue of A. This error estimator has been introduced for the first time in [4].

To simplify the notation we define for any subset of elements $S \subset \mathcal{T}_n$

$$\eta_{j,n,S}^2 := \sum_{\tau \in S} \eta_{j,n,\tau}^2 \ , \quad \tilde{\eta}_{j,n,S}^2 := \sum_{\tau \in S} \tilde{\eta}_{j,n,\tau}^2 \ .$$

Furthermore, let introduce the residual of (3.2).

Definition 4.5: For all $v \in H_\pi^1(\Omega)$ we define the residual of the variational formulation as

$$\langle \mathrm{Res}(v), w \rangle := a_\kappa(v,w) - \Lambda(v)\ b(v,w) \ , \quad \text{for all } w \in H_\pi^1(\Omega) \ ,$$

where $\Lambda(v)$ is the Raileight quotient of v.

Recall the Scott-Zhang quasi-interpolation operator [38] $I_n : H^1(\Omega) \to V_n^p$, which satisfies, for any $v \in H^1(\Omega)$:

$$\|v - I_n v\|_{0,\tau} \lesssim H_\tau \|v\|_{1,\omega(\tau)}, \qquad \text{and} \qquad \|v - I_n v\|_{0,f} \lesssim H_f^{\frac{1}{2}} \|v\|_{1,\omega(f)} \ , \tag{4.3}$$

where $\omega(\tau)$ (respectively $\omega(f)$) denotes the union of all elements sharing at least a vertex with τ (resp. f).

Theorem 4.6 (Upper bound for the residual): Let $(\lambda_{j,n}, u_{j,n})$ be a computed eigenpair, then we have that

$$\|\mathrm{Res}(u_{j,n})\|_{(H_\pi^1(\Omega))'} := \sup_{0 \neq w \in H_\pi^1(\Omega)} \frac{|\langle \mathrm{Res}(u_{j,n}), w \rangle|}{\|w\|_{\kappa,A,B}} \lesssim \eta_{j,n} \ .$$

Proof. From the definition of $\mathrm{Res}(u_{j,n})$ we have that if $w_n \in V_n^p$ then

$$\langle \mathrm{Res}(u_{j,n}), w_n \rangle = 0 \ .$$

So for all $w \in H_\pi^1(\Omega)$ we have that

$$\langle \mathrm{Res}(u_{j,n}), w \rangle = \langle \mathrm{Res}(u_{j,n}), w - w_n \rangle \ ,$$

where w_n is the projection of w on V_n^p using Scott-Zhang projection operator. Then by integration by parts and using (4.3) we have:

$$\begin{aligned} \langle \mathrm{Res}(u_{j,n}), w \rangle &= \sum_{\tau \in \mathcal{T}_n} \Big(\int_\tau -R_I(u_{j,n}, \lambda_{j,n})(\overline{w - w_n})\, dx + \frac{1}{2} \int_{\partial\tau} R_F(u_{j,n})(\overline{w - w_n})\, ds \Big) \\ &\lesssim \sum_{\tau \in \mathcal{T}_n} \eta_{j,n,\tau} \|w\|_{1,\omega(\tau)} \lesssim \eta_{j,n} \|w\|_1 \ , \end{aligned}$$

where the hidden constant in the last inequality may depend on the minimum angle in the mesh. Finally, the result is achieved using (3.5) that shows $\|w\|_1 \lesssim \|w\|_{\kappa,A,B}$.
■

Corollary 4.7: Let $(\lambda_{j,n}, u_{j,n})$ be a computed eigenpair, then we have that

$$\|\mathrm{Res}(u_{j,n})\|_{(H_\pi^1(\Omega))'} := \sup_{0 \neq w \in H_\pi^1(\Omega)} \frac{|\langle \mathrm{Res}(u_{j,n}), w \rangle|}{\|w\|_{\kappa,A,B}} \lesssim \tilde{\eta}_{j,n} \ .$$

Proof. Since $\eta_{j,n}$ and $\tilde{\eta}_{j,n}$ are equal up to multiplication by a constant (independent of the mesh), the result is straightforward. ∎

In order to prove the stability of the error estimators, we are going to use bubble functions, which are in general smooth and positive real valued functions with compact supports and bounded by 1 in the L^∞ norm. We define for any edge f the set Δ_f, which is the union of the two elements sharing f. In particular we need for any element τ a real-valued bubble function ψ_τ with support in τ which vanishes on the edges of τ and for any edge f, we need a real-valued bubble function ψ_f that vanishes outside the closure of Δ_f. In [41, Lemma 3.3], such bubble functions ψ_τ, ψ_f are constructed using polynomials. Moreover, it is proven in [41] that ψ_τ, ψ_f satisfy the following properties:

Proposition 4.8: There are constants, which only depend on the regularity of the mesh $\mathcal{T}_n$ and on the value p, such that the inequalities on an element τ

$$\|v\|_{0,\tau} \lesssim \|\psi_\tau^{1/2} v\|_{0,\tau}, \tag{4.4}$$

$$|\psi_\tau v|_{1,\tau} \lesssim H_\tau^{-1} \|v\|_{0,\tau}, \tag{4.5}$$

and on a edge f

$$\|\omega\|_{0,f} \lesssim \|\psi_f^{1/2}\, \omega\|_{0,f}, \tag{4.6}$$

$$|\psi_f\, \omega|_{1,\Delta_f} \lesssim H_f^{-1/2} \|\omega\|_{0,f}, \tag{4.7}$$

$$\|\psi_f\, \omega\|_{0,\Delta_f} \lesssim H_f^{1/2} \|\omega\|_{0,f}, \tag{4.8}$$

hold for all $\tau \in \mathcal{T}_n$, all $f \in \mathcal{F}_n$, for all polynomials $v \in V_n^p$ and for all polynomials $\omega = v'|_f$ for some $v' \in V_n^p$.

Theorem 4.9 (Stability of the standard error estimator): For any mesh $\mathcal{T}_n$ and for any $(\lambda_{j,n}, u_{j,n})$ eigenpair computed on $\mathcal{T}_n$, we have that

$$\eta_{j,n,\tau} \lesssim (H_\tau(\lambda_{j,n} + \sigma + 1) + 1)\|u_{j,n}\|_{\kappa,A,B,\omega(\tau)}\ , \quad \text{for all } \tau \in \mathcal{T}_n\ , \tag{4.9}$$

and that

$$\eta_{j,n} \lesssim C_{\eta_j}\ , \tag{4.10}$$

where the constant C_{η_j} depends on the index j and on $H_0^{\max}$.

Proof. Let $w_\tau = \psi_\tau R_I(u_{j,n}, \lambda_{j,n})$, where ψ_τ is a bubble function with support in the interior of the element τ, then from (4.4):

$$\begin{aligned}
\|R_I(u_{j,n},\lambda_{j,n})\|_{0,\tau}^2 &\lesssim \|\psi_\tau^{1/2} R_I(u_{j,n},\lambda_{j,n})\|_{0,\tau}^2 = \int_\tau R_I(u_{j,n},\lambda_{j,n})\overline{w_\tau}\,dx \\
&= \int_\tau \Big((\nabla + i\kappa)\cdot(A(\nabla + i\kappa)u_{j,n}) + \lambda_{j,n}\, u_{j,n} B\Big)\overline{w_\tau}\,dx \\
&= \int_\tau -A(\nabla + i\kappa)u_{j,n}\cdot(\nabla - i\kappa)\overline{w_\tau} + \lambda_{j,n}\, u_{j,n}\, B\, \overline{w_\tau}\,dx \\
&= \int_\tau -A(\nabla + i\kappa)u_{j,n}\cdot(\nabla - i\kappa)\overline{w_\tau} - \sigma\, u_{j,n}\, B\, \overline{w_\tau}\,dx \\
&\quad + \int_\tau \sigma\, u_{j,n}\, B\, \overline{w_\tau} + \lambda_{j,n}\, u_{j,n}\, B\, \overline{w_\tau}\,dx\ ,
\end{aligned}$$

where we used integration by parts. Then using Cauchy Schwarz, (3.5), the fact that $\psi_\tau \leq 1$ and (4.5) we have

$$\begin{aligned}
\|R_I(u_{j,n},\lambda_{j,n})\|_{0,\tau}^2 &\leq \|u_{j,n}\|_{\kappa,A,B,\tau}\|w_\tau\|_{\kappa,A,B,\tau} + (\lambda_{j,n}+\sigma)\|u_{j,n}\|_{0,B,\tau}\|w_\tau\|_{0,B,\tau} \\
&\lesssim \big((H_\tau^{-1}+1)\|u_{j,n}\|_{\kappa,A,B,\tau} + (\lambda_{j,n}+\sigma)\|u_{j,n}\|_{0,\tau}\big)\|R_I(u_{j,n},\lambda_{j,n})\|_{0,\tau} \\
&\lesssim \big(H_\tau^{-1}+\lambda_{j,n}+\sigma+1\big)\|u_{j,n}\|_{\kappa,A,B,\tau}\ \|R_I(u_{j,n},\lambda_{j,n})\|_{0,\tau}. \qquad (4.11)
\end{aligned}$$

Now, let $w_f = \psi_f R_F(u_{j,n})$, where ψ_f is a bubble function of the face f and with support Δ_f, then from (4.6):

$$\begin{aligned}
\|R_F(u_{j,n})\|_{0,f}^2 &\lesssim \|\psi_f^{1/2}R_F(u_{j,n})\|_{0,f}^2 = \int_f R_F(u_{j,n})\overline{w_f}\,ds \\
&= \int_{\Delta_f} R_I(u_{j,n},\lambda_{j,n})\overline{w_f}\,dx + a_\kappa(u_{j,n},w_f) - \lambda_{j,n}b(u_{j,n},w_f)\ ,
\end{aligned}$$

where we used integration by parts. Then using Cauchy Schwarz and (4.11) we have

$$\begin{aligned}
\|R_F(u_{j,n})\|_{0,f}^2 &\leq \|R_I(u_{j,n},\lambda_{j,n})\|_{0,\Delta_f}\|w_f\|_{0,\Delta_f} + \|u_{j,n}\|_{\kappa,A,B,\Delta_f}\|w_f\|_{\kappa,A,B,\Delta_f} \\
&\quad + (\lambda_{j,n}+\sigma)\|u_{j,n}\|_{0,\Delta_f}\|w_f\|_{0,\Delta_f} \\
&\lesssim \big(H_\tau^{-1}+\lambda_{j,n}+\sigma+1\big)\|u_{j,n}\|_{\kappa,A,B,\Delta_f}\|w_f\|_{0,\Delta_f} + \|u_{j,n}\|_{\kappa,A,B,\Delta_f}\|w_f\|_{\kappa,A,B,\Delta_f} \\
&\quad + (\lambda_{j,n}+\sigma)\|u_{j,n}\|_{0,\Delta_f}\|w_f\|_{0,\Delta_f}\ .
\end{aligned}$$

Finally using (4.7) and (4.8) we have

$$\begin{aligned}
\|R_F(u_{j,n})\|_{0,f}^2 &\lesssim \Big(H_f^{-1/2}+H_f^{1/2}+(\lambda_{j,n}+\sigma+H_\tau^{-1})H_f^{1/2} \\
&\quad + (\lambda_{j,n}+\sigma)H_f^{1/2}\Big)\|u_{j,n}\|_{\kappa,A,B,\Delta_f}\|R_F(u_{j,n})\|_{0,f}\ . \qquad (4.12)
\end{aligned}$$

So, putting together (4.11) and (4.12) we got

$$\eta_{j,n,\tau}^2 \lesssim (H_\tau(\lambda_{j,n}+\sigma+1)+1)^2\|u_{j,n}\|_{\kappa,A,B,\omega(\tau)}^2\ ,$$

that is (4.9).

In order to prove (4.10) we just have to sum (4.9) over all elements in $\mathcal{T}_n$ and we use (3.15), (3.17) and the minimum-maximum principle (3.12):

$$\begin{aligned}
\eta_{j,n}^2 &\lesssim (H_n^{\max}(\lambda_{j,n}+\sigma+1)+1)^2\|u_{j,n}\|_{\kappa,A,B}^2 = (H_n^{\max}(\lambda_{j,n}+\sigma)+1)^2(\lambda_{j,n}+\sigma) \\
&\lesssim (H_n^{\max}(\lambda_j+(H_n^{\max})^{2s}+\sigma+1)+1)^2(\lambda_j+(H_n^{\max})^{2s}+\sigma) \\
&\leq (H_0^{\max}(\lambda_j+(H_0^{\max})^{2s}+\sigma+1)+1)^2(\lambda_j+(H_0^{\max})^{2s}+\sigma)\ .
\end{aligned}$$

■

Corollary 4.10: For any mesh $\mathcal{T}_n$ and for any $(\lambda_{j,n},u_{j,n})$ computed eigenpair using $\mathcal{T}_n$, we have that

$$\tilde{\eta}_{j,n,\tau} \lesssim (H_\tau(\lambda_{j,n}+\sigma+1)+1)\|u_{j,n}\|_{\kappa,A,B,\omega(\tau)}\ , \quad \text{for all } \tau\in\mathcal{T}_n\ ,$$

and that

$$\tilde{\eta}_{j,n} \lesssim C_{\tilde{\eta}_j} \ ,$$

where the constant $C_{\tilde{\eta}_j}$ depends on the considered eigenvalue and on $H_0^{\max}$.

Proof. Since $\eta_{j,n}$ and $\tilde{\eta}_{j,n}$ are equal up to multiplication by a constant (independent of the mesh), the result is straightforward. ∎

5 Convergence

In this section we prove the convergence of the adaptive finite element method for both eigenvalue and eigenfunctions. In order to make the proof working, we need some mild assumptions on the subdivision strategy for elements and on the marking strategy.

Algorithm 1 Convergent adaptive algorithm

$(\lambda_{j,n}, u_{j,n}, \mathcal{T}_n) := \text{ConvAFEM}(\kappa, \mathcal{T}_0, j)$
$n = 0$
repeat
  Compute the j-th eigenpair $(\lambda_{j,n}, u_{j,n})$ with quasimomentum κ on $\mathcal{T}_n$
  Compute $\eta_{j,n,\tau}$ for all $\tau \in \mathcal{T}_n$
  Mark the elements using the marking strategy
  Refine the mesh $\mathcal{T}_n$ and construct $\mathcal{T}_{n+1}$
  $n = n + 1$
until

Algorithm 2 Adaptivity algorithm

$(\lambda_{j,n}, u_{j,n}, \mathcal{T}_{j,n}) := \text{AFEM}(\kappa, \mathcal{T}_0, j, \theta, \text{tol}, \max_n)$
$n = 0$
repeat
  Compute the j-th eigenpair $(\lambda_{j,n}, u_{j,n})$ with quasimomentum κ on $\mathcal{T}_n$
  Compute $\eta_{j,n,\tau}$ for all $\tau \in \mathcal{T}_n$
  Mark the elements using the marking strategy (Definition 5.2)
  Refine the mesh $\mathcal{T}_n$ and construct $\mathcal{T}_{n+1}$
  $n = n + 1$
until $\eta_{j,n} \leq \text{tol}$ OR $n \geq \max_n$

More precisely we use the so called "red-refinement" procedure to refine the marked elements, which consists in splitting the marked elements in four smaller elements. This procedure could generate a non-conforming mesh due to the presence of hanging nodes. To recover the conformity, the closure of the mesh is computed, i.e, all the elements with a hanging node on one of their edges are split in two smaller elements using the "green-refinement" strategy. Let introduce the following notation: all the meshes before the application of the closure are denoted with a tilde, e.g. $\tilde{\mathcal{T}}_n$, instead all the conforming meshes resulting from the application of the closure are denote as before, e.g. $\mathcal{T}_n$.

Let's explain in more detail the refinement procedure, which is the standard red-green refinement which keeps the refined meshes shape regular. The initial mesh $\mathcal{T}_0$ is assumed to be conforming and we also set $\tilde{\mathcal{T}}_0 \equiv \mathcal{T}_0$. Then during the first iteration of Algorithm 1 the marking procedure marks some elements of $\mathcal{T}_0$ to be refined, all the marked elements and all the elements with more than one marked neighbors are refined using the "red-refinement" strategy. The resulting mesh $\tilde{\mathcal{T}}_1$ could be

not conforming since there could be some hanging nodes, but there are no elements with more than 1 hanging node. After the application of the closure to $\tilde{\mathcal{T}}_1$, a conforming mesh $\mathcal{T}_1$ results.

From the second iteration on of Algorithm 1 the refining procedure is more complicated because in order to keep the mesh shape regular, the closure have to be undone before the refinement is applied. So, without loss in generality let suppose that $n \geq 2$, then the marks on the elements of $\mathcal{T}_n$ are passed on the elements of $\tilde{\mathcal{T}}_n$ for which we can distinguish two different cases:

1. All elements $\tilde{\tau} \in \tilde{\mathcal{T}}_n$ that are also marked elements in $\mathcal{T}_n$, are marked;
2. All elements $\tilde{\tau} \in \tilde{\mathcal{T}}_n$ that have been split in two children elements by the closure and such that at least one of their two children is marked, are marked.

Then the "red-refinement" strategy is applied to the marked elements of $\tilde{\mathcal{T}}_n$ and to all elements with more than one refined neighbors, resulting in the mesh $\tilde{\mathcal{T}}_{n+1}$. Finally the application of the closure to $\tilde{\mathcal{T}}_{n+1}$ gives us $\mathcal{T}_{n+1}$.

The red and the green subdivision strategies satisfy the following properties: any element τ is subdivided into elements $\tau'_1, \cdots, \tau'_{n_\tau}$ such that

$$\tau = \tau'_1 \cup \cdots \cup \tau'_{n_\tau} \ ,$$

$$|\tau| = |\tau'_1| + \cdots + |\tau'_{n_\tau}| \ ,$$

$$\exists \ \underline{s} > 0, \overline{s} < 1 : \underline{s}|\tau| \leq |\tau'_i| \leq \overline{s}|\tau| \ ,$$

for all $i = 1, \cdots, n_\tau$, with $n_\tau = 2$ for the green refinement and $n_\tau = 4$ for the red.

The algorithm that we use in practice is Algorithm 2, where the only difference with Algorithm 1 is the presence of the stopping condition that stops the computation either when the number of iterations has exceeded $\max_n \in \mathbb{N}$ or when the estimation of the error $\eta_{j,n}$ is below a positive tolerance tol $\in \mathbb{R}$. We use the same algorithms also with the a posteriori error estimator $\tilde{\eta}_{j,n}$.

In order to prove the convergence we need the following weak assumption on the marking strategy which has been already used in [18]:

Assumption 5.1: We assume that the marking strategy is such that at least one element of $\mathcal{T}_n$ holding the largest value for the error estimator is marked for refinement.

In all numerics we use the following marking strategy that clearly satisfies the assumption:

Definition 5.2 (Marking Strategy): Given a parameter $0 < \theta < 1$, the procedure is: mark the elements in a minimal subset $\mathcal{M}_n$ of $\mathcal{T}_n$ such that

$$\eta_{j,n,\mathcal{M}_n} \geq \ \theta \ \eta_{j,n} \ . \tag{5.1}$$

Also when the "modified" error estimator $\tilde{\eta}_{j,n}$ is used, an analogous marking strategy is employed.

Before presenting the most technical results, we would like to show that for fixed j the sequence of computed eigenvalues $\{\lambda_{j,n}\}_{n\in\mathbb{N}}$ always converges to a point $\lambda_{j,\infty}$ and that within the sequence of computed eigenfunctions $\{u_{j,n}\}_{n\in\mathbb{N}}$ there is a converging subsequence that converges to the point $u_{j,\infty}$. What here is missing in order to have a complete convergence result for the adaptive finite element method (AFEM) is the proof that the limit pair $(\lambda_{j,\infty}, u_{j,\infty})$ is an eigenpair of the continuous problem. This is what is proved in the rest of the section.

Theorem 5.3 (Convergence of a subsequence): For any sequence of meshes $\{\mathcal{T}_n\}_{n\in\mathbb{N}}$ generated by Algorithm 1 for a given j, we have that there exists $\lambda_{j,\infty}$ such that the sequence of computed eigenvalues $\lambda_{j,n}$ converges to $\lambda_{j,\infty} \geq 0$ when n goes to infinity. Moreover, there is a subsequence $\{u_{j,n_m}\}_{m\in\mathbb{N}_0}$ of the sequence of corresponding eigenfunctions $\{u_{j,n}\}_{n\in\mathbb{N}_0}$ that converges to a function $u_{j,\infty} \in H^1_\pi(\Omega)$, with $\|u_{j,\infty}\|_0 = 1$.

Proof. We know already that the sequence $\{\lambda_{j,n}\}_{n\in\mathbb{N}_0}$ is bounded from below by 0 and that it is a non-increasing sequence. Then we can conclude that the limit exists and we denote it by $\lambda_{j,\infty}$.

From problem (3.8) we have that

$$\|u_{j,n}\|^2_{\kappa,A,B} \;=\; a_\kappa(u_{j,n},u_{j,n}) \;+\; \sigma\; b(u_{j,n},u_{j,n}) \;=\; (\lambda_{j,n}+\sigma)\; b(u_{j,n},u_{j,n}) \;=\; \lambda_{j,n}+\sigma\;, \qquad (5.2)$$

which implies that the sequence $\{\|u_{j,n}\|^2_{\kappa,A,B}\}_{n\in\mathbb{N}_0}$ converges to $\lambda_{j,\infty}+\sigma$. Therefore we have that the sequence $\{u_{j,n}\}_{n\in\mathbb{N}_0}$ is bounded in $H^1_\pi(\Omega)$. Then there exists a subsequence $\{u_{j,n_p}\}_{p\in\mathbb{N}_0}$ converging weakly to a function $u_{j,\infty}\in H^1_\pi(\Omega)$ respect to the norm of $\|\cdot\|_{\kappa,A,B}$. Using the fact that the imbedding $H^1_\pi(\Omega)\subset L^2(\Omega)$ is compact, there is a subsequence of $\{u_{j,n_p}\}_{p\in\mathbb{N}_0}$, which we denote $\{u_{j,n_m}\}_{m\in\mathbb{N}_0}$, such that

$$u_{j,n_m} \;\to\; u_{j,\infty} \quad \text{in } L^2(\Omega)\;. \qquad (5.3)$$

Since $\|u_{j,n_m}\|_{\kappa,A,B}$ is a subsequence of $\|u_{j,n}\|_{\kappa,A,B}$, we have that $\|u_{j,n_m}\|_{\kappa,A,B}\to\|u_{j,\infty}\|_{\kappa,A,B}$. This together with (5.3), gives us

$$u_{j,n_m} \;\to\; u_{j,\infty} \quad \text{in } H^1_\pi(\Omega)\;.$$

■

Definition 5.4: Considering the sequence of meshes $\{\tilde{\mathcal{T}}_n\}_{n\in\mathbb{N}_0}$ constructed using Algorithm 1, we define

$$\tilde{\mathcal{T}}_n^+ := \{\tau\in\tilde{\mathcal{T}}_n : \tau\in\tilde{\mathcal{T}}_m, \forall m\geq n\}\;,$$

the set of all elements that are never refined again and

$$\tilde{\mathcal{T}}_n^0 := \tilde{\mathcal{T}}_n \setminus \tilde{\mathcal{T}}_n^+\;,$$

the set of elements that are going to be refined. We also introduce

$$\Omega_n^+ := \cup_{\tau\in\tilde{\mathcal{T}}_n^+}\omega(\tau)\;,$$

and

$$\Omega_n^0 := \cup_{\tau\in\tilde{\mathcal{T}}_n^0}\omega(\tau)\;.$$

We introduce the mesh size function $\tilde{H}_n\in L^\infty(\Omega)$ associated with the mesh $\tilde{\mathcal{T}}_n$ such that for any point x in the interior of τ, with $\tau\in\tilde{\mathcal{T}}_n$, $\tilde{H}_n(x)=\tilde{H}_\tau$. Since the function $\tilde{H}_n$ is not uniquely defined on the edges of the mesh, we are going to restrict its domain to $\Omega/\tilde{\Sigma}_n$, where $\tilde{\Sigma}_n$ is the skeleton of the mesh $\tilde{\mathcal{T}}_n$, which has 2-dimensional Lebesgue measure zero. Moreover the limiting skeleton $\tilde{\Sigma}_\infty=\cup_{n\in\mathbb{N}}\tilde{\Sigma}_n$ has 2-dimensional Lebesgue measure zero, too, so it is straightforward to see that, thanks to the applications of the refinement procedures, for any point $x\in\Omega/\tilde{\Sigma}_\infty$ the sequence $\{\tilde{H}_n(x)\}_{n\in\mathbb{N}_0}$ is monotonically non-increasing and bounded from below by 0. So we can define for any point $x\in\Omega/\tilde{\Sigma}_\infty$:

$$\tilde{H}_\infty(x) := \lim_{n\to\infty}\tilde{H}_n(x)\;.$$

The next lemma has been already proved in [30, Lemma 4.3] for a conforming sequence of meshes. The fact that the sequence $\{\tilde{\mathcal{T}}_n\}_{n\in\mathbb{N}_0}$ is constituted by non-conforming meshes has no effect on the proof itself.

Lemma 5.5: Considering the sequence of mesh size functions $\{\tilde{H}_n\}_{n\in\mathbb{N}_0}$ and its pointwise limit $\tilde{H}_\infty$ we have that

$$\lim_{n\to\infty}\|\tilde{H}_n-\tilde{H}_\infty\|_{\infty,\Omega} \;=\; 0\;.$$

Lemma 5.6: Denoting by $\mathcal{X}_{\Omega_n^0}$ the characteristic function of Ω_n^0 we have that

$$\lim_{n\to\infty} \|\tilde{H}_n \mathcal{X}_{\Omega_n^0}\|_\infty \;=\; 0 \;.$$

Proof. For any $\tau \in \mathcal{T}_n^0$ we have by definition that

$$\tilde{H}_\infty \leq \overline{s}^{1/2} \tilde{H}_n \;,$$

in τ. Then in τ

$$\tilde{H}_n \leq \alpha(\tilde{H}_n - \tilde{H}_\infty) \;,$$

with $\alpha = 1/(1-\overline{s}^{1/2})$. Since $\tau \in \mathcal{T}_n^0$ is arbitrary and also taking into account the shape regularity that impose to neighbors elements similar sizes, this implies

$$\|\tilde{H}_n \mathcal{X}_{\Omega_n^0}\|_\infty \leq \|\alpha(\tilde{H}_n - \tilde{H}_\infty)\mathcal{X}_{\Omega_n^0}\|_\infty \leq \alpha \|\tilde{H}_n - \tilde{H}_\infty\|_\infty \;,$$

which converges to 0 by the previous lemma. ∎

Lemma 5.7 (Decay of the estimator on marked elements): Let $\{u_{j,n_m}\}_{m\in\mathbb{N}_0}$ be the convergent subsequence of $\{u_{j,n}\}_{n\in\mathbb{N}_0}$ as in Theorem 5.3. Then,

$$\lim_{m\to\infty} \max_{\tau\in\mathcal{M}_{n_m}} \eta_{j,n_m,\tau} = 0 \;.$$

Proof. Let $\tau^{\max} \in \mathcal{M}_{n_m}$ be the element where the maximum is reached. So we have from Theorem 4.9:

$$\eta_{j,n_m,\tau^{\max}} \lesssim \|u_{j,n_m}\|_{\kappa,A,B,\omega(\tau^{\max})} \leq \|u_{j,n_m} - u_{j,\infty}\|_{\kappa,A,B,\Omega} + \|u_{j,\infty}\|_{\kappa,A,B,\omega(\tau^{\max})} \;.$$

We already know that the first term on the rhs tends to zero from Theorem 5.3. Also the second term tends to zeros since Lemma 5.6 and Assumption 5.1

$$|\omega(\tau^{\max})| \lesssim H^2_{\tau^{\max}} \leq \tilde{H}^2_{\tilde{\tau}^{\max}} \leq \|\tilde{H}_{n_m} \mathcal{X}_{\Omega^0_{n_m}}\|^2_{\infty,\Omega} \to 0 \;,$$

where the element $\tilde{\tau}^{\max}$ is the element in $\tilde{\mathcal{T}}_{n_m}$ either that coincides with $\tau^{\max}$ or the element that contains $\tau^{\max}$ in the case that $\tau^{\max}$ has been split by the closure procedure. ∎

Theorem 5.8 (Weak convergence of the residual): Let $\{u_{j,n_m}\}_{m\in\mathbb{N}_0}$ be the convergent subsequence of $\{u_{j,n}\}_{n\in\mathbb{N}_0}$ as in Theorem 5.3. Then,

$$\lim_{m\to\infty} \langle \mathrm{Res}(u_{j,n_m}), v\rangle = 0 \;, \quad \text{for all } v \in H^1_\pi(\Omega) \;.$$

Proof. Let's start proving the result for $v \in H^2(\Omega) \cap H^1_\pi(\Omega)$. Let $k \in \mathbb{N}_0$ and $n_m > k$ then we have

$$\tilde{\mathcal{T}}_k^+ \subset \tilde{\mathcal{T}}_{n_m}^+ \subset \tilde{\mathcal{T}}_{n_m} \;.$$

Let v_{n_m} the Lagrange interpolation of v, so using Theorem 4.6 we have:

$$|\langle \mathrm{Res}(u_{j,n_m}), v\rangle| = |\langle \mathrm{Res}(u_{j,n_m}), v - v_{n_m}\rangle| \lesssim \sum_{\tau\in\mathcal{T}_{n_m}} \eta_{j,n_m,\tau} \|v - v_{n_m}\|_{\kappa,A,B,\omega(\tau)} \;. \tag{5.4}$$

Using the relation between the elements of the two meshes $\mathcal{T}_{n_m}$ and $\tilde{\mathcal{T}}_{n_m}$ we can rewrite (5.4) as

$$|\langle \mathrm{Res}(u_{j,n_m}), v\rangle| \lesssim \sum_{\tilde{\tau}\in\tilde{\mathcal{T}}_{n_m}} \eta_{j,n_m,\tilde{\tau}} \|v - v_{n_m}\|_{\kappa,A,B,\omega(\tilde{\tau})} \;, \tag{5.5}$$

where either $\eta_{j,n_m,\tilde{\tau}}$ is the same as $\eta_{j,n_m,\tau}$ if the two elements τ and $\tilde{\tau}$ coincide or $\eta^2_{j,n_m,\tilde{\tau}} = \eta^2_{j,n_m,\tau_1} + \eta^2_{j,n_m,\tau_2}$ is the sum of the residuals of the two elements τ_1, τ_2 in $\mathcal{T}_{n_m}$ in which $\tilde{\tau}$ has been split during the closure of the mesh. Then

$$\begin{aligned}
|\langle \mathrm{Res}(u_{j,n_m}), v\rangle| &\lesssim \sum_{\tilde{\tau}\in\tilde{\mathcal{T}}_{n_m}} \eta_{j,n_m,\tilde{\tau}} \|v - v_{n_m}\|_{\kappa,A,B,\omega(\tilde{\tau})} \\
&= \sum_{\tilde{\tau}\in\tilde{\mathcal{T}}_k^+} \eta_{j,n_m,\tilde{\tau}} \|v - v_{n_m}\|_{\kappa,A,B,\omega(\tilde{\tau})} + \sum_{\tilde{\tau}\in\tilde{\mathcal{T}}_{n_m}\setminus\tilde{\mathcal{T}}_k^+} \eta_{j,n_m,\tilde{\tau}} \|v - v_{n_m}\|_{\kappa,A,B,\omega(\tilde{\tau})} \\
&\lesssim \eta_{j,n_m,\tilde{\mathcal{T}}_k^+} \|v - v_{n_m}\|_{\kappa,A,B,\Omega_k^+} + \eta_{j,n_m,\tilde{\mathcal{T}}_{n_m}\setminus\tilde{\mathcal{T}}_k^+} \|v - v_{n_m}\|_{\kappa,A,B,\Omega_k^0} \ ,
\end{aligned}$$

where we used also the fact that $\Omega^o_{n_m} \subset \Omega^o_k$ since $\mathcal{T}_k^+ \subset \mathcal{T}^+_{n_m}$. So form Theorem 4.9 we have that $\eta_{j,n_m,\tilde{\mathcal{T}}_{n_m}\setminus\tilde{\mathcal{T}}_k^+} \leq \eta_{j,n_m} \leq C_{\eta_j}$, then using interpolation estimates and the fact that in the interior of any $\tau \in \mathcal{T}_{n_m}$ $\tilde{H}_{n_m} \leq \tilde{H}_k$ we have:

$$|\langle \mathrm{Res}(u_{j,n_m}), v\rangle| \lesssim (\eta_{j,n_m,\tilde{\mathcal{T}}_k^+} H^{\max}_{n_m} + C_{\eta_{n_m}} \|\tilde{H}_k \mathcal{X}_{\Omega_k^0}\|_\infty)\|v\|_2 \ .$$

So for any $\epsilon > 0$ we have that we can choose the k introduced before such that by Lemma 5.6 we have:

$$C_{\eta_j} \|\tilde{H}_k \mathcal{X}_{\Omega_k^0}\|_\infty \leq \epsilon \ .$$

On the other hand by Assumption 5.1

$$\begin{aligned}
\eta_{n_m,\tilde{\mathcal{T}}_k^+} &\leq (\#\tilde{\mathcal{T}}_k^+)^{1/2} \mathrm{max}_{\tilde{\tau}\in\tilde{\mathcal{T}}_k^+} \eta_{j,n_m,\tilde{\tau}} \leq (\#\tilde{\mathcal{T}}_k^+)^{1/2} \mathrm{max}_{\tau\in\mathcal{T}_k^+} \eta_{j,n_m,\tau} \\
&\leq (\#\tilde{\mathcal{T}}_k^+)^{1/2} \mathrm{max}_{\tau\in\mathcal{M}_{n_m}} \eta_{n_m,\tau} \ .
\end{aligned}$$

So we can choose a $K > k$ such that for all $n_m > K$ we have from Lemma 5.7:

$$\eta_{n_m,\tilde{\mathcal{T}}_k^+} \leq \epsilon \ .$$

Finally, since H^2 is dense in H^1_π, we can extend the result. ∎

Theorem 5.9: Under the same assumptions as in Theorem 5.3, we have that the limiting pair $(\lambda_{j,\infty}, u_{j,\infty})$ is an eigenpair of the problem:

$$\left.\begin{array}{lcll} a_\kappa(u_{j,\infty}, v) & = & \lambda_{j,\infty} b(u_{j,\infty}, v) \ , & \text{for all} \quad v \in H^1_\pi(\Omega) \\ \|u_{j,\infty}\|_{0,B,\Omega} & = & 1 \ . & \end{array}\right\} \tag{5.6}$$

Proof. Denoting with $\{u_{j,n_m}\}_{m\in\mathbb{N}_0}$ the converging subsequence in Theorem 5.3, we have

$$\begin{aligned}
&|\langle \mathrm{Res}(u_{j,\infty}), v\rangle| = |\langle \mathrm{Res}(u_{j,\infty}) - \mathrm{Res}(u_{j,n_m}), v\rangle + \langle \mathrm{Res}(u_{j,n_m}), v\rangle| \\
&\leq |a_\kappa(u_{j,\infty} - u_{j,n_m}, v)| + |b(\lambda_{j,\infty}u_{j,\infty} - \lambda_{n_m}u_{j,n_m}, v)| + |\langle \mathrm{Res}(u_{j,n_m}), v\rangle| \\
&\leq \|u_{j,\infty} - u_{j,n_m}\|_{\kappa,A,B}\|v\|_{\kappa,A,B} + \|\lambda_{j,\infty}u_{j,\infty} - \lambda_{n_m}u_{j,n_m}\|_{0,B}\|v\|_{0,B} + |\langle \mathrm{Res}(u_{j,n_m}), v\rangle| \ .
\end{aligned}$$

which converges to 0 thanks to Theorem 5.3 and Theorem 5.8. ∎

Theorem 5.10 (Convergence result): For any j, let $\{\lambda_{j,n}, u_{j,n}\}_{n\in\mathbb{N}_0}$ denote the whole sequence of computed eigenpairs obtained with Algorithm 1. Then there exists an eigenvalue λ of the continuous problem (3.2) such that

$$\lim_{n\to\infty} \lambda_{j,n} = \lambda \ ,$$

and

$$\lim_{n\to\infty} \mathrm{dist}(u_{j,n}, E(\lambda))_{\kappa,A,B} = 0 \ .$$

Proof. Taking $\lambda := \lambda_\infty$ and by Theorem 5.3 we have that $lim_{n\to\infty}\lambda_{j,n} = \lambda$, also from Theorem 5.9 we can conclude that λ is an eigenvalue of (3.2). In order to prove $\lim_{n\to\infty} \mathrm{dist}(u_{j,n}, E(\lambda))_{\kappa,A,B} = 0$ we argue by contradiction. Supposing that the result were not true, then there would exist a subsequence $\{u_{j,n_r}\}_{m\in\mathbb{N}_0}$ of $\{u_{j,n}\}_{n\in\mathbb{N}_0}$ such that

$$\mathrm{dist}(u_{j,n_r}, E(\lambda))_{\kappa,A,B} > \epsilon \ , \quad \forall r \in \mathbb{N}_0 \ , \tag{5.7}$$

where ϵ is a positive real number.

Then applying the same arguments as in Theorem 5.3 and in Theorem 5.9, we would be able to extract a subsequence $\{u_{j,n_m}\}_{m\in\mathbb{N}_0}$ of $\{u_{j,n_r}\}_{m\in\mathbb{N}_0}$ converging to some function $u'_{j,\infty}$ which would still be in the eigenspace $E(\lambda)$ since the corresponding sequence of eigenvalues converges to λ. Now we have a contradiction because there is a subsequence of $\{u_{j,n_m}\}_{m\in\mathbb{N}_0}$ converging into $E(\lambda)$ even if (5.7).

■

Remark 5.11: Theorem 5.10 shows that a sequence of computed eigenvalues converges to a true eigenvalue λ, but it doesn't show to what eigenvalue the sequence converges. Suppose the pathological case where the initial mesh $\mathcal{T}_0$ is orthogonal to the eigenspace $E(\lambda_j)$ for some $j \leq \dim V_0^p$, also suppose that all meshes in the sequence $\{\mathcal{T}_n\}_{n\in\mathbb{N}}$ are orthogonal to $E(\lambda_j)$. In this case running Algorithm 1 or Algorithm 2 with such an index j generates a sequence of computed eigenvalues $\lambda_{j,n}$ that approximate some eigenvalues higher in the spectrum than λ_j and also the sequence would converge to some eigenvalue higher in the spectrum than λ_j. Such situations will not occur if the initial mesh $\mathcal{T}_0$ is fine enough such that all eigenspaces of eigenvalues $\lambda_1, \ldots, \lambda_j$ are well represented in V_n^p with the correct multiplicity.

6 An efficient and convergent method to compute the bands

It is very common in practice the need to compute the extrema of a band in the spectrum. For example, in presence of a gap in the spectrum, the minimum of the band above the gap and the maximum of the band below the gap will define the size of the gap. Also when the supercell framework [39] is used to search for trapped modes, the extrema of the band corresponding to a trapped mode, will assess the quality of the computation. If the band is too wide, the supercell should be increased in order to obtain more accurate results [22, Section 7]. The global extrema of a band are not easy to find, all methods to search for them based on Newton method could converge to local extrema and not to the global ones, for some starting points. Instead a method that computes and converges to the whole band would not suffer of this drawback.

The most trivial way to try to approximate an entire band in the spectrum is to choose as many values of quasimomentum κ as possible and for each value of κ run Algorithm 2 starting from the same initial mesh. This method is very inefficient because, from the theory, [8, 28] it is clear that the bands in the spectrum are continuous, in the sense that each eigenvalue as a function of κ is continuous. So, it is reasonable to suppose that, for close values of κ, the corresponding eigenvalues in the same band are very close, too. Moreover, the adaptive method should produce very similar

meshes for close enough values of κ. This suggests a more efficient way to approximate bands, in which information from consecutive runs of Algorithm 2 for different values of κ are shared.

In this section we are going to describe such an efficient method to compute bands in the spectrum. By efficient we mean that this method needs fewer mesh refinements to reach the same approximation of a band over a set of values of quasimomentum κ, compared to Algorithm 1 applied to each value of κ in the set individually. Moreover, we are going to show that the sequence of approximated bands $\mathcal{C}_{j,m}$ computed with our efficient method converges to the true band $\mathcal{C}_j$.

Let $\mathcal{G}_0$ be a conforming and shape regular mesh of triangles constructed on the reduced first Brillouin zone $\mathcal{K}_{\text{red}}$, which is a subset of $\mathcal{K}$ [2, 25]. We can restrict to $\mathcal{K}_{\text{red}}$ because it comes from the theory that the extrema of any band over $\mathcal{K}$ are also reached in $\mathcal{K}_{\text{red}}$. In the following we are going to construct a sequence of meshes on $\mathcal{K}_{\text{red}}$ starting with the mesh $\mathcal{G}_0$ and where $\mathcal{G}_{m+1}$ is the resulting mesh after all the elements in $\mathcal{G}_m$ have been refined as described in Figure 1. It is important to understand that the meshes $\mathcal{G}_m$ are different from the meshes $\mathcal{T}_n$, since the formers are subdivision of the reduced first Brillouin zone $\mathcal{K}_{\text{red}}$, while $\mathcal{T}_n$ are subdivision of the primitive cell Ω. Moreover, we denote by $\mathcal{N}_m$ the set of all the nodes in the mesh $\mathcal{G}_m$ and with $\mathcal{N}_\infty = \bigcup_{m\geq 0} \mathcal{N}_m$.

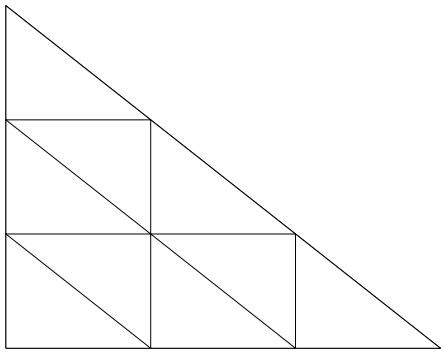

Fig. 1: A reference element of a mesh $\mathcal{G}_m$ split in 9 elements of $\mathcal{G}_{m+1}$.

Let's introduce the notation $(\lambda_{j,m}^\kappa, u_{j,m}^\kappa)$ and $\mathcal{T}_m^\kappa$ to denote the computed eigenpair of index j and the mesh used to compute it for the value of the quasimomentum $\kappa \in \mathcal{N}_m$. Thanks to the particular refining procedure that we have adopted for $\mathcal{G}_m$, each point $\kappa \in \mathcal{N}_{m+1}$ has a unique "father" $\kappa' \in \mathcal{N}_m$, where the father of the node κ is the node κ' closest to κ in the reference element. In the case that $\kappa \in \mathcal{N}_{m+1} \cap \mathcal{N}_m$ then the father is $\kappa' \equiv \kappa$. The relation is explained graphically in Figure 2.

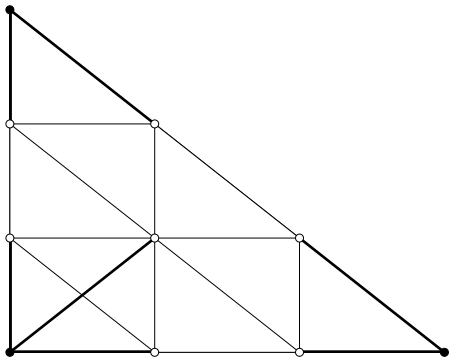

Fig. 2: A refined reference element of a mesh $\mathcal{G}_m$, where the black dots are the "father" nodes and the white dots are the "children". The thick lines link the children to their fathers.

Our efficient method to approximate bands is illustrated in Algorithm 3. This algorithm works in two stages A and B. In the stage A, which is the external repeat-until loop with counter m, the algorithm constructs the sequence of meshes $\mathcal{G}_m$ on the reduced first Brillouin zone $\mathcal{K}_{\text{red}}$. At each iteration a finer mesh $\mathcal{G}_{m+1}$ is constructed refining the previous mesh $\mathcal{G}_m$ using the refinement procedure illustrated in Figure 1. Moreover, each iteration of stage A constructs an approximation $\mathcal{C}_{j,m}$ of the band of interest using stage B, which is described next. In the stage B, which is the inner for-all-do loop, many sequences of adapted meshes on the primitive cell Ω are constructed, each sequence corresponds to a different node $\kappa \in \mathcal{N}_m$. The purpose of this stage is to apply our AFEM to approximate the eigenpair of interest for each value of the quasimomentum $\kappa \in \mathcal{N}_m$ starting from the last adapted mesh constructed for the father of κ during the previous iteration of stage A. To this end we define the function FatherMesh which takes as argument a point $\kappa \in \mathcal{N}_m$ and it returns the mesh $\mathcal{T}_{m-1}^{\kappa'}$ which is the finest mesh generated by the AFEM procedure at the previous iteration where κ' is the father of κ. Any run of Algorithm 3 may consist in many iterations of stages A and B. The Algorithm 3 is efficient in approximating bands, since, for each node $\kappa \in \mathcal{N}_m$, the adaptive procedure, which is used to further improve the approximation of the eigenpair, starts from the already adapted mesh for the father node from the previous iteration of stage A. In this way we take advantage of the fact that eigenpairs in the same band for close values of the quasimomentum are very close, too.

Finally, we have to define some parameters in order to use Algorithm 3. These parameters are: θ, which have already been introduced for Algorithm 2 to tune the marking strategy; an integer value $\max_{\text{it}}$ greater than 0, which sets the maximum number of refinements and which plays the same role of $\max_n$ in Algorithm 2; an initial mesh $\mathcal{T}_0$ on the primitive cell Ω; an initial mesh $\mathcal{G}_0$ on $\mathcal{K}_{\text{red}}$; another integer value $\max_m$ greater than 0; and finally a finite sequence $\{\text{tol}_s\}$ of length $\max_m$ of real values, where $0 < \text{tol}_{s+1} < \text{tol}_s < \cdots < \text{tol}_0$, which prescribe the wanted tolerance for the approximated band $\mathcal{C}_m$, for each iteration of stage A.

Algorithm 3 Efficient method to compute bands

$\mathcal{C}_{j,m} := \mathrm{Band}(\mathcal{G}_0, \mathcal{T}_0, \max_m, \{\mathrm{tol}_s\}, \theta, j, \max_{\mathrm{it}})$
for all $\kappa \in \mathcal{N}_0$ **do**
  $\mathcal{T}_0^\kappa := \mathcal{T}_0$
  $\mathcal{C}_{j,0}(\kappa) := 0$
end for
$m = 0$
repeat
  for all $\kappa \in \mathcal{N}_m$ **do**
    $(\lambda_{j,m+1}^\kappa, u_{j,m+1}^\kappa, \mathcal{T}_{m+1}^\kappa) = \mathrm{AFEM}(\kappa, \mathrm{FatherMesh}(\kappa), j, \theta, \mathrm{tol}_m, \max_{\mathrm{it}})$
    $\mathcal{C}_{j,m+1}(\kappa) := \lambda_{j,m+1}^\kappa$
  end for
  Refine the mesh $\mathcal{G}_m$ and construct $\mathcal{G}_{m+1}$
  $m = m + 1$
until $m \geq \max_m$

Algorithm 3 is convergent in the sense that, if its outer repeat-until loop is run infinitely many times, $\mathcal{C}_m$ will converge to the true band. To prove this statement we are going to suppose to be able to run Algorithm 3 with $\max_m = \infty$ and with tol_m values forming a strictly monotone decreasing sequence converging to 0, in this way the outer loop of Algorithm 3 becomes an infinite loop.

Let $\mathcal{W}_m$ be the finite dimensional space of elementwise linear functions on the mesh $\mathcal{G}_m$, and $\mathcal{W}_\infty$ the limit of $\mathcal{W}_m$ when m goes to infinity. The computed bands $\mathcal{C}_{j,m}$ in Algorithm 3 are function in $\mathcal{W}_m$ satisfying the relation

$$\mathcal{C}_{j,m+1}(\kappa) := \lambda_{j,m+1}^\kappa \ , \quad \forall \kappa \in \mathcal{N}_m \ .$$

The next lemma is straightforward, since it is an application of Theorem 5.10.

Lemma 6.1: For any $\kappa \in \mathcal{N}_\infty$ and under the assumption that the initial mesh $\mathcal{T}_0$ is fine enough as in Remark 5.11, we have that $\mathcal{C}_{j,m}(\kappa)$ converges to the true value $\mathcal{C}_j(\kappa)$.

Proof. Let's assume that m is the minimum value such that $\kappa \in \mathcal{N}_m$, then in view of Algorithm 3 with $\max_m = \infty$ we have that the subroutine AFEM is applied infinitely many times to the point κ. This is equivalent to apply Algorithm 1 to the point κ, then the convergence of $\mathcal{C}_{j,m}(\kappa) \equiv \lambda_{j,m}^\kappa$ to $\mathcal{C}_j(\kappa) \equiv \lambda_j^\kappa$ comes as a consequence of Theorem 5.10. ∎

Theorem 6.2 (Convergence to the true band): Under the assumption that the initial mesh $\mathcal{T}_0$ is fine enough as in Remark 5.11, we have that the sequence of computed bands $\mathcal{C}_{j,m}$ constructed by Algorithm 3 with $\max_m = \infty$ converges to the true band $\mathcal{C}$.

Proof. For any $\kappa \in \mathcal{N}_\infty$ let us denote by m' the minimum value such that $\kappa \in \mathcal{N}_{m'}$. Now, using Lemma 6.1, we have that the sequence formed by $\mathcal{C}_{j,m}(\kappa)$, for any $m \geq m'$, converges to $\mathcal{C}_j(\kappa)$ when m goes to infinity. So this implies that, for any $\kappa \in \mathcal{N}_\infty$, $\mathcal{C}_{j,m}(\kappa)$ converges to $\mathcal{C}_j(\kappa)$. Because the set of points $\mathcal{N}_\infty$ is dense in $\mathcal{K}_{\mathrm{red}}$, we conclude that $\mathcal{C}_{j,m}$ converges pointwise almost everywhere to $\mathcal{C}_j$. Furthermore, $\mathcal{C}_j$ is a continuous function, as well as all the functions in the sequence $\mathcal{C}_{j,m} \in \mathcal{W}_m$, so the pointwise convergence on a dense set of points is enough to imply the uniform convergence.

∎

The central result of this section comes straightforwardly from the previous theorem:

Theorem 6.3 (Convergence of the extrema): Under the assumption that the initial mesh $\mathcal{T}_0$ is fine enough as in Remark 5.11, we have that:

$$\lim_{m \to \infty} \max_{\kappa \in \mathcal{N}_m} \mathcal{C}_m \ = \ \max_{\kappa \in \mathcal{K}_{\mathrm{red}}} \mathcal{C} \ ,$$

and

$$\lim_{m\to\infty} \min_{\kappa\in\mathcal{N}_m} \mathcal{C}_m \;=\; \min_{\kappa\in\mathcal{K}_{\rm red}} \mathcal{C}\ .$$

7 Numerics

In this section we show some numerical results related to the TE case mode computed using Algorithm 2 and Algorithm 3. We assume that A is piecewise constant and $B = 1$, this leads typically to localized singularities in the gradient of the eigenfunctions at corner points of the interface between different values of A.

The results below are computed using linear elements. The algorithms have been implemented in Fortran90 with the auxiliary of ARPACK [29] to compute the eigenpairs and of the HSL library [37] to solve linear systems within the ARPACK solver.

7.1 TE case problem on periodic medium

We first consider the TE problem for a periodic medium with square inclusions. The unit cell is the unit square with a square inclusion of side 0.5 centered inside it. We choose A to take the value 1 inside the inclusion and the value 0.05 outside it. This is a realistic example, since expected jumps in ε of real PCs are of this order.

It has already been shown in [22] that the "modified" error estimator performs better than the "standard" one in terms of number of degrees of freedom (DOFs) versus $|\lambda_j - \lambda_{j,n}|$. Now we want to explore the convergence aspects of these a posteriori error estimators.

So, in order to understand the convergence rate of the method, we would like to test numerically if the decay of the error for eigenvalues can be well approximated by the following formula:

$$|\lambda_j - \lambda_{j,n}| \;\leq\; C\gamma^{2n}\ , \tag{7.1}$$

where n is the iterations counter in Algorithm 2 and C and γ are constants. This formula comes from [20] where it has been shown to hold for standard elliptic problems with discontinuous coefficients and using a slightly different adaptive algorithm, which is not oscillation-free. For a sequence of computed eigenpairs on a sequence of adapted meshes either using the "standard" or the "modified" error estimator we estimate numerically the quantity γ using the formula: $\gamma := \sqrt{|\lambda_j - \lambda_{j,n}|/|\lambda_j - \lambda_{j,n-1}|}$. In Table 1 and Table 2 we present the computed values of γ for both the "standard" and the "modified" error estimator for two different values of the quasimomentum κ. The approximated eigenvalues for both values of the quasimomentum belong to the second band in the spectrum, which is strictly positive for all values of quasimomentum. As can be seen the values of γ for the two error estimators are very similar, even if the "modified" one needs less DOFs to reach the same accuracy, compared to the "standard" one. This is another clue that the "modified" error estimator performs better.

	$\eta_{j,n}$			$\tilde{\eta}_{j,n}$		
n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	γ	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	γ
1	0.0584	400	-	0.0584	400	-
2	0.0543	434	0.9643	0.0425	486	0.8535
3	0.0414	535	0.8732	0.0330	640	0.8808
4	0.0314	728	0.8707	0.0231	900	0.8374
5	0.0232	1071	0.8597	0.0139	1356	0.7756
6	0.0155	1584	0.8183	0.0105	1772	0.8669
7	0.0103	2039	0.8140	0.0080	2406	0.8755
8	0.0083	2722	0.8963	0.0058	3437	0.8525
9	0.0064	3764	0.8785	0.0039	4958	0.8213
10	0.0049	5331	0.8784	0.0027	6458	0.8287
11	0.0028	7342	0.7568	0.0022	8358	0.8979
12	0.0022	9593	0.8820	0.0017	11101	0.8831
13	0.0018	12626	0.9012	0.0013	15277	0.8613
14	0.0014	16997	0.8859	0.0009	20688	0.8234
15	0.0011	22833	0.8836	0.0006	26334	0.8460
16	0.0006	29583	0.7715	0.0005	33218	0.9013
17	0.0005	37643	0.8830	0.0004	42896	0.8850
18	0.0004	48507	0.8947	0.0003	56756	0.8619
19	0.0003	63516	0.8911	0.0002	74796	0.8398

Tab. 1: Values of γ for $\kappa = (0, 0)$ with $\theta = 0.5$.

	$\eta_{j,n}$			$\tilde{\eta}_{j,n}$		
n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	γ	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	γ
1	0.0505	400	-	0.0505	400	-
2	0.0473	436	0.9685	0.0363	502	0.8483
3	0.0391	586	0.9086	0.0276	749	0.8721
4	0.0319	850	0.9041	0.0176	1092	0.7983
5	0.0244	1223	0.8747	0.0122	1398	0.8342
6	0.0158	1686	0.8032	0.0091	1910	0.8599
7	0.0090	2296	0.7576	0.0071	2818	0.8846
8	0.0082	3220	0.9517	0.0054	3816	0.8695
9	0.0071	4480	0.9333	0.0036	4984	0.8200
10	0.0057	5820	0.8969	0.0026	6228	0.8504
11	0.0040	7622	0.8357	0.0020	9134	0.8744
12	0.0025	9748	0.7924	0.0016	12505	0.8999
13	0.0022	12883	0.9247	0.0012	16186	0.8750
14	0.0019	17383	0.9478	0.0009	20162	0.8370
15	0.0016	22344	0.9168	0.0006	24200	0.8652
16	0.0012	27970	0.8726	0.0005	32822	0.8673
17	0.0008	35172	0.8195	0.0004	44932	0.9016
18	0.0006	42886	0.8216	0.0003	57553	0.8861
19	0.0005	55426	0.9354	0.0002	71512	0.8491

Tab. 2: Values of γ for $\kappa = (\pi, \pi)$ with $\theta = 0.5$.

In Figure 3 we depict the mesh coming from the fourth iteration of Algorithm 2 with $\theta = 0.5$. As can be seen the corners of the inclusion are much more refined than the rest of the domain. In Figure 4 we depict the eigenfunction corresponding to the eigenvalue in the second band of the problem with quasimomentum $\kappa = (0, 0)$.

7.2 TE mode problem on supercell

The spectra of periodic media are characterized by band gaps, but, for many applications, the employment of media with band gaps is not enough. Commonly it is necessary to create eigenvalues inside the gaps in the spectra of the media. The importance of these eigenvalues is due to the fact that electromagnetic waves, which have frequencies corresponding to these eigenvalues, may remain trapped inside the defects [14, 16] and they decay exponentially away from the defects. The common way to create such eigenvalues is by introducing a localized defect in the periodic structures [16, 15]. Such localized defects do not change the bands of the essential spectrum [15, Theorem 1].

In the next set of experiments we continue to work with the TE case problem and we shall use the "supercell method" [39] to compute the modes arising from the defect. The supercell method takes the defect problem (which is no longer periodic) and approximates it by a "nearby problem" in which the defect is surrounded by a finite number of layers of the original periodic medium, which is then truncated and repeated periodically, so that we get a new artificial periodic problem where each cell has a defect surrounded by some periodic layers.

We shall compute defect modes for the problem introduced in §7.1 using a supercell with two layers of periodic structure surrounding the defect, as in Figure 5. This new medium (since it is again infinitely periodic) could have new bands in its spectrum due to the defect. However it is also known that as the number of periodic layers increases, the bands in the gaps shrink exponentially quickly to the eigenvalues of the original defective material [39].

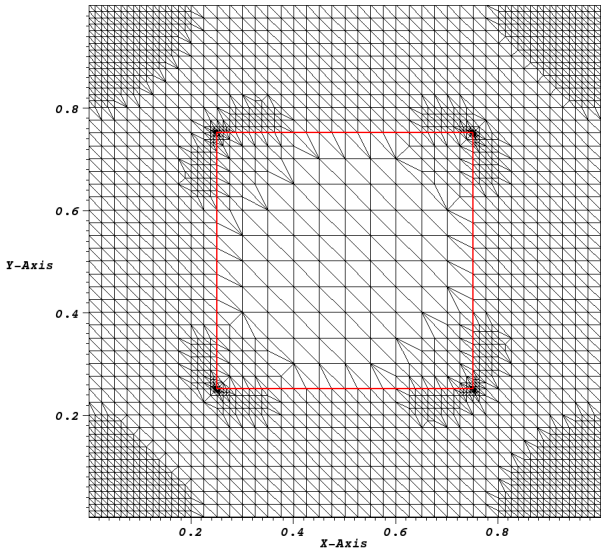

Fig. 3: A refined mesh coming from the adaptivity FEM for the TE mode problem with $\kappa = (0, 0)$ and using $\tilde{\eta}_{j,n}$.

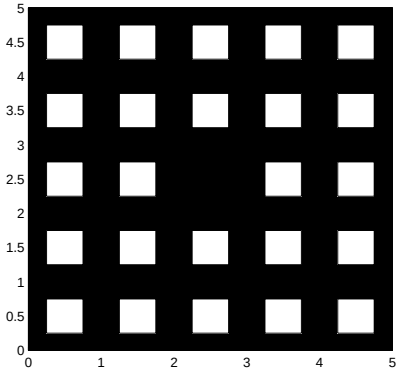

Fig. 5: The structure of the supercell used for the computations.

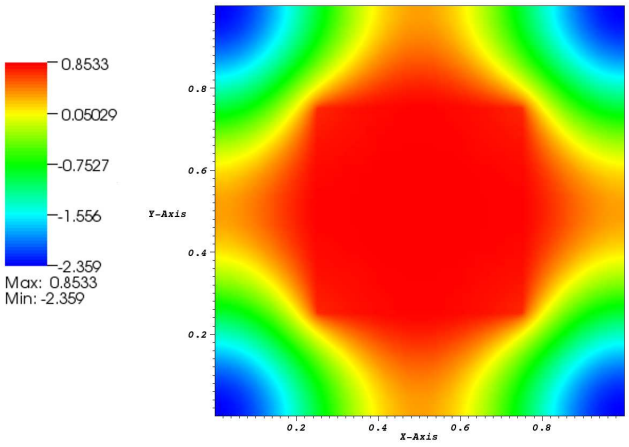

Fig. 4: The eigenfunction of the eigenvalue in the second band of the TE mode problem with quasi-momentum $\kappa = (0, 0)$.

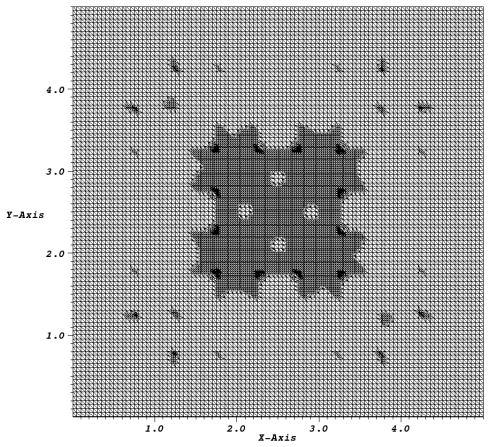

Fig. 6: An adapted mesh for a trapped eigenvalue of the TE case problem on a supercell with quasi-momentum $\kappa = (0, 0)$.

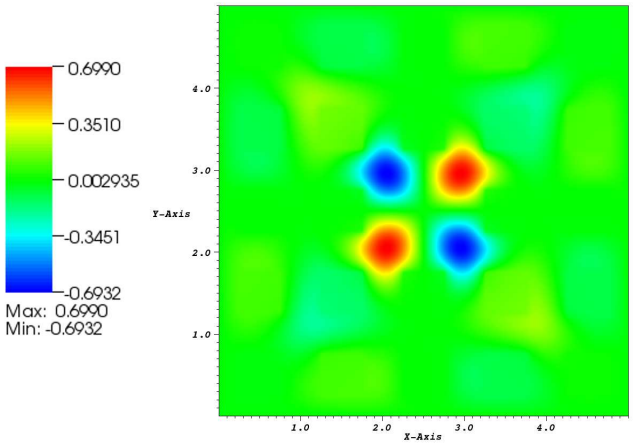

Fig. 7: A picture of the eigenfunction trapped in the defect of the TE case problem on a supercell with quasimomentum $\kappa = (0, 0)$.

We have already reported in [22] that also in the supercell setting the "modified" error estimator works better than the "standard" one in terms of number of degrees of freedom (DOFs) versus $|\lambda_j - \lambda_{j,n}|$, so the results below are mainly about the convergence of the method for trapped modes in supercell structures.

In Figure 6 we depict the mesh coming from the fourth iteration of Algorithm 2 with $\theta = 0.5$. As can be seen there is a lot of refinement around the defect, especially around the corners of the inclusions. Away from the defect there is just a bit of refinement which is again around the corners of the inclusions, the reason why the refinement is so concentrated in the defect and the reason why the corners of the inclusions away from the defect seem to not show important singularities, is because the trapped mode has a fast decay outside the defect that flattens down the singularities that it encounters, see Figure 7, where we depict the eigenfunction corresponding to the mode trapped inside the defect that we have computed.

Also in the supercell setting we can compute numerically the values of γ as in (7.1) in order to better understand the convergence of the method. The computed values are reported in Table 3 and Table 4 using both error estimators and for two values of the quasimomentum. As before, also in this setting, the values of γ for the two error estimators are very similar, even if the "modified" one needs less DOFs to reach the same accuracy, compared to the "standard" one. This is another clue that the "modified" error estimator performs better.

	$\eta_{j,n}$			$\tilde{\eta}_{j,n}$		
n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	γ	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	γ
1	0.0228	10000	-	0.0228	10000	-
2	0.0178	10180	0.8836	0.0162	10480	0.8432
3	0.0148	10873	0.9111	0.0126	11378	0.8814
4	0.0118	12116	0.8935	0.0091	12850	0.8487
5	0.0084	14160	0.8414	0.0065	14808	0.8453
6	0.0061	17128	0.8527	0.0048	19350	0.8592
7	0.0046	22784	0.8744	0.0037	25562	0.8734
8	0.0036	30364	0.8830	0.0027	33366	0.8627
9	0.0026	40791	0.8544	0.0019	43964	0.8364
10	0.0018	54071	0.8297	0.0013	59826	0.8351
11	0.0013	72860	0.8451	0.0010	79716	0.8864
12	0.0010	97840	0.8825	0.0008	105876	0.8661
13	0.0008	130455	0.8638	0.0006	137420	0.8583

Tab. 3: Values of γ for $\kappa = (0, 0)$ with $\theta = 0.5$ for the supercell problem.

	$\eta_{j,n}$			$\tilde{\eta}_{j,n}$		
n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	γ	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	γ
1	0.0164	10000	-	0.0164	10000	-
2	0.0143	10346	0.9345	0.0119	10620	0.8502
3	0.0140	10864	0.9882	0.0095	12101	0.8932
4	0.0122	11804	0.9358	0.0067	14221	0.8428
5	0.0106	13367	0.9320	0.0051	18184	0.8700
6	0.0091	15950	0.9265	0.0037	23310	0.8489
7	0.0080	19553	0.9354	0.0028	31076	0.8770
8	0.0066	24075	0.9077	0.0022	41046	0.8763
9	0.0053	31329	0.8982	0.0015	55164	0.8421
10	0.0044	39981	0.9149	0.0011	74861	0.8580
11	0.0032	50614	0.8430	0.0008	98964	0.8447
12	0.0027	66315	0.9179	0.0006	131051	0.8901
13	0.0022	84636	0.9125	0.0005	170567	0.8617

Tab. 4: Values of γ for $\kappa = (\pi/5, \pi/5)$ with $\theta = 0.5$ for the supercell problem.

7.3 Computation of the band of a trapped mode

Finally, we present some numerical results using Algorithm 3 to approximate the band of the trapped mode already analysed in this section. We are going to compare the results from Algorithm 3 against the results from Algorithm 2 applied to each considered value of the quasimomentum and always starting from the same mesh. In particular we are interested in comparing the computational costs of these two approaches in terms of number of mesh refinements #ref.

The starting mesh $\mathcal{G}_0$ for Algorithm 3 contains just one element (i.e., 3 nodes) as big as $\mathcal{K}_{\text{red}}$ for the considered supercell. In this numerical experiment we are going to construct just one refinement of $\mathcal{G}_0$, namely $\mathcal{G}_1$ which contains 10 nodes, i.e., we set $\max_m = 1$. Moreover, we set $\max_{\text{it}} = 20$ and $\theta = 0.5$ Also for Algorithm 2 applied to each node of $\mathcal{G}_1$ we are going to set $\max_{\text{it}} = 20$ and

$\theta = 0.5$. To make the comparison fair both algorithms start form the same mesh $\mathcal{T}_0$ and the same final tolerance for the residuals is set for both of them which are called tol_1 in Algorithm 3 and tol in Algorithm 2.

In Table 5 we collected the results for two different final tolerances of 0.3 and 0.15 which roughly correspond to $|\lambda_j^\kappa - \lambda_{j,n}^\kappa| = 0.005$ and $|\lambda_j^\kappa - \lambda_{j,n}^\kappa| = 0.001$ for all considered values of κ. The difference between the number of refinements is a clear indication of the efficiency of Algorithm 3.

Algorithm 3		**Standard adaptivity**	
$\{\mathrm{tol}_s\}$	#ref	tol	#ref
0.4 0.3	45	0.3	68
0.3 0.15	97	0.15	123

Tab. 5: Comparison between Algorithm 3 and the standard adaptive method, both applied to the band of the trapped mode for the TE case problem on a supercell.

References

[1] H. Ammari and F. Santosa, Guided waves in a photonic bandgap structure with a line defect, *SIAM J. Appl. Math.* **64** (2004) 2018–2033.

[2] N. W. Ashcroft, N. D. Mermin *Solid State Physics*, (Brooks/Cole, 1976).

[3] W. Axmann and P. Kuchment, An efficient finite element method for computing spectra of photonic and acoustic band-gap materials, *J. Comput. Physics* **150** (1999) 468–481.

[4] C. Bernardi and R.Verfürth, Adaptive finite element methods for elliptic equations with non-smooth coefficients, *Numer. Math.* **85** (2000) 579–608.

[5] D. Boffi, M. Conforti and L. Gastaldi, Modified edge finite elements for photonic crystals, *Numer. Math.* **105** (2006) 249–266.

[6] Y. Cao, Z. Hou and Y. Liu, Convergence problem of plane-wave expansion method for photonic crystals, *Physics Letters A* **327** (2004) 247–253.

[7] C. Carstensen and J. Gedicke, An oscillation-free adaptive FEM for symmetric eigenvalue problems, Preprint 489, DFG Research Center Matheon, Strasse des 17.Juni 136, D-10623 Berlin, 2008.

[8] S. J. Cox and D. C. Dobson, Maximizing band gaps in two-dimensional photonic crystals, *SIAM J. Appl. Math.* **59** (1999) 2108–2120.

[9] X. Dai, J. Xu and A. Zhou, Convergence and optimal complexity of adaptive finite element eigenvalue computations, *Numer. Math.* **110** (2008) 313–355.

[10] D. C. Dobson, An Efficient Method for Band Structure Calculations in 2D Photonic Crystals, *J. Comp. Phys.* **149** (1999) 363–376.

[11] D. C. Dobson, J. Gopalakrishnan and J. E. Pasciak, An efficient method for band structure calculations in 3D photonic crystals, *J. Comput. Phys.* **161** (2000) 668–679.

[12] C. Engström, C. Hafner and K. Schmidt, Computations of lossy Bloch waves in two-dimensional photonic crystals, *J. Comput. Theor. Nanosci.* **6** (2009) 775–783.

[13] A. Figotin and V. Goren, Resolvent method for computations of localized defect modes of H-polarization in two-dimensional photonic crystals, *Phys. Rev. E* **64** (2001) 1–16.

[14] A. Figotin and V. Gorentsveig, Localized electromagnetic waves in a layered periodic dielectric medium with a defect, *Phys. Rev. B* **58** (1998) 180–188.

[15] A. Figotin and A. Klein, Localized classical waves created by defects, *J. Stat. Phys.* **86** (1997) 165–177.

[16] A. Figotin and A. Klein, Midgap defect modes in dielectric and acoustic media, *SIAM J. Appl. Math.* **58** (1998) 1748–1773.

[17] E. M. Garau and P. Morin, Convergence and quasi-optimality of adaptive FEM for Steklov eigenvalue problems, *IMA J Numer Anal* (2010), to appear.

[18] E. M. Garau, P. Morin and C. Zuppa, Convergence of adaptive finite element methods for eigenvalue problems, *Math. Models Methods Appl. Sci.* **19** (2009) 721–747.

[19] S. Giani, *Convergence of adaptive finite element methods for elliptic eigenvalue problems with application to photonic crystals*, PhD Thesis, University of Bath, 2008.

[20] S. Giani and I. G. Graham, A convergent adaptive method for elliptic eigenvalue problems, *SIAM J. Numer. Anal.* **47** (2009) 1067–1091.

[21] S. Giani and I G. Graham, A convergent adaptive method for elliptic eigenvalue problems and numerical experiments, Bath Institute for Complex Systems Preprint number 14/08 , University of Bath, 2008.

[22] S. Giani and I. G. Graham, Adaptive finite element methods for computing band gaps in photonic crystals, *Numerische Mathematik*, submitted.

[23] B. Hiett, *Photonic Crystal modelling using finite element analysis*, PhD Thesis, University of Southampton, 2000.

[24] J. D. Joannopoulos and S. G. Johnson, Block-iterative frequency-domain methods for Maxwell's equations in a planewave basis, *Optics Express* **8** (2001) 173–190.

[25] J. Joannopoulos, S. Johnson, J. Winn, R. Meade, *Photonic Crystals: Molding the Flow of Light* (Princeton University Press, 2008).

[26] A. Klöckner, *On the computation of maximally localized Wannier functions*, PhD Thesis, Karlsruhe University, 2004.

[27] P. Kuchment, *Floquet Theory for Partial Differential Equations*, (Birkhauser Verlag, 1993).

[28] P. Kuchment, *The mathematics of photonic crystals*, SIAM, Frontiers Appl. Math. **22** (2001) 207–272.

[29] R. B. Lehoucq, D. C. Sorensen, and C. Yang, ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods, (SIAM, 1998).

[30] P. Morin, K. Siebert and A. Veeser, A basic convergence result for conforming adaptive finite elements *Math. Models Methods Appl. Sci.* **18** (2008) 707–737.

[31] R. A. Norton, *Numerical computation of band gaps in photonic crystal fibres*, University of Bath, 2008.

[32] R. Norton and R. Scheichl, Convergence analysis of planewave expansion methods for Schroedinger operators with discontinuous periodic potentials, *SIAM Journal on Numerical Analysis* **47** (2010) 4356–4380.

[33] G. J. Pearce, T. D. Hedley and D. M. Bird, Adaptive curvilinear coordinates in a plane-wave solution of Maxwell's equations in photonic crystals, *Physical Review B* **71** (2005), 2005.

[34] K. Sakoda, *Optical Properties of Photonic Crystals*, (Springer-Verlag, 2001).

[35] K. Schmidt and R. Kappeler, Efficient Computation of Photonic Crystal waveguide modes with dispersive material, *Opt. Express* **18** (2010) 7307–7322.

[36] K. Schmidt and P. Kauf, Computation of the band structure of two-dimensional photonic crystals with *hp* finite elements , *Comput Meth Appl Mech Eng* (2008), to appear.

[37] J. A. Scott, Sparse Direct Methods: An Introduction, *Lecture Notes in Physics* **535** (2000).

[38] R. L. Scott and S. Zhang, Finite element interpolation of nonsmooth functions satisfying boundary conditions, *Math Comp* **54** (1990) 483–493.

[39] S. Soussi, Convergence of the supercell method for defect modes calculations in photonic crystals, *SIAM J. Numer. Anal.* **43** (2005) 1175–1201.

[40] G. Strang and G. J. Fix, *An Analysis of the Finite Element Method* (Prentice-Hall, 1973).

[41] R. Verfürth, *A Review of a Posteriori Error Estimation and Adaptive Mesh Refinement Techniques* (Wiley-Teubner ,1996).