Numerische Mathematik manuscript No.
(will be inserted by the editor)

S. Giani · I.G. Graham

Adaptive finite element methods for computing band gaps in photonic crystals

the date of receipt and acceptance should be inserted later

Abstract In this paper we propose and analyse adaptive finite element methods for computing the band structure of 2D periodic photonic crystals. The problem can be reduced to the computation of the discrete spectra of each member of a family of periodic Hermitian eigenvalue problems on a unit cell, parametrised by a two-dimensional parameter - the quasimomentum. These eigenvalue problems involve non-coercive elliptic operators with generally discontinuous coefficients and are solved by adaptive finite elements. We propose an error estimator of residual type and show it is reliable and efficient for each eigenvalue problem in the family. In particular we prove that if the error estimator converges to zero then the distance of the computed eigenfunction from the true eigenspace also converges to zero and the computed eigenvalue converges to a true eigenvalue with double the rate. We also prove that if the distance of a computed sequence of approximate eigenfunctions from the true eigenspace approaches zero, then so must the error estimator. The results hold for eigenvalues of any multiplicity. We illustrate the benefits of the resulting adaptive method in practice, both for fully periodic structures and also for the computation of eigenvalues in the band gap of structures with defect, using the supercell method.

MSC2010 Subject Classification: 65M50, 65M60, 65F15

1 Introduction

Photonic crystals (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; see [26,38,30,2] for an introduction. In this paper we consider only 2D PCs, whose behaviour is periodic in the plane determined by two orthogonal directions, and is constant in the direction normal to this plane.

The propagation of light in any kind of PC is governed by Maxwell's equations. In 2D PCs, the 3D Maxwell's equations reduce to a two-dimensional one-component wave equation, which determines either the electric field or the magnetic field. Because the problem is periodic, the Floquet transform [30,29] can be applied to split each mode into a family of eigenvalue problems on a unit cell Ω of the periodic medium with periodic boundary conditions. This family is parameterised by the quasimomentum κ, which varies in the first Brillouin zone - for a definition see § 2. All eigenvalue problems in the family have the weak form: *seek eigenpairs of the form* $(\lambda, u) \in \mathbb{C} \times H^1_\pi(\Omega)$*, with* u *appropraitely normalised, such that*

$$\int_\Omega ((\nabla + i\kappa)v)^* A(\nabla + i\kappa)u \;=\; \lambda \int_\Omega Bu\bar{v} \quad \text{in } \Omega, \text{ for all } v \in H^1_\pi(\Omega), \tag{1.1}$$

S. Giani (Corresponding Author)
School of Mathematical Sciences, University of Nottingham, University Park, Nottingham, NG7 2RD, UK.
E-mail: Stefano.Giani@nottingham.ac.uk, Telephone:+44 (0) 115 84 67916 , Fax: +44 (0) 115 95 13837

I.G. Graham
Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK.
E-mail: I.G.Graham@bath.ac.uk

where Ω is the primitive cell of the photonic crystal and $H^1_\pi(\Omega)$ is the space all functions of $H^1(\Omega)$ satisfying periodic boundary conditions on $\partial\Omega$. Here, the (generally) matrix-valued function A is real symmetric and uniformly positive definite, i.e.,

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

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.,

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

We note that the eigenvalue problem, subject to the normalisation constraint on u, is a nonlinear problem for the unknown pair (λ, u).

In the theory in this paper 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. However the algorithm will still run even if these constraints are not satisfied 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"), for example [37,26,11,34,36]. This method exploits the periodicity in the PC and uses modern highly tuned FFT algorithms to obtain fast implementations. However the overall rate of convergence of approximate spectra to true spectra is slow because the jumps in the dielectric destroy the exponential accuracy which is achieved by Fourier spectral methods for smooth problems. Methods for accelerating the convergence by artificially smoothing the jumps in the dielectric have also been proposed. These converge quickly to a solution which contains a smoothing error and it turns out to be impossible to recover overall exponential accuracy by this method - see [34–36] for a complete analysis. Other spectral methods include [17] which uses an expansion in terms of eigenfunctions for the crystal without any defects. Semi-analytical methods which impose considerable limitations on the geometry of the crystal are also considered, for example, in [18].

We use adaptive finite element methods because they provide flexible solvers for PDE eigenvalue problems and are able to deal optimally with the heterogeneous media problems encountered in PC models. There are already a number of papers about low order finite element methods for PCs [4,10,14,15,24,28] and most recently there has been considerable interest in p and hp methods, with the latter having the potential to obtain exponential accuracy [16,32,39,40] . Accurate computations based on *a priori hp* refinement strategies are shown in [39,40]. However, as far as we are aware, until now no one has used adaptivity based on *a posteriori* error estimates on these problems.

Mesh adaptivity based on *a posteriori* error estimates has been widely used to improve the accuracy of numerical solutions of PDEs (e.g. [1]). Recently the question of convergence of h-adaptive methods for elliptic eigenvalue problems has received intensive interest. One of the first proofs was in [22], but this is only for eigenproblems based on coercive bilinear forms. As we shall see the Hermitian form on the left-hand side of the PC eigenvalue problem (1.1) is not coercive for all values of the quasimomentum κ, so new methods of analysis are required. Some of the methods presented in this paper were first developed in the PhD thesis [21], where the convergence of adaptive methods for PCs was also discussed. Some previous numerical experiments were reported in [23]. Recently there is much interest in adaptive methods for PDE eigenvalue problems in general - see for example [12,33] for other applications.

The outline of the paper is as follows. The next section - §2 - briefly describes how problem (1.1) is derived from Maxwell's equations. Here we also prove some basic properties of the Hermitian form in (1.1) and we introduce the finite element discretization. Then §3 proves some basic a priori estimates for finite element approximation of PC eigenvalue problems. These are derived from the classical literature and are essential for the main results of this paper which are contained in §§4 and 5. To give a flavour of the main results, let $(\lambda_{j,n}, u_{j,n})$ denote a computed finite element eigenpair of (1.1) (where $u_{j,n}$ is a finite element function and $\lambda_{j,n}$ approximates a true eigenvalue λ_j of arbitrary multiplicity), then in Definition 4.3 we define an a posteriori error estimator $\eta_{j,n}$ (being a sum of computable contributions from each mesh element), and in Theorems 4.7 and 4.9 we prove that

$$\text{dist}(u_{j,n}, E_1(\lambda_j)) \leq C\eta_{j,n} \quad \text{and} \quad |\lambda_{j,n} - \lambda_j| \leq C\eta^2_{j,n} \,, \tag{1.4}$$

with C independent of the mesh, where $E_1(\lambda_j)$ denotes the unit ball in the exact eigenspace corresponding to λ_j and the distance is measured in an energy inner product related to the Hermitian form in (1.1) (see Lemma 2.1). Recalling that nonlinearity of the eigenvalue problem (1.1), it is not surprising that elementary a posteriori error estimates usually involve additional terms on the right hand side. However, due to the a priori results in §3 these are rigorously shown to be of higher order and so do not appear in our estimates.

By (1.4), the eigenfunction and eigenvalue error both approach zero if the estimator $\eta_{j,n} \to 0$. The converse is established in §5, i.e., if the eigenfunction and eigenvalue errors both converge to zero, then so does the error estimate $\eta_{j,n}$. (This is known as "efficiency".) Finally, numerical experiments illustrating the results with our method, compared to more standard FEM methods, are collected in §6. 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 the analysis of adaptive finite element methods for PC applications.

2 Photonic crystal eigenvalue problem and numerical method

In general, PCs are of practical interest because of their band gap properties - i.e., monochromatic electromagnetic waves of certain frequencies may not propagate inside them. Since fabrication is simpler in 2D than in 3D and since the 2D case still includes many important applications, (e.g., [27]), considerable numerical interest has focussed on the 2D case - e.g. [4,11,14,17,32,34,39,40] - and the present paper obtains the first rigorous theory for adaptive finite element methods in this case.

The mathematical development (see e.g. [30]) begins with the eigenvalue problem for Maxwell's equations

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

where $\mathbf{E}_\omega$ is the electric field, $\mathbf{H}_\omega$ is the magnetic field, ε and μ are, respectively, the dielectric permittivity and magnetic permeability tensors, and c is the speed of light in a vacuum. We assume the medium is periodic in the (x, y) plane and is constant in the third (z) direction and that the material is non-magnetic (so $\mu = 1$). The problem (2.1) splits naturally into two independent problems, called transverse magnetic (TM) and transverse electric (TE) modes, as explained in [30]. On the assumption that the medium is isotropic (so ε is scalar-valued), the problems are

$$\Delta u_\omega + \frac{\omega^2}{c^2}\varepsilon u_\omega = 0 \quad \text{(TM case)} \ , \tag{2.2}$$

and

$$\nabla \cdot \frac{1}{\varepsilon}(\nabla u_\omega) + \frac{\omega^2}{c^2} u_\omega = 0, \quad \text{(TE case)} \ . \tag{2.3}$$

Both problems (2.2) and (2.3) may be written in the abstract form as that of seeking (λ, u) with $u \neq 0$ such that

$$\nabla \cdot (A \nabla u) + \lambda B u \ = \ 0 \ . \tag{2.4}$$

The anisotropic case (where ε is a tensor) may also be included in this formulation - see e.g. [32]. Since A or B may be discontinuous, (2.4) has to be understood in an appropriate weak form. So far (2.4) is posed over all of $\mathbb{R}^2$, with periodic data.

A 2D periodic medium can be described using a lattice $L := \{\mathbf{R} = n_1 \mathbf{r}_1 + n_2 \mathbf{r}_2 \ , \ n_1, n_2 \in \mathbb{Z}\}$, where $\{\mathbf{r}_1, \mathbf{r}_2\}$ is a basis for $\mathbb{R}^2$. The *(Wigner-Seitz) primitive cell* for L is the set Ω of all points in $\mathbb{R}^2$ which are closer to $\mathbf{0}$ than to any other point in L - see [3]. When $\overline{\Omega}$ is translated through all $\mathbf{R} \in L$, we obtain a covering of $\mathbb{R}^2$ with overlap of measure 0. The *reciprocal lattice* for L is the lattice $\hat{L}$ generated by a basis $\{\mathbf{k}_1, \mathbf{k}_2\}$, chosen so that $\mathbf{r}_i \cdot \mathbf{k}_j = 2\pi\delta_{i,j}$, $i, j = 1, 2$, where $\delta_{i,j}$ is the Kronecker delta and the primitive cell for the reciprocal lattice is called the *first Brillouin zone*, which we denote here by $\mathcal{K}$ [3].

For example, if L is the square lattice generated by $\{\mathbf{e}_1, \mathbf{e_2}\}$ (where $\mathbf{e}_i$ are the standard basis functions in $\mathbb{R}^2$), then $\Omega = [-0.5, 0.5]^2$, $\hat{L}$ is generated by $\{2\pi\mathbf{e}_1, 2\pi\mathbf{e_2}\}$ and the first Brillouin zone is $\mathcal{K} = [-\pi, +\pi]^2$. Such square lattices are used in all numerical experiments in Section 6.

The Floquet transform - see, e.g. [30] - may them be used to show the equivalence of the problem (2.4) to a family of problems on the primitive cell Ω parametrized by quasimomentum $\kappa \in \mathcal{K}$. This is the family

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

where $\tilde{u}$ is the Floquet transform of u and λ is the corresponding eigenvalue which now depends on κ. This equation should again be understood in the weak form - a rigorous derivation can be found for example in [9]. In order to recover the spectrum of the problem (2.4), it is sufficient to compute the union of all the spectra of the problems in the family (2.5) for all $\kappa \in \mathcal{K}$, and these problems have discrete spectrum since the domain Ω is compact. For more details see [30, page 19]. Writing (2.5) in weak form gives precisely (1.1).

Throughout $L^2(\Omega)$ denotes the usual space of square integrable complex valued functions equipped with the weighted norm

$$\|f\|_{0,B} \ = \ b(f, f)^{1/2} \ , \quad b(f, g) := \int_\Omega B f \bar{g} \ . \tag{2.6}$$

$H^1(\Omega)$ denotes the usual space of functions in $L^2(\Omega)$ with square integrable gradient, with H^1-norm denoted $\|f\|_1$, and $H^1_\pi(\Omega)$ denotes the subspace of functions in $f \in H^1(\Omega)$ which satisfy periodic boundary conditions on $\partial\Omega$. We will also need the fractional order spaces $H^{1+s}(\Omega)$, $s \in [0,1]$. When we want to restrict these norms to a measurable subset $S \subseteq \Omega$, we write $\|f\|_{0,B,S}$, $\|f\|_{1,S}$, etc.

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*

$$\left.\begin{array}{ll} a_\kappa(u_j, v) = \lambda_j\ b(u_j, v)\ , & \text{for all} \quad v \in H^1_\pi(\Omega) \\ \|u_j\|_{0,B} \ = 1 & \end{array}\right\} \tag{2.7}$$

where

$$a_\kappa(u,v) := \int_\Omega ((\nabla + i\kappa)v(x))^* A(x)((\nabla + i\kappa)u(x))\ . \tag{2.8}$$

It is easy to see that a_κ is a Hermitian form on $H^1_\pi(\Omega)$. which is bounded on $H^1(\Omega)$ independently of $\kappa \in \mathcal{K}$. Moreover by the positive definiteness of A assumed in (1.2), we have

$$a_\kappa(u,u) \ \geq\ \underline{a} \int_\Omega |(\nabla + i\kappa)u|^2 \geq 0\ , \quad \text{for all } u \in H^1_\pi(\Omega)\ . \tag{2.9}$$

Thus the spectrum of (2.7) is real and non-negative

However $a_\kappa(u,u)$ is not always strictly positive (for $u \neq 0$), since if $\kappa = (0,0)$ then $a_\kappa(1,1) = 0$. Thus we introduce the shifted Hermitian form:

$$(u,v)_{\kappa,A,B} \ :=\ a_\kappa(u,v) + \ \sigma\ b(u,v)\ , \quad \text{for all } u,v \in H^1_\pi(\Omega)\ , \tag{2.10}$$

with a fixed shift

$$\sigma := \max_{\kappa\in\mathcal{K}} |\kappa|^2 \underline{a}/\underline{b}\ +\ 1\ . \tag{2.11}$$

As the following result shows, this shifted form is coercive on $H^1_\pi(\Omega)$ (i.e., $(u,u)_{\kappa,A,B}/\|u\|_1^2$ is bounded below by a positive constant for all $u \in H^1_\pi(\Omega)$). This shifted form is used in the theory below, but is never used in computations.

Lemma 2.1 $(\cdot,\cdot)_{\kappa,A,B}$ *is an inner product on* $H^1_\pi(\Omega)$ *and we denote the induced norm by* $\|\cdot\|_{\kappa,A,B}$,

Proof. We shall show that

$$\|u\|^2_{\kappa,A,B} \ =\ (u,u)_{\kappa,A,B} \ \geq\ c_a \|u\|_1^2, \quad \text{for all } \kappa \in \mathcal{K}, \quad u \in H^1_\pi(\Omega)\ , \tag{2.12}$$

when $c_a = \min\{\underline{a}/2, \underline{b}\}$. Since $(\cdot,\cdot)_\kappa$ is a Hermitian form on $H^1_\pi(\Omega)$, this proves the result.

By definition of $a_\kappa(\cdot,\cdot)$, we have:

$$\begin{aligned} a_\kappa(u,u) &= \int_\Omega ((\nabla u)^* A \nabla u) \ \ +\ \ (\kappa^T A \kappa)|u|^2 \ \ +\ \ i\{((\nabla u)^* A\kappa)u - (\kappa^T A \nabla u)\overline{u}\} \\ &= \int_\Omega (\nabla u)^* A \nabla u \ \ +\ \ (\kappa^T A\kappa)|u|^2 \ \ -\ \ 2\,\mathrm{Im}\{((\nabla u)^* A\kappa)u\}\ . \end{aligned}$$

It is straightforward to show that

$$\mathrm{Im}\,\{((\nabla u)^* A\kappa)u\} \ \ \leq\ |(\nabla u)^* A\kappa|\ |u| \ \leq\ \{(\nabla u)^* A \nabla u\}^{1/2}\ \left\{\kappa^T A\kappa\right\}^{1/2}\ |u|\ ,$$

and by an application of Cauchy-Schwarz in $L_2(\Omega)$ we obtain

$$\int_\Omega \mathrm{Im}\,\{((\nabla u)^* A\kappa)u\} \ \ \leq\ \left\{\int_\Omega (\nabla u)^* A\nabla u\right\}^{1/2}\ \left\{\int_\Omega (\kappa^T A\kappa)\ |u|^2\right\}^{1/2}\ .$$

Thus calling $\alpha = \left\{\int_\Omega \nabla u^* A \nabla u\right\}^{1/2}$, and $\beta = \left\{\int_\Omega (\kappa^T A\kappa)|u|^2\right\}^{1/2}$ we have from the arithmetic-geometric mean inequality, i.e $2\alpha\beta \leq \delta\alpha^2 + \delta^{-1}\beta^2$, that for any $\delta \in (0,1)$

$$a_\kappa(u,u) \ \geq\ \alpha^2 + \beta^2 - 2\alpha\beta \ \geq\ (1-\delta)\alpha^2 + (1-\delta^{-1})\beta^2$$

Hence, for any $\sigma \in \mathbb{R}$ we have

$$\begin{aligned} a_\kappa(u,u) \ +\ \sigma\ b(u,u) \geq\ & (1-\delta)\underline{a}\ |u|_1^2 + \left((1-\delta^{-1})\underline{a}|\kappa|^2 + \sigma\underline{b}\right)\ \|u\|_0^2 \\ & \geq \min\{(1-\delta)\underline{a}, (1-\delta^{-1})\underline{a}|\kappa|^2 + \sigma\underline{b}\}\|u\|_1^2\ . \end{aligned}$$

Now choosing $\delta = 1/2$ and since $\sigma = \underline{a}\max_{\kappa\in\mathcal{K}}|\kappa|^2/\underline{b}+1$ we see that

$$\min\{(1-\delta)\underline{a},(1-\delta^{-1})\underline{a}|\kappa|^2+\sigma\underline{b}\} = \min\{\underline{a}/2,-\underline{a}|\kappa|^2+\sigma\underline{b}\} \geq \min\{\underline{a}/2,\underline{b}\} = c_a .$$

■

Now, to discretize (2.7), let $\mathcal{T}_n$, $n = 1, 2, \ldots$ denote a family of conforming, shape-regular (see, e.g., [1]) and periodic triangular meshes on Ω. These meshes may be computed adaptively. With H_τ denoting the diameter of element τ, we define $H_n^{\max} := \max_{\tau\in\mathcal{T}_n}\{H_\tau\}$. On any mesh $\mathcal{T}_n$ we denote by $V_n \subset H^1_\pi(\Omega)$ the finite dimensional space of continuous functions which are affine on each element $\tau \in \mathcal{T}_n$. The discrete formulation of problem (2.7) is: *seek eigenpairs of the form* $(\lambda_{j,n}, u_{j,n}) \in \mathbb{R}\times V_n$ *such that*

$$\left.\begin{array}{ll} a_\kappa(u_{j,n},v_n) = \lambda_{j,n}\, b(u_{j,n},v_n)\ , & \text{for all} \quad v_n \in V_n \\ \|u_{j,n}\|_{0,B} = 1 & \end{array}\right\} \tag{2.13}$$

3 A priori convergence results

In this section we gather together some a priori estimates for PC eigenvalue problems. These results are mostly classical so we only give a few details for results which are not easily found in the literature. Suitable references are [5–7, 44]. With the shift σ from (2.11), the shifted versions of problems (2.7) and (2.13) are:
Seek eigenpairs of the form $(\zeta_j, u_j) \in \mathbb{R}\times H^1_\pi(\Omega)$ *such that*

$$\left.\begin{array}{ll} a_\kappa(u_j,v) + \sigma\, b(u_j,v) = \zeta_j\, b(u_j,v)\ , & \text{for all} \quad v \in H^1_\pi(\Omega) \\ \|u_j\|_{0,B} = 1\ ; & \end{array}\right\} \tag{3.1}$$

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

$$\left.\begin{array}{ll} a_\kappa(u_{j,n},v_n) + \sigma\, b(u_{j,n},v_n) = \zeta_{j,n}\, b(u_{j,n},v_n)\ , & \text{for all} \quad v_n \in V_n \\ \|u_{j,n}\|_{0,B} = 1\ . & \end{array}\right\} \tag{3.2}$$

The following proposition is self-evident:

Proposition 3.1 *The eigenpairs of* (2.7) *and* (3.1) *are in one-one correspondence. In fact,* (u_j, λ_j) *is an eigenpair of* (2.7) *if and only if* (u_j, ζ_j)*, with* $\zeta_j = \lambda_j + \sigma$*, is an eigenpair of* (3.1)*. Similarly* $(u_{j,n}, \lambda_{j,n})$ *is an eigenpair of* (2.13) *if and only if* $(u_{j,n}, \zeta_{j,n})$*, with* $\zeta_{j,n} = \lambda_{j,n} + \sigma$*, is an eigenpair of* (3.2)*.*

It follows from Lemma 2.1 that all eigenvalues of (3.1) and all $N = \dim V_n$ eigenvalues of (3.2) are positive. We can order them as $0 < \zeta_1 \leq \zeta_2 \ldots$ and $0 < \zeta_{1,n} \leq \zeta_{2,n} \ldots \leq \lambda_{N,n}$. Moreover, we know (e.g. [6]) that $\zeta_{j,n} \to \zeta_j$, for any j, as $H_n^{\max} \to 0$ and (by the minimax principle) that $\zeta_{j,n}$ is monotone non-increasing, i.e.

$$\zeta_{j,n} \geq \zeta_{j,m} \geq \zeta_j\ , \quad \text{for all} \quad j = 1,\ldots,N, \quad \text{and all} \quad m \geq n\ . \tag{3.3}$$

Hence $\lambda_{j,n} \to \lambda_j$, for any j, as $H_n^{\max} \to 0$ and

$$\lambda_{j,n} \geq \lambda_{j,m} \geq \lambda_j\ , \quad \text{for all} \quad j = 1,\ldots,N, \quad \text{and all} \quad m \geq n\ . \tag{3.4}$$

Let u_j and $u_{j,n}$ be any normalised eigenvectors of (2.7) and (2.13). Then

$$\begin{aligned} a_\kappa(u_j-u_{j,n},u_j-u_{j,n}) &= a_\kappa(u_j,u_j)+a_\kappa(u_{j,n},u_{j,n})-2\mathrm{Re}\{a_\kappa(u_j,u_{j,n})\} \\ &= \lambda_j+\lambda_{j,n}-2\lambda_j\ \mathrm{Re}\{b(u_j,u_{j,n})\} \\ &= (\lambda_{j,n}-\lambda_j)+2\lambda_j\ (1-\mathrm{Re}\{b(u_j,u_{j,n})\}) \\ &= (\lambda_{j,n}-\lambda_j)+\lambda_j\ b(u_j-u_{j,n},u_j-u_{j,n})\ . \end{aligned} \tag{3.5}$$

Combining this with (3.4), we obtain

$$a_\kappa(u_j-u_{j,n},u_j-u_{j,n}) = |a_\kappa(u_j-u_{j,n},u_j-u_{j,n})| = |\lambda_j-\lambda_{j,n}| + \lambda_j\ \|u_j-u_{j,n}\|^2_{0,B}\ . \tag{3.6}$$

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 3.2 *Given a function $v \in L^2(\Omega)$ and a finite dimensional subspace $\mathcal{P} \subset L^2(\Omega)$, 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 Lemma 2.1.

Now let λ_j be any eigenvalue of (2.7), let $E(\lambda_j)$ denote the (finite dimensional) space spanned by the eigenfunctions of λ_j and set $E_1(\lambda_j) = \{u \in E(\lambda_j) : \|u\|_{0,B} = 1\}$. Let T_{λ_j} denote the orthogonal projection of H^1_π onto $E(\lambda_j)$ with respect to the inner product $(\cdot,\cdot)_{\kappa,A,B}$ defined in (2.10).

Lemma 3.3 *Let $(\lambda_{j,n}, u_{j,n})$ be an eigenpair of* (2.13)*. Then*

$$\|u_{j,n} - u_j\|_{0,B} = \operatorname{dist}(u_{j,n}, E_1(\lambda_j))_{0,B}\ , \tag{3.7}$$

if and only if

$$\|u_{j,n} - u_j\|_{\kappa,A,B} = \operatorname{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B}\ . \tag{3.8}$$

Proof. Since $E(\lambda_j)$ is finite dimensional, the minimizers in (3.7) and (3.8) exist. Moreover

$$0 \;=\; (T_{\lambda_j} w, (I-T_{\lambda_j})v)_{\kappa,A,B} \;=\; (\lambda_j+\sigma)\, b(T_{\lambda_j} w, (I-T_{\lambda_j})v) \quad \text{for all} \quad v,w \in L^2_B(\Omega)\cap H^1_\pi(\Omega)\ . \tag{3.9}$$

Hence for any $v_j \in E(\lambda_j)$ we have the decomposition

$$u_{j,n} - v_j \;=\; (I-T_{\lambda_j})u_{j,n} \;+\; T_{\lambda_j}(u_{j,n}-v_j) \;=\; (I-T_{\lambda_j})u_{j,n} \;+\; (T_{\lambda_j}u_{j,n} - v_j)\ ,$$

which is orthogonal both with respect to $(\cdot,\cdot)_{\kappa,A,B}$ and $(\cdot,\cdot)_{0,B}$. Thus

$$\begin{aligned}
\|u_{j,n}-v_j\|^2_{0,B} \;&=\; \|(I-T_{\lambda_j})u_{j,n}\|^2_{0,B} + \|T_{\lambda_j}u_{j,n}-v_j\|^2_{0,B}\ ,\\
\|u_{j,n}-v_j\|^2_{\kappa,A,B} \;&=\; \|(I-T_{\lambda_j})u_{j,n}\|^2_{\kappa,A,B} + \|T_{\lambda_j}u_{j,n}-v_j\|^2_{\kappa,A,B}\ .
\end{aligned}$$

Hence u_j satisfies (3.8) if and only if it minimizes $\|T_{\lambda_j}u_{j,n} - v_j\|^2_{\kappa,A,B}$. The latter quantity is equal to $(\lambda_j-\sigma)\|T_{\lambda_j}u_{j,n}-v_j\|^2_{0,B}$ and hence u_j satisfies (3.8) if and only if it satisfies (3.7). ∎

In order to make further progress we need some assumption on regularity of solutions of elliptic problems associated with $(\cdot,\cdot)_{\kappa,A,B}$.

Assumption 3.4 *We assume that there exists a constant $C_{\rm ell} > 0$ and $s \in (0,1]$ with the following property. For $f \in L^2(\Omega)$, if $v := \mathcal{S}f \in H^1_\pi(\Omega)$ solves the problem $(v,w)_{\kappa,A,B} = b(f,w)$ for all $w \in H^1_\pi(\Omega)$, then*

$$\|\mathcal{S}f\|_{1+s} \le C_{\rm ell}\|f\|_{0,B}\ , \tag{3.10}$$

where $\|\cdot\|_{1+s}$ is the norm in the Sobolev space $H^{1+s}(\Omega)$.

This is a standard assumption which is satisfied in a wide number of applications such as problems with discontinuous coefficients (see eg. [22] for more references).

From now on we shall let C denote a generic constant which may depend on the true eigenvalues and vectors of (2.7) and other constants introduced above, but is always independent of n.

Theorem 3.5 *Suppose $1 \le j \le \dim V_n$. Let λ_j be an eigenvalue of* (2.7) *with corresponding eigenspace $E(\lambda_j)$ of any (finite) dimension and let $(\lambda_{j,n}, u_{j,n})$ be an eigenpair of* (2.13)*. Then, for $H_n^{\max}$ sufficiently small,*

(i)

$$|\lambda_j - \lambda_{j,n}| \;\le\; (\operatorname{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B})^2; \quad \textit{and} \quad |\lambda_j-\lambda_{j,n}| \;\le\; C(H_n^{\max})^{2s}; \tag{3.11}$$

(ii)

$$\operatorname{dist}(u_{j,n}, E_1(\lambda_j))_{0,B} \;\le\; C(H_n^{\max})^s \operatorname{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B}\ ; \tag{3.12}$$

(iii)

$$\operatorname{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B} \;\le C(H_n^{\max})^s\ . \tag{3.13}$$

Proof. First consider part (i). Since $\lambda_j \geq 0$ and $\sigma > 0$, the first estimate in (3.11) follows directly from (3.6). To obtain the second estimate in (3.11), we recall a standard error estimate for elliptic eigenvalues (see e.g. [6, (1.1)]) which, appled to problems (3.1) and (3.2), gives

$$\lambda_{j,n} - \lambda_j \;=\; (\lambda_{j,n} + \sigma) - (\lambda_j + \sigma) \;\leq\; C \sup_{u \in E_1(\lambda_j)} \inf_{v_n \in V_n} \|u - v_n\|^2_{\kappa,A,B} \;.$$

Combining this with standard finite element error estimates and recalling (3.4), we get

$$|\lambda_{j,n} - \lambda_j| \quad \leq\; C(H_n^{\max})^{2s} \sup_{u \in E_1(\lambda_j)} \|u\|^2_{1+s}, \tag{3.14}$$

For $u \in E_1(\lambda_j)$, Assumption 3.4 implies $\|u\|_{1+s} \;\leq\; C_{ell}(\lambda_j + \sigma)\|u\|_{0,B} \;\leq\; C_{ell}(\lambda_j + \sigma)$, which yields the result.

To obtain (ii), we use the following estimate [6, (3.31a)]:

$$\frac{\|T_{\lambda_j} u_{j,n} - u_{j,n}\|_{0,B}}{\|T_{\lambda_j} u_{j,n} - u_{j,n}\|_{\kappa,A,B}} \;\leq\; C\eta_n \;, \quad \text{where} \quad \eta_n \;=\; \sup_{\substack{g \in L^2(\Omega) \\ \|g\|_{0,B}=1}} \inf_{\chi \in V_n} \|\mathcal{S}g - \chi\|_{\kappa,A,B} \;, \tag{3.15}$$

and $\mathcal{S}$ is the solution operator defined in Assumption 3.4. Analogously to (3.14) we have $\eta_n \leq C(H_n^{\max})^s$ and hence (3.15) implies

$$\begin{aligned} \|T_{\lambda_j} u_{j,n} - u_{j,n}\|_{0,B} \;&\leq\; C(H_n^{\max})^s \|T_{\lambda_j} u_{j,n} - u_{j,n}\|_{\kappa,A,B} \\ &=\; C(H_n^{\max})^s \mathrm{dist}(u_{j,n}, E(\lambda_j))_{\kappa,A,B} \\ &\leq\; C(H_n^{\max})^s \mathrm{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B} \;, \end{aligned} \tag{3.16}$$

where we used the inclusion $E_1(\lambda_j) \subset E(\lambda_j)$. Since $\|u_{j,n}\|_{0,B} = 1$, (3.16) also implies that

$$\begin{aligned} \Big| \|T_{\lambda_j} u_{j,n}\|_{0,B} - 1 \Big| \;&\leq\; \|T_{\lambda_j} u_{j,n} - u_{j,n}\|_{0,B} \\ &\leq\; C(H_n^{\max})^s \mathrm{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B} \;. \end{aligned} \tag{3.17}$$

Combining (3.16) and (3.17), we obtain

$$\begin{aligned} \mathrm{dist}(u_{j,n}, E_1(\lambda_j))_{0,B} \;&\leq\; \left\| \frac{T_{\lambda_j} u_{j,n}}{\|T_{\lambda_j} u_{j,n}\|_{0,B}} - u_{j,n} \right\|_{0,B} \\ &\leq\; \Big\| T_{\lambda_j} u_{j,n} - u_{j,n} \Big\|_{0,B} + \Big| 1 - \|T_{\lambda_j} u_{j,n}\|^{-1}_{0,B} \Big| \; \|T_{\lambda_j} u_{j,n}\|_{0,B} \\ &=\; \Big\| T_{\lambda_j} u_{j,n} - u_{j,n} \Big\|_{0,B} + \Big| \|T_{\lambda_j} u_{j,n}\|_{0,B} - 1 \Big| \\ &\leq\; C(H_n^{\max})^s \mathrm{dist}(u_{j,n} \,, E_1(\lambda_j))_{\kappa,A,B} \;. \end{aligned}$$

which is (3.12).

Finally, for part (iii), we note that (3.6), Lemma 3.3 and (3.11) imply ,

$$\mathrm{dist}(u_{j,n}, E_1(\lambda_j))^2_{\kappa,A,B} \;\leq\; C(H_n^{\max})^{2s} \;+\; \lambda_j \, \mathrm{dist}(u_{j,n}, E_1(\lambda_j))^2_{0,B} \tag{3.18}$$

which, via (3.12), implies (3.13).

■

4 A posteriori error estimator and reliability

Our a posteriori error estimator is presented in (4.1) below. Its most important characteristics are *reliability* and *efficiency*. In broad terms reliability means that the ratio of the actual error to the error estimator is bounded above by a positive constant independent of the mesh, while efficiency means that this ratio is bounded below by a positive constant independent of the mesh. We prove reliability and efficiency for (4.1) in this and the following sections.

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

The residual estimator $\eta_{j,n}$ is defined as a sum of norms of element residuals and edge residuals, which are all computable quantities. We denote by $\mathcal{F}_n$ the set of all the edges (including boundary edges) of the elements of the mesh $\mathcal{T}_n$. For $f \in \mathcal{F}_n$, we denote by $\tau_1(f)$ and $\tau_2(f)$, the two elements sharing $f \in \mathcal{F}_n$ and we let H_f denote the length of f. We let $\mathbf{n}_f$ denote the unit normal on the edge f, which is assumed to point from $\tau_1(f)$ into $\tau_2(f)$. To simplify the notation, we define the functional $[\cdot]_f$ as follows

Definition 4.2 *We can define for any function $g : \Omega \to \mathbb{C}$ which is continuous on each element of the mesh $\mathcal{T}_n$ and for any $f \in \mathcal{F}_n$*

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

Definition 4.3 (Residual Estimator) *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:*

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

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

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

(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:

$$\eta_{j,n} := \left\{ \sum_{\tau\in\mathcal{T}_n} H_\tau^2 \|R_I(u_{j,n},\lambda_{j,n})\|_{0,\tau}^2 \;+\; \sum_{f\in\mathcal{F}_n} H_f \|R_F(u_{j,n})\|_{0,f}^2 \right\}^{1/2} . \tag{4.1}$$

In Theorem 4.8 we prove reliability of the estimator $\eta_{j,n}$ for eigenfunctions, and in Theorem 4.9 we prove reliability of the estimator $\eta_{j,n}^2$ for eigenvalues. (The appearance of the square in the latter estimator reflects the known higher rate of convergence for eigenvalues in the a priori estimates in §3.) The proofs of these theorems require first proving Theorems 4.6 and 4.7, in which additional terms $G_{j,n}$ and $G'_{j,n}$ appear on the right-hand side. These terms, which we subsequently show are genuinely higher order, reflect the non-linearity of the eigenvalue problem, as mentioned above.

In order to prove reliability in Theorem 4.6 and Theorem 4.7, we need two preliminary lemmas:

Lemma 4.4 *Let $(\lambda_{j,n}, u_{j,n})$ be an eigenpair of the discrete problem (2.13) and (λ_j, u_j) be an eigenpair of the continuous problem (2.7). Then denoting by $e_{j,n} := u_j - u_{j,n}$, we have*

$$b(\lambda_j u_j - \lambda_{j,n}u_{j,n}, e_{j,n}) \;=\; \frac{1}{2}(\lambda_j + \lambda_{j,n})\; b(e_{j,n}, e_{j,n}) \;+\; \mathrm{i}(\lambda_{j,n} - \lambda_j)\mathrm{Im}\; b(u_j, u_{j,n}). \tag{4.2}$$

Proof. Using the sesquilinearity of $b(\cdot,\cdot)$ and exploiting the fact that $(\lambda_{j,n}, u_{j,n})$ and (λ_j, u_j) are respectively two normalized eigenpairs of (2.13) and of (2.7), we have

$$\begin{aligned} b(\lambda_j u_j - \lambda_{j,n}u_{j,n}, e_{j,n}) &= b(\lambda_j u_j - \lambda_{j,n}u_{j,n}, u_j) \;-\; b(\lambda_j u_j - \lambda_{j,n}u_{j,n}, u_{j,n}) \\ &= \lambda_j + \lambda_{j,n} \;-\; \lambda_{j,n}\,\overline{b(u_j, u_{j,n})} \;-\; \lambda_j\; b(u_j, u_{j,n}) \\ &= (\lambda_j + \lambda_{j,n})(1 - \mathrm{Re}\; b(u_j, u_{j,n})) \;+\; i(\lambda_{j,n} - \lambda_j)\mathrm{Im}\; b(u_j, u_{j,n}) \ . \end{aligned} \tag{4.3}$$

Another use of sesquilinearity gives us:

$$\begin{aligned} b(e_{j,n}, e_{j,n}) &= \; b(u_j, u_j) \;+\; b(u_{j,n}, u_{j,n}) \;-\; b(u_j, u_{j,n}) \;-\; \overline{b(u_j, u_{j,n})} \\ &= 2 - 2\mathrm{Re}\; b(u_j, u_{j,n}) \ . \end{aligned} \tag{4.4}$$

The insertion of (4.4) into (4.3) concludes the proof.

■

Lemma 4.5 *Let $(\lambda_{j,n}, u_{j,n})$ be an eigenpair of problem (2.13) and let (λ_j, u_j) be an eigenpair of problem (2.7). Then, for any $v \in H^1_\pi(\Omega)$,*

$$a_\kappa(u_j - u_{j,n}, v) \;-\; b(\lambda_j u_j - \lambda_{j,n} u_{j,n}, v) = \sum_{\tau\in\mathcal{T}_n} \int_\tau R_I(u_{j,n}, \lambda_{j,n})\overline{v} - \sum_{f\in\mathcal{F}_n} \int_f R_F(u_{j,n})\overline{v}\ . \tag{4.5}$$

Proof. The result is obtained by integration by parts. We start from the left-most term in (4.5). Using the fact that (λ_j, u_j) is an eigenpair of (2.7) yields

$$\begin{aligned} a_\kappa(u_j - u_{j,n}, v) &= \; a_\kappa(u_j, v) \;-\; a_\kappa(u_{j,n}, v) \;=\; \lambda_j\; b(u_j, v) \;-\; a_\kappa(u_{j,n}, v) \\ &= \lambda_{j,n}\; b(u_{j,n}, v) \;-\; a_\kappa(u_{j,n}, v) \;+\; b(\lambda_j u_j - \lambda_{j,n} u_{j,n}, v)\ . \end{aligned} \tag{4.6}$$

Now apply element-wise integration by parts to $a_\kappa(u_{j,n}, v)$ in (4.6), yielding:

$$\begin{aligned} a_\kappa(u_j - u_{j,n}, v) = &\sum_{\tau\in\mathcal{T}_n} \int_\tau \Big((\nabla + i\kappa)\cdot A(\nabla + i\kappa) u_{j,n} + \lambda_{j,n} B\; u_{j,n} \Big)\overline{v} \\ &- \sum_{f\in\mathcal{F}_n} \int_f [\mathbf{n}_f \cdot A(\nabla + i\kappa) u_{j,n}]_f\; \overline{v} \;+\; b(\lambda_j u_j - \lambda_{j,n} u_{j,n}, v)\ . \end{aligned}$$

■

We now use these lemmas to prove reliability for eigenfunctions. Recall the Scott-Zhang quasi-interpolation operator $I_n : H^1(\Omega) \to V_n$ (defined in [42]), which satisfies, for any $v \in H^1(\Omega)$:

$$\|v - I_n v\|_{0,\tau} \lesssim H_\tau \|v\|_{1,\omega(\tau)}, \qquad \text{and} \qquad \|v - I_n v\|_{0,f} \lesssim H_f^{\frac{1}{2}} \|v\|_{1,\omega(f)}\ , \tag{4.7}$$

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

Theorem 4.6 (Reliability for eigenfunctions) *Let $(\lambda_{j,n}, u_{j,n})$ be a computed eigenpair with $\lambda_{j,n}$ converging to an eigenvalue λ_j of (2.7). Then*

$$\operatorname{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B} \;\lesssim\; \eta_{j,n} \;+\; G_{j,n}, \tag{4.8}$$

where

$$G_{j,n} \;=\; \frac{1}{2}(\lambda_j + \lambda_{j,n} + 2\sigma)\frac{\operatorname{dist}(u_{j,n}, E_1(\lambda_j))^2_{0,B}}{\operatorname{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B}}. \tag{4.9}$$

Proof. Given $u_{j,n}$, define $u_j \in E_1(\lambda_j)$ to simultaneously minimize (3.7) and (3.8) in Lemma 3.3. Again, we define $e_{j,n} := u_j - u_{j,n}$.

Note first that, since (λ_j, u_j) and $(\lambda_{j,n}, u_{j,n})$ respectively solve the eigenvalue problems (2.7) and (2.13), we have, for all $w_n \in V_n$,

$$\begin{aligned} \|e_{j,n}\|^2_{\kappa,A,B} &= a_\kappa(e_{j,n}, e_{j,n} - w_n) \;+\; a_\kappa(u_j, w_n) \;-\; a_\kappa(u_{j,n}, w_n) \;+\; \sigma\; b(e_{j,n}, e_{j,n}) \\ &= a_\kappa(e_{j,n}, e_{j,n} - w_n) \;+\; b(\lambda_j u_j - \lambda_{j,n} u_{j,n}, w_n) \;+\; \sigma\; b(e_{j,n}, e_{j,n}) \\ &= a_\kappa(e_{j,n}, e_{j,n} - w_n) \;-\; b(\lambda_j u_j - \lambda_{j,n} u_{j,n}, e_{j,n} - w_n) \\ &\quad + b(\lambda_j u_j - \lambda_{j,n} u_{j,n}, e_{j,n}) \;+\; \sigma\; b(e_{j,n}, e_{j,n})\ . \end{aligned} \tag{4.10}$$

Looking first at the final two terms in (4.10) we see from Lemma 4.4

$$\begin{aligned} b(\lambda_j u_j - \lambda_{j,n} u_{j,n}, e_{j,n}) \;+\; \sigma\; b(e_{j,n}, e_{j,n}) = \;&\frac{1}{2}(\lambda_j + \lambda_{j,n} + 2\sigma)\; b(e_{j,n}, e_{j,n}) \\ &+ \mathrm{i}(\lambda_{j,n} - \lambda_j)\ \mathrm{Im}\ b(u_j, u_{j,n})\ . \end{aligned} \tag{4.11}$$

Combining this with Lemma 4.5 in (4.10) we get:

$$\begin{aligned} \|e_{j,n}\|^2_{\kappa,A,B} = &\sum_{\tau\in\mathcal{T}_n} \int_\tau R_I(u_{j,n}, \lambda_{j,n})\overline{(e_{j,n} - w_n)} \\ &- \sum_{f\in\mathcal{F}_n} \int_f R_F(u_{j,n})\overline{(e_{j,n} - w_n)} \\ &+ \frac{1}{2}(\lambda_j + \lambda_{j,n} + 2\sigma)\; b(e_{j,n}, e_{j,n}) \;+\; \mathrm{i}(\lambda_{j,n} - \lambda_j)\ \mathrm{Im}\ b(u_j, u_{j,n})\ . \end{aligned} \tag{4.12}$$

Taking the real part of (4.12) and applying the triangle inequality, yields

$$\begin{aligned}\|e_{j,n}\|^2_{\kappa,A,B} \leq & \left| \sum_{\tau\in\mathcal{T}_n} \int_\tau R_I(u_{j,n},\lambda_{j,n})\overline{(e_{j,n}-w_n)} \right| \\ & + \left| \sum_{f\in\mathcal{F}_n} \int_f R_F(u_{j,n})\overline{(e_{j,n}-w_n)} \right| \ + \ \frac{1}{2}(\lambda_j+\lambda_{j,n}+2\sigma) b(e_{j,n},e_{j,n}). \end{aligned} \tag{4.13}$$

In particular we are allowed to choose $w_n = I_n e_{j,n}$ where I_n is the interpolation operator defined above, with properties (4.7). Substituting this in (4.13) and using Cauchy-Schwarz, together with (4.7), we obtain:

$$\begin{aligned}\|e_{j,n}\|^2_{\kappa,A,B} \leq & \sum_{\tau\in\mathcal{T}_n} \|R_I(u_{j,n},\lambda_{j,n})\|_{0,\tau}\|e_{j,n}-I_n e_{j,n}\|_{0,\tau} \\ & + \sum_{f\in\mathcal{F}_n} \|R_F(u_{j,n})\|_{0,f}\|e_{j,n}-I_n e_{j,n}\|_{0,f} \ + \ \frac{1}{2}(\lambda_j+\lambda_{j,n}+2\sigma)\ b(e_{j,n},e_{j,n}) \\ \lesssim & \sum_{\tau\in\mathcal{T}_n} H_\tau \|R_I(u_{j,n},\lambda_{j,n})\|_{0,\tau}\|e_{j,n}\|_{1,\omega(\tau)} \\ & + \sum_{f\in\mathcal{F}_n} H_f^{1/2}\|R_F(u_{j,n})\|_{0,f}\|e_{j,n}\|_{1,\omega(f)} \ + \ \frac{1}{2}(\lambda_j+\lambda_{j,n}+2\sigma)\ b(e_{j,n},e_{j,n}). \end{aligned} \tag{4.14}$$

Since (by an argument analogous to the proof of Lemma 2.1), $\|e_{j,n}\|_{1,\omega(\tau)} \lesssim \|e_{j,n}\|_{\kappa,A,B,\omega(\tau)}$ and $\|e_{j,n}\|_{1,\omega(f)} \lesssim \|e_{j,n}\|_{\kappa,A,B,\omega(f)}$, another application of the Cauchy-Schwarz inequality yields

$$\begin{aligned}\|e_{j,n}\|^2_{\kappa,A,B} \ \lesssim \ & \eta_{j,n}\Bigg\{ \sum_{\tau\in\mathcal{T}_n} \|e_{j,n}\|^2_{\kappa,A,B,\omega(\tau)} \ + \ \sum_{f\in\mathcal{F}_n} \|e_{j,n}\|^2_{\kappa,A,B,\omega(f)} \Bigg\}^{1/2} \\ & + \tfrac{1}{2}(\lambda_j+\lambda_{j,n}+2\sigma)\ b(e_{j,n},e_{j,n}) \\ \lesssim \ & \eta_{j,n}\|e_{j,n}\|_{\kappa,A,B} \ + \ \frac{1}{2}(\lambda_j+\lambda_{j,n}+2\sigma)\ \|e_{j,n}\|^2_{0,B}\ . \end{aligned} \tag{4.15}$$

Finally, in order to conclude the proof we just have to divide both sides of (4.15) by $\|e_{j,n}\|_{\kappa,A,B}$, and recall Lemma 3.3.

■

The next theorem, which is similar to Theorem 4.6, shows the reliability for eigenvalues.

Theorem 4.7 (Reliability for eigenvalues) *Under the same assumptions as in Theorem 4.6, we have:*

$$|\lambda_{j,n}-\lambda_j| \ \lesssim \ \eta^2_{j,n} \ + \ G'_{j,n}\ ,$$

where

$$G'_{j,n} \ = \ \frac{1}{2}\eta_{j,n}(\lambda_j+\lambda_{j,n}+2\sigma)\frac{\mathrm{dist}(u_{j,n},E_1(\lambda_j))^2_{0,B}}{\mathrm{dist}(u_{j,n},E_1(\lambda_j))_{\kappa,A,B}} \ + \ \frac{1}{2}(\lambda_{j,n}-\lambda_j+2\sigma)\mathrm{dist}(u_{j,n},E_1(\lambda_j))^2_{0,B}\ .$$

Proof. With $u_j, u_{j,n}$ and $e_{j,n}$ as in the proof of Theorem 4.6, we use (3.6) to obtain

$$|\lambda_{j,n}-\lambda_j| = a_\kappa(e_{j,n},e_{j,n}) \ - \ \lambda_j\ b(e_{j,n},e_{j,n})\ . \tag{4.16}$$

Hence noticing that $a_\kappa(e_{j,n},e_{j,n}) \ \leq \ a_\kappa(e_{j,n},e_{j,n})+\sigma b(e_{j,n},e_{j,n}) \ = \ \|e_{j,n}\|^2_{\kappa,A,B} \ = \ \mathrm{dist}(u_{j,n},E_1(\lambda_j))_{\kappa,A,B}$, and substituting (4.8) into (4.16) we obtain

$$\begin{aligned}|\lambda_{j,n}-\lambda_j| \ \leq \ & (\eta_{j,n}+G_{j,n})\mathrm{dist}(u_{j,n},E_1(\lambda_j))_{\kappa,A,B} - \lambda_j \mathrm{dist}(u_{j,n},E_1(\lambda_j))^2_{0,B} \\ = \ & \eta_{j,n}\ \mathrm{dist}(u_{j,n},E_1(\lambda_j))_{\kappa,A,B} \ + \ \frac{1}{2}(\lambda_{j,n}+\lambda_j+2\sigma)\ \mathrm{dist}(u_{j,n},E_1(\lambda_j))^2_{0,B} \\ & - \ \lambda_j\ \mathrm{dist}(u_{j,n},E_1(\lambda_j))^2_{0,B} \\ = \ & \eta_{j,n}\ \mathrm{dist}(u_{j,n},E_1(\lambda_j))_{\kappa,A,B} \ + \ \frac{1}{2}(\lambda_{j,n}-\lambda_j+2\sigma)\ \mathrm{dist}(u_{j,n},E_1(\lambda_j))^2_{0,B} \end{aligned}$$

Then using (4.8) again we have

$$|\lambda_{j,n} - \lambda_j| \lesssim \eta_{j,n}^2 + \frac{1}{2}\eta_{j,n}(\lambda_{j,n} + \lambda_j + 2\sigma)\frac{\mathrm{dist}(u_{j,n}, E_1(\lambda_j))_{0,B}^2}{\mathrm{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B}}$$
$$+\frac{1}{2}(\lambda_{j,n} - \lambda_j + 2\sigma)\ \mathrm{dist}(u_{j,n}, E_1(\lambda_j))_{0,B}^2.$$

■

Now the two final results of this section show that $G_{j,n}$ in Theorem 4.6 and $G'_{j,n}$ in Theorem 4.7 are indeed "higher order terms".

Theorem 4.8 *Under the same assumptions of Theorem 4.6 we have that if $H_n^{\max}$ is small enough, then*

$$\mathrm{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B} \lesssim \eta_{j,n}\ . \tag{4.17}$$

Proof. Again write $e_{j,n} := u_j - u_{j,n}$, where $u_j \in E_1(\lambda_j)$ is the simultaneous minimizer of (3.7), (3.8). From Theorem 4.6 we have

$$\mathrm{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B} \lesssim \eta_{j,n} + G_{j,n}\ . \tag{4.18}$$

Now, applying Theorem 3.5(ii) we have

$$G_{j,n} = \frac{1}{2}(\lambda_j + \lambda_{j,n} + 2\sigma)\ \frac{\mathrm{dist}(u_{j,n}, E_1(\lambda_j))_{0,B}^2}{\mathrm{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B}}$$
$$\lesssim \frac{1}{2}(\lambda_j + \lambda_{j,n} + 2\sigma)(H_n^{max})^{2s}\ \mathrm{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B}\ . \tag{4.19}$$

Supposing that $H_n^{\max}$ is small enough, we obtain $\lambda_{j,n} \lesssim \lambda_j$ and

$$G_{j,n} \lesssim (\lambda_j + \sigma)\ (H_n^{max})^{2s}\mathrm{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B} < \frac{1}{2}\mathrm{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B}\ .$$

Then from (4.18), we have $\mathrm{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B} \lesssim \eta_{j,n}$, as required. ■

Theorem 4.9 *Under the same assumptions as Theorem 4.8 we have:*

$$|\lambda_{j,n} - \lambda_j| \lesssim \eta_{j,n}^2\ .$$

Proof. Again write $e_{j,n} := u_j - u_{j,n}$, where $u_j \in E_1(\lambda_j)$ is the simultaneous minimizer of (3.7), (3.8). Then we have, from (4.16),

$$|\lambda_{j,n} - \lambda_j| = a_\kappa(e_{j,n}, e_{j,n}) - \lambda_j\ b(e_{j,n}, e_{j,n}) \leq a_\kappa(e_{j,n}, e_{j,n})\ . \tag{4.20}$$

Noticing that $a_\kappa(e_{j,n}, e_{j,n}) \leq \mathrm{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B}^2$ and substituting (4.17) in (4.20) we obtain the result. ■

5 Efficiency

While the reliability estimates in the previous section show the error is bounded above by a positive constant times an error estimator as the mesh is refined, the "global efficiency" estimate, which we obtain in this section (Corollary 5.6), obtains a corresponding lower bound. In order to prove Corollary 5.6, we need first a weaker result called "local efficiency", which is obtained in Lemma 5.4.

We shall use bubble functions, which are smooth and positive real valued functions with support on an element and are bounded by 1 in the L^∞ norm. They are constructed using polynomials and so satisfy inverse estimates which are collected in the next proposition. 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 boundary of τ and for any edge f, and we need a real-valued bubble function ψ_f with support in Δ_f and which vanishes on the boundary of Δ_f. In [8, p.587] - see also [45, Lemma 3.3] - such bubble functions ψ_τ, ψ_f are constructed which satisfy the following properties:

Proposition 5.1 *There are constants, which only depend on the regularity of the mesh $\mathcal{T}_n$, such that*

$$\|v\|_{0,\tau} \lesssim \|\psi_\tau^{1/2} v\|_{0,\tau}, \tag{5.1}$$

$$|\psi_\tau v|_{1,\tau} \lesssim H_\tau^{-1} \|v\|_{0,\tau}, \tag{5.2}$$

$$\|w\|_{0,f} \lesssim \|\psi_f^{1/2}\ w\|_{0,f}, \tag{5.3}$$

$$|\psi_f\ w|_{1,\Delta_f} \lesssim H_f^{-1/2} \|w\|_{0,f}, \tag{5.4}$$

$$\|\psi_f\ w\|_{0,\Delta_f} \lesssim H_f^{1/2} \|w\|_{0,f}, \tag{5.5}$$

hold for all $\tau \in \mathcal{T}_n$, all $f \in \mathcal{F}_n$, and for all polynomials v and w.

In the next two lemmas we bound the L^2 norms of the residuals R_I and R_F on τ (defined in Definition 4.3 above) in terms of the energy norm of the error on τ.

Lemma 5.2 *Let $(\lambda_{j,n}, u_{j,n})$ be an eigenpair of (2.13) and (λ_j, u_j) be an eigenpair of (2.7). Then for any element $\tau \in \mathcal{T}_n$ we have*

$$H_\tau \|R_I(u_{j,n}, \lambda_{j,n})\|_{0,\tau} \lesssim \|A^{1/2}(\nabla + i\kappa)(u_j - u_{j,n})\|_{0,\tau} + H_\tau \|\lambda_{j,n} u_{j,n} - \lambda_j u_j\|_{0,B,\tau}\ . \tag{5.6}$$

Proof. Let ψ_τ be the bubble function introduced above and set $w_\tau = \psi_\tau\ R_I(u_{j,n}, \lambda_{j,n})$. Because we are using linear elements, and since A, B are assumed to be constant in the interior of each element, the residual R_I is a linear function on τ. This fact together with (5.1) and the positivity of ψ_τ leads to

$$\begin{aligned}
\|R_I(u_{j,n}, \lambda_{j,n})\|_{0,\tau}^2 \lesssim \|\psi_\tau^{1/2} R_I(u_{j,n}, \lambda_{j,n})\|_{0,\tau}^2 &= \int_\tau \psi_\tau |R_I(u_{j,n}, \lambda_{j,n})|^2 \\
&= \int_\tau R_I(u_{j,n}, \lambda_{j,n}) \overline{w}_\tau \\
&= \int_\tau \left((\nabla + i\kappa) \cdot A(\nabla + i\kappa) u_{j,n} \ +\ \lambda_{j,n}\ B\ u_{j,n} \right) \overline{w}_\tau\ .
\end{aligned} \tag{5.7}$$

Hence integrating by parts and using the fact that w_τ vanishes on $\partial\tau$, we get

$$\|R_I(u_{j,n}, \lambda_{j,n})\|_{0,\tau}^2 \lesssim -a_\kappa(u_{j,n}, w_\tau) \ +\ \lambda_{j,n} b(u_{j,n}, w_\tau).$$

Since u_j satisfies (2.7) and since $\omega_\tau \in H_0^1(\tau) \subset H_\pi^1(\Omega)$, we have

$$\|R_I(u_{j,n}, \lambda_{j,n})\|_{0,\tau}^2 \lesssim -a_\kappa(u_{j,n} - u_j, w_\tau) \ +\ b(\lambda_{j,n} u_{j,n} - \lambda_j u_j, w_\tau)\ .$$

Hence by the Cauchy-Schwarz inequality we obtain

$$\begin{aligned}
\|R_I(u_{j,n}, \lambda_{j,n})\|_{0,\tau}^2 &\lesssim \left\|A^{1/2}(\nabla + i\kappa)(u_j - u_{j,n})\right\|_{0,\tau} \left\|A^{1/2}(\nabla - i\kappa)\overline{w}_\tau\right\|_{0,\tau} \\
&\quad +\ \|\lambda_{j,n} u_{j,n} - \lambda_j u_j\|_{0,B,\tau}\ \|w_\tau\|_{0,B,\tau} \\
&\lesssim \left\|A^{1/2}(\nabla + i\kappa)(u_j - u_{j,n})\right\|_{0,\tau}\ \|w_\tau\|_{1,\tau} \\
&\quad +\ \|\lambda_{j,n} u_{j,n} - \lambda_j u_j\|_{0,B,\tau}\ \|w_\tau\|_{0,B,\tau}.
\end{aligned} \tag{5.8}$$

For the final step we use the definition of w_τ and (5.2) to obtain from (5.8):

$$\begin{aligned}
\|R_I(u_{j,n}, \lambda_{j,n})\|_{0,\tau}^2 \lesssim \Big[& H_\tau^{-1}\ \left\|A^{1/2}(\nabla + i\kappa)(u_j - u_{j,n})\right\|_{0,\tau} \\
& + \|\lambda_{j,n} u_{j,n} - \lambda_j u_j\|_{0,B,\tau} \Big] \|R_I(u_{j,n}, \lambda_{j,n})\|_{0,\tau},
\end{aligned}$$

then multiplying each side by $H_\tau \|R_I(u_{j,n}, \lambda_{j,n})\|_{0,\tau}^{-1}$ yields the result. ∎

Lemma 5.3 *Under the same conditions as Lemma 5.2, for any f in $\mathcal{F}_n$*

$$H_f^{1/2}\|R_F(u_{j,n})\|_{0,f} \lesssim \sum_{\tau\in\Delta_f}\left(\|A^{1/2}(\nabla+i\kappa)(u_j-u_{j,n})\|_{0,\tau}+H_f\ \|\lambda_{j,n}u_{j,n}-\lambda_j u_j\|_{0,B,\tau}\right). \tag{5.9}$$

Proof. Let ψ_f be as in Proposition 5.1, and set $w_f := \psi_f\, R_F(u_{j,n})$. Applying (5.3), recalling that w_f vanishes on all edges except f and then using Lemma 4.5, we obtain

$$\begin{aligned}\|R_F(u_{j,n})\|_{0,f}^2 \lesssim \|\psi_f^{1/2}R_F(u_{j,n})\|_{0,f}^2 &= \int_f R_F(u_{j,n})\overline{w}_f = \sum_{f\in\mathcal{F}_n}\int_f R_F(u_{j,n})\overline{w}_f\\ &= \sum_{\tau\in\Delta_f}\int_\tau R_I(u_{j,n},\lambda_{j,n})\overline{w}_f - a_\kappa(u_j-u_{j,n},w_f)+b(\lambda_j u_j-\lambda_{j,n}u_{j,n},w_f).\end{aligned} \tag{5.10}$$

Then, using the Cauchy-Schwarz inequality on (5.10), we get:

$$\begin{aligned}\|R_F(u_{j,n})\|_{0,f}^2 \lesssim \sum_{\tau\in\Delta_f}&\|R_I(u_{j,n},\lambda_{j,n})\|_{0,\tau}\ \|w_f\|_{0,\tau}\\ &+\|A^{1/2}(\nabla+i\kappa)(u_j-u_{j,n})\|_{0,\Delta_f}\ \|A^{1/2}(\nabla-i\kappa)\overline{w_f}\|_{0,\Delta_f}\\ &+\|\lambda_{j,n}u_{j,n}-\lambda_j u_j\|_{0,B,\Delta_f}\ \|w_f\|_{0,B,\Delta_f}.\end{aligned} \tag{5.11}$$

Now, we have to estimate each of the three terms on the right-hand side of (5.11). The first term can be treated using (5.5) and (5.6):

$$\begin{aligned}&\sum_{\tau\in\Delta_f}\|R_I(u_{j,n},\lambda_{j,n})\|_{0,\tau}\ \|w_f\|_{0,\tau} \lesssim H_f^{1/2}\sum_{\tau\in\Delta_f}\|R_I(u_{j,n},\lambda_{j,n})\|_{0,\tau}\ \|R_F(u_{j,n})\|_{0,f}\\ &\lesssim H_f^{1/2}\ \|R_F(u_{j,n})\|_{0,f}\sum_{\tau\in\Delta_f}\left(H_\tau^{-1}\|A^{1/2}(\nabla+i\kappa)(u_j-u_{j,n})\|_{0,\tau}+\|\lambda_{j,n}u_{j,n}-\lambda_j u_j\|_{0,B,\tau}\right).\end{aligned} \tag{5.12}$$

To treat the second term on the right hand side of (5.11), note that we can use (5.4) and (5.5) to obtain:

$$\begin{aligned}\|A^{1/2}(\nabla-i\kappa)\overline{w_f}\|_{0,\Delta_f} &\lesssim \|w_f\|_{0,\Delta_f}+|w_f|_{1,\Delta_f}\\ &\lesssim \left(H_f^{1/2}+H_f^{-1/2}\right)\|R_F(u_{j,n})\|_{0,f}\end{aligned} \tag{5.13}$$

To treat the last term on the right hand side of (5.11), note that by (5.5),

$$\|w_f\|_{0,B,\Delta_f} \lesssim \|w_f\|_{0,\Delta_f} \lesssim H_f^{1/2}\ \|R_F(u_{j,n})\|_{0,f}. \tag{5.14}$$

Now substituting (5.12), (5.13) and (5.14) in (5.11) we get:

$$\begin{aligned}\|R_F(u_{j,n})\|_{0,f}^2 \lesssim \|R_F(u_{j,n})\|_{0,f}\Big[&(H_f^{1/2}+H_f^{-1/2})\textstyle\sum_{\tau\in\Delta_f}\|A^{1/2}(\nabla+i\kappa)(u_j-u_{j,n})\|_{0,\tau}\\ &+H_f^{1/2}\ \|\lambda_{j,n}u_{j,n}-\lambda_j u_j\|_{0,B,\tau}\Big].\end{aligned}$$

To conclude the proof we have to multiply both sides by $H_f^{1/2}\|R_F(u_{j,n})\|_{0,f}^{-1}$ and note that $H_f H_\tau^{-1}\lesssim 1$. ∎

In Lemma 5.4 we prove a local version of the efficiency, this result is extended to the whole domain Ω in Theorem 5.5.

Lemma 5.4 (Local efficiency) *Under the same conditions as Lemma 5.2, define*

$$\eta_{j,n,f}^2 := \sum_{\tau\in\Delta_f} H_\tau^2\ \|R_I(u_{j,n},\lambda_{j,n})\|_{0,\tau}^2 + H_f\ \|R_F(u_{j,n})\|_{0,f}^2.$$

Then

$$\eta_{j,n,f}^2 \lesssim \sum_{\tau\in\Delta_f}\left(\|A^{1/2}(\nabla+i\kappa)(u_j-u_{j,n})\|_{0,\tau}^2+H_\tau^2\|\lambda_{j,n}u_{j,n}-\lambda_j u_j\|_{0,B,\tau}^2\right). \tag{5.15}$$

Proof. Combine the results from Lemma 5.2 and Lemma 5.3. ∎

Theorem 5.5 (Global efficiency) *Under the same assumptions as Lemma 5.2, suppose also that $u_j \in E_1(\lambda_j)$ minimizes the distance in Lemma 3.3. Then*

$$\eta_{j,n}^2 \lesssim \ \mathrm{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B}^2 \ + \ \|H_\tau(\lambda_{j,n}u_{j,n} - \lambda_j u_j)\|_{0,B}^2 \ . \tag{5.16}$$

Proof. Summing (5.15) over all edges f and recalling (4.1) yields

$$\eta_{j,n}^2 \lesssim \sum_{f\in\mathcal{F}_n} \eta_{j,n,f}^2 \ \lesssim \ \sum_{f\in\mathcal{F}_n} \left\{ \sum_{\tau\in\Delta_f} \left(\|A^{1/2}(\nabla + i\kappa)(u_j - u_{j,n})\|_{0,\tau}^2 \ + \ H_\tau^2 \|\lambda_{j,n}u_{j,n} - \lambda_j u_j\|_{0,B,\tau}^2 \right) \right\}. \tag{5.17}$$

The subsets Δ_f, for each value of f, are not all disjoint, but the maximum number of overlapping subdomains Δ_f at any point in the interior of an element is 3. So (5.17) yields the result. ∎

The following corollary explains why Theorem 5.5 really is a statement about global efficiency.

Corollary 5.6 *Under the same assumptions as Theorem 5.5 and with the extra assumption that $H_n^{\max}$ is small enough, we have*

$$\eta_{j,n} \lesssim \ \mathrm{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B} \ .$$

Proof. By Theorem 5.5 (recalling that $\|u_{j,n}\|_{0,B} = 1$), and then Theorem 3.5, we obtain

$$\begin{aligned} \eta_{j,n}^2 \ &\lesssim \ \|u_j - u_{j,n}\|_{\kappa,A,B}^2 \ + \ \left(H_n^{\max}\right)^2 \left(|\lambda_{j,n} - \lambda_j|^2 \ + \ \lambda_j^2 \ \|u_{j,n} - u_j\|_{0,B}^2\right) \\ &\lesssim \ \|u_j - u_{j,n}\|_{\kappa,A,B}^2 \ + \ \left(H_n^{\max}\right)^2 \left(\|u_j - u_{j,n}\|_{\kappa,A,B}^4 \ + \ \lambda_j^2 \ \left(H_n^{\max}\right)^{2s} \|u_j - u_{j,n}\|_{\kappa,A,B}^2\right) \\ &\lesssim \ \left(1 + \left(H_n^{\max}\right)^{2+2s}\right) \ \|u_j - u_{j,n}\|_{\kappa,A,B}^2 \ \lesssim \ \mathrm{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B}^2 \ , \end{aligned}$$

and the result follows. ∎

The next corollary is very important for computations, since it proves that $\eta_{j,n} \to 0$ is equivalent to convergence of the computed eigenpair in an appropriate sense.

Corollary 5.7 *Let $(\lambda_{j,n}, u_{j,n})$ be a computed eigenpair and assume also that $H_n^{\max}$ is small enough.*
(i) If $\eta_{j,n} \to 0$ as $n \to \infty$, then both $\mathrm{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B}$ and $|\lambda_{j,n} - \lambda_j|$ tend to zero;
(ii) If $\mathrm{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B} \to 0$ as $n \to \infty$, then both $\lambda_{j,n} \to \lambda_j$ and $\eta_{j,n} \to 0$ as $n \to \infty$.

Proof. Part (i) follows directly from Theorems 4.8 and 4.9. To obtain (ii), notice that if $\mathrm{dist}(u_{j,n}, E_1(\lambda_j))_{\kappa,A,B} \to 0$, then by Theorem 3.5 we have $\lambda_{j,n} \to \lambda_j$ and by Corollary 5.6, we also have $\eta_{j,n} \to 0$ as $n \to \infty$.

∎

6 Adaptive FEM and numerical experiments

In this section we present an adaptive algorithm and study numerically its performance for various problems related to the TE case mode of problem (1.1). In this case A is piecewise constant, $B = 1$ and there are typically localized singularities in the gradient of the eigenfunctions at corner points of the interface in the dielectric ε, leading to a strong need for adaptivity. We shall use the a posteriori error estimator $\eta_{j,n}$ introduced in §4 (which we shall refer to as the "standard" estimator), and we shall compare the results to those using a slightly different estimator, below referred to as the "modified" estimator, and defined by

$$\tilde{\eta}_{j,n} := \left\{ \sum_{\tau\in\mathcal{T}_n} H_\tau^2 \alpha_\tau^{-1} \|R_I(u_{j,n}, \lambda_{j,n})\|_{0,\tau}^2 \ + \ \sum_{f\in\mathcal{F}_n} H_f \alpha_f^{-1} \|R_F(u_{j,n})\|_{0,f}^2 \right\}^{1/2}, \tag{6.1}$$

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. Since $\eta_{j,n}$ and $\tilde{\eta}_{j,n}$ are equal up to multiplication by a constant (independent of the mesh), all the results in §§4 and 5 also hold for $\tilde{\eta}_{j,n}$. We shall see below that in some cases $\tilde{\eta}_{j,n}$ performs much better than $\eta_{j,n}$. An error estimator similar to $\tilde{\eta}_{j,n}$ for elliptic PDEs with discontinuous coefficients is presented in [8], where also its *robustness* with respect to the jumps in A is proved. In this work we observe that with fixed A, and for some values of quasimomentum κ, the modified estimator performs better than the standard estimator. However for other values of κ the two estimators perform similarly. This observation merits further investigation, but to avoid making the paper longer we do not discuss it further here.

Our adaptivity algorithm uses the following standard marking strategy.

Definition 6.1 (Marking Strategy) *Given a parameter $0 < \theta < 1$, the procedure is: mark the elements in a minimal subset $\hat{\mathcal{M}}_n$ of $\mathcal{T}_n$ such that*

$$\left(\sum_{\tau \in \hat{\mathcal{M}}_n} \eta_{j,n,\tau}^2 \right)^{1/2} \geq \theta \, \eta_{j,n} \, , \tag{6.2}$$

where $\eta_{j,n,\tau}$ is:

$$\eta_{j,n,\tau}^2 := H_\tau^2 \| R_I(u_{j,n}, \lambda_{j,n}) \|_{0,\tau}^2 \; + \; \sum_{f \in \partial\tau} \frac{1}{2} H_f \| \; R_F(u_{j,n}) \|_{0,f}^2 \, . \tag{6.3}$$

It is straightforward to see that $\left(\sum_{\tau \in \mathcal{T}_n} \eta_{j,n,\tau}^2 \right)^{1/2} = \eta_{j,n}$. Also when the "modified" error estimator $\tilde{\eta}_{j,n}$ is used an analogous marking strategy is employed.

Our adaptive algorithm is given in Algorithm 1 and requires specification of the two parameters; tol (the accuracy tolerance) and $\max_n$ (the maximum number of allowed mesh refinements). For the refinement step in the algorithm we have used standard "red refinement" (see, e.g., [13]). Eigenpairs are computed via Arnoldi's method using ARPACK [31] with the associated linear systems implemented by the sparse direct solver ME27 from the HSL archive [41,25].

Algorithm 1 Adaptivity algorithm

Require: $\mathcal{T}_0$, j, κ
  $n = 0$
  repeat
    Compute $(\lambda_{j,n}, u_{j,n})$ on $\mathcal{T}_n$
    Compute $\eta_{j,n,\tau}$ for all $\tau \in \mathcal{T}_n$
    Mark the elements using the marking strategy (Definition 6.1)
    Refine the mesh $\mathcal{T}_n$ and construct $\mathcal{T}_{n+1}$
    $n = n + 1$
  until $\eta_{j,n} \leq$ tol OR $n \geq \max_n$

6.1 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 dielectric properties of real photonic crystals are of this order. The jump in the value of A could produce a jump in the gradient of the eigenfunctions across the boundaries of the subdomains. As above, the eigenfunctions lie in $H^{s+1}(\Omega)$, with $s > 1/2 - \varepsilon$, for all $\varepsilon > 0$ in general. However, since we resolve exactly the interface, we see a convergence speed coming from the regularity of the eigenfunctions in each subdomain, which is $u \in H^{s+1}(\Omega_i)$ where $s > 2/3$. From Theorem 3.5(i,iii) we have that using uniform refinement, the rate of convergence for eigenvalues should be at least $\mathcal{O}(H_n^{\max})^{2s}$.

Tables 1 and 2 illustrate the performance of the standard and modified error estimators for computing the smallest non-zero eigenvalue of (1.1) in the case of quasimomentum $\kappa = (0, 0)$. Here n is the refinement number as in Algorithm 1 and $\beta = -\log(|\lambda_j - \lambda_{j,n}|/|\lambda_j - \lambda_{j,n-1}|)/\log(\#\text{DOFs}_n/\#\text{DOFs}_{n-1})$ is a computed estimate of the convergence rate. Tables 3 and 4 give the analogous results for quasimomentum $\kappa = (\pi, \pi)$. We can see that in both cases the adaptive methods perform better than the uniform refinements, however the "modified" error estimator performs even better than the "standard" one, in fact for both values of κ less DOFs are necessary for the "modified" error estimator compared to the "standard" one to reach the same accuracy. In fact this observation holds for any κ which is far enough from the origin. and this is the main reason behind the introduction of the error estimator $\tilde{\eta}_{j,n}$. For this problem the exact eigenvalues λ are unknown, so in all four tables the errors which are displayed are computed using very accurate approximations of the exact eigenvalues, computed on a very fine mesh involving about a million of DOFs.

Theorem 4.9 shows that for sufficiently fine meshes (apart form a hidden constant), $\eta_{j,n}^2$ provides an upper bound for the eigenvalue error. This is also true for $\tilde{\eta}_{j,n}$ by the remarks above. To numerically investigate the implications of this result, we approximate numerically the hidden constant $C_r = |\lambda_j - \lambda_{j,n}|/\eta_{j,n}^2$ in Theorem 4.9. Similarly, we compute $\tilde{C}_r = |\lambda_j - \lambda_{j,n}|/\tilde{\eta}_{j,n}^2$. As can be seen in Tables 5 and 6, the computed

Uniform			$\eta_{j,n}$							
			$\theta = 0.5$				$\theta = 0.8$			
n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	β	n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	β
1	0.0584	400	1	0.0584	400	-	1	0.0584	400	-
2	0.0188	1600	6	0.0155	1584	0.9623	3	0.0187	1460	0.8798
3	0.0063	6400	9	0.0064	3764	1.0277	5	0.0048	5670	1.0025
4	0.0021	25600	13	0.0018	12626	1.0541	6	0.0021	10711	1.3050
5	0.0007	102400	16	0.0006	29583	1.1846	8	0.0005	40698	1.0864

Table 1 Comparison for $\kappa = (0, 0)$ and with $j = 2$ between the uniform refinement and the adaptive method with the "standard" error estimator.

Uniform			$\tilde{\eta}_{j,n}$							
			$\theta = 0.5$				$\theta = 0.8$			
n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	β	n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	β
1	0.0584	400	1	0.0584	400	-	1	0.0584	400	-
2	0.0188	1600	5	0.0139	1356	1.1746	3	0.0138	1452	1.1165
3	0.0063	6400	8	0.0058	3437	0.9360	5	0.0032	5824	1.0478
4	0.0021	25600	12	0.0017	11101	1.0522	6	0.0018	11342	0.8904
5	0.0007	102400	15	0.0006	26334	1.1829	7	0.0007	23044	1.2318

Table 2 Comparison for $\kappa = (0, 0)$ and with $j = 2$ between the uniform refinement and the adaptive method with the "modified" error estimator.

Uniform			$\eta_{j,n}$							
			$\theta = 0.5$				$\theta = 0.8$			
n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	β	n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	β
1	0.0505	400	1	0.0505	400	-	1	0.0505	400	-
2	0.0155	1600	6	0.0158	1686	0.8086	4	0.0089	2922	0.8718
3	0.0050	6400	11	0.0040	7622	0.9073	5	0.0053	6264	0.6742
4	0.0016	25600	15	0.0016	22344	0.8396	7	0.0015	24110	0.9299
5	0.0005	102400	19	0.0005	55426	1.3181	9	0.0004	86668	1.0845

Table 3 Comparison for $\kappa = (\pi, \pi)$ and with $j = 2$ between the uniform refinement and the adaptive method with the "standard" error estimator.

Uniform			$\tilde{\eta}_{j,n}$							
			$\theta = 0.5$				$\theta = 0.8$			
n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	β	n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	β
1	0.0505	400	1	0.0505	400	-	1	0.0505	400	-
2	0.0155	1600	5	0.0122	1398	1.1314	3	0.0118	1546	1.0727
3	0.0050	6400	9	0.0036	4984	0.9626	5	0.0028	6348	1.0228
4	0.0016	25600	12	0.0016	12505	0.8736	6	0.0015	14749	0.7578
5	0.0005	102400	17	0.0005	32822	1.2407	8	0.0003	57480	1.1161

Table 4 Comparison for $\kappa = (\pi, \pi)$ and with $j = 2$ between the uniform refinement and the adaptive method with the "modified" error estimator.

values of C_r and $\tilde{C}_r$ remain almost constant as the mesh is refined and also they do not seem to be affected by variations in the value of κ. This implies that both the error estimators $\eta_{j,n}$ and $\tilde{\eta}_{j,n}$ decay in the same way as the true error, which is important in practice since it means that $\eta_{j,n}$ and $\tilde{\eta}_{j,n}$ can be used as an indicator of the size of the true error, even when the true error is not available. However, it is easy to see that the value of $\tilde{C}_r$ doesn't change as much as the value of C_r, this suggests that the "modified" error estimator follows better the behavior of the true error. Also the "modified" error estimator performs better than the "standard" one because for the same n, the true error $|\lambda_j - \lambda_{j,n}|$ is smaller using the "modified" error estimator. In Figure 1 we depict the mesh coming from the fourth iteration of Algorithm 1 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 2 we depict the eigenfunction corresponding to the smallest positive eigenvalue of the problem with quasimomentum $(0, 0)$.

n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	$\eta^2_{j,n}$	C_r	$\lvert\lambda_j - \lambda_{j,n}\rvert$	$\tilde{\eta}^2_{j,n}$	$\tilde{C}_r$
1	0.0584	0.1126	0.5182	0.0584	1.2280	0.0475
2	0.0543	0.0974	0.5571	0.0425	0.9520	0.0447
3	0.0414	0.0751	0.5513	0.0330	0.6746	0.0489
4	0.0314	0.0538	0.5830	0.0231	0.4848	0.0477
5	0.0232	0.0371	0.6242	0.0139	0.3172	0.0439
6	0.0155	0.0253	0.6135	0.0105	0.2378	0.0440
7	0.0103	0.0191	0.5398	0.0080	0.1752	0.0457
8	0.0083	0.0142	0.5807	0.0058	0.1266	0.0460
9	0.0064	0.0103	0.6168	0.0039	0.0900	0.0437
10	0.0049	0.0074	0.6618	0.0027	0.0671	0.0402
11	0.0028	0.0053	0.5342	0.0022	0.0511	0.0425
12	0.0022	0.0040	0.5504	0.0017	0.0386	0.0439
13	0.0018	0.0030	0.5877	0.0013	0.0290	0.0434
14	0.0014	0.0023	0.6122	0.0009	0.0215	0.0396

Table 5 Comparison for $\kappa = (0, 0)$ and with $j = 2$ between the "standard" error estimator and the "modified" error estimator with $\theta = 0.5$.

n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	$\eta^2_{j,n}$	C_r	$\lvert\lambda_j - \lambda_{j,n}\rvert$	$\tilde{\eta}^2_{j,n}$	$\tilde{C}_r$
1	0.0505	0.1629	0.3098	0.0505	1.2271	0.0411
2	0.0473	0.1337	0.3538	0.0363	0.9866	0.0368
3	0.0391	0.1020	0.3832	0.0276	0.7095	0.0389
4	0.0319	0.0750	0.4257	0.0176	0.4690	0.0375
5	0.0244	0.0548	0.4462	0.0122	0.3453	0.0355
6	0.0158	0.0395	0.3988	0.0091	0.2696	0.0336
7	0.0090	0.0285	0.3172	0.0071	0.1997	0.0355
8	0.0082	0.0225	0.3641	0.0054	0.1466	0.0365
9	0.0071	0.0175	0.4079	0.0036	0.1060	0.0340
10	0.0057	0.0135	0.4248	0.0026	0.0809	0.0322
11	0.0040	0.0103	0.3901	0.0020	0.0627	0.0318
12	0.0025	0.0079	0.3175	0.0016	0.0480	0.0336
13	0.0022	0.0063	0.3406	0.0012	0.0366	0.0338
14	0.0019	0.0051	0.3818	0.0009	0.0279	0.0310

Table 6 Comparison for $\kappa = (\pi, \pi)$ and with $j = 2$ between the "standard" error estimator and the "modified" error estimator with $\theta = 0.5$.

6.2 TE mode problem on supercell

The spectra of photonic crystals typically contain band gaps, but, for many applications, the identification of 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 [18,20] 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 — see [20] and [19, Theorem 2]. Such localized defects do not change the bands of the essential spectrum [19, Theorem 1].

In the next set of experiments we continue to work with the TE case problem and we shall use the "supercell method" [43] 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 §6.1 using a supercell with two or more layers of periodic structure surrounding the defect. (In Figure 3 we depict the unit cell with two layers added). This new medium (since it is again infinitely periodic) has a new band in its spectrum caused by the defect. However it is also known ([43]) that as the number of periodic layers increases, and under some conditions, the band shrinks exponentially quickly to the eigenvalue of the original defective material.

In order to compute good approximations of these trapped modes, it is not only necessary to compute accurately the TE case problem on supercells, but also it is necessary to use enough layers of periodic structure around the defect to ensure that the band in the supercell problem is sufficiently narrow. Ideally, the error in the approximation of the eigenvalue problem and the diameter of the defect band should have the same order.

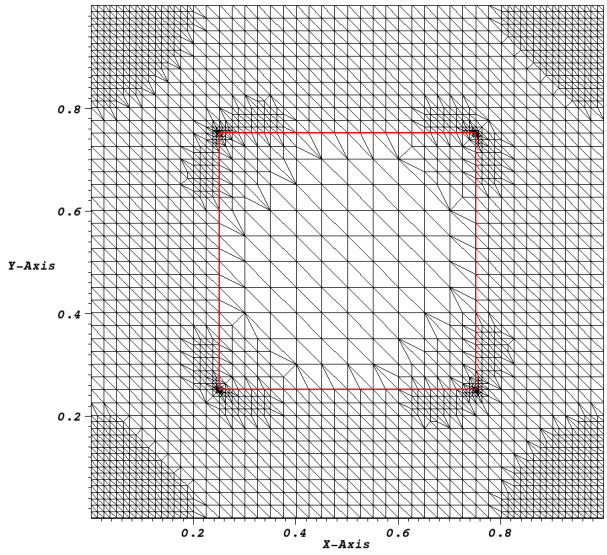

Fig. 1 A refined mesh coming from the adaptive FEM for the TE mode problem with $\kappa = (0, 0)$ and using $\eta_{j,n}$, with $j = 2$.

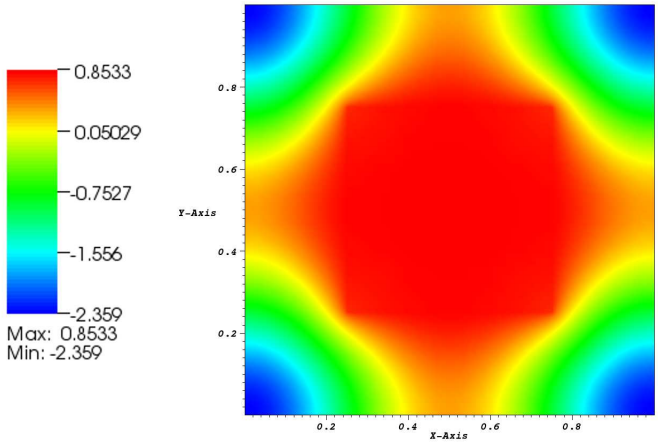

Fig. 2 The eigenfunction with index $j = 2$ of the TE mode problem with quasimomentum $\kappa = (0, 0)$.

Just to give an idea of the size of the defect band as a function of the number of layers of periodic structure around the defect, Table 7, gives the diameters of the defect bands for different sizes of the supercell computed using the "exact" values of the trapped eigenvalues computed on a very fine mesh at 55 different points of the first Brillouin zone.

In Tables 8-11 and Figures 4-5 the performance of the two error estimators are compared with uniform refinement for computing a trapped mode for different values of the quasimomentum on a supercell with 2 layers of periodic medium, whose first Brillouin zone is $[-\pi/5, \pi/5]^2$. As can be seen in the case of supercells and trapped modes we have that both the "standard" and the "modified" error estimators give greater orders of convergence compared to uniform refinement.

For this problem the difference in the accuracy between our method and the uniform refinement method is much more striking compared to the previous example. The reason is not only that the adaptive method refines around the corners, where the singularities are, but also, because the most part of the "energy" of

Number of Layers	Diameter defect band
2	0.3008
3	0.0295
4	0.0154

Table 7 Size of the defect band as function of the number of layers of periodic structure around the defect.

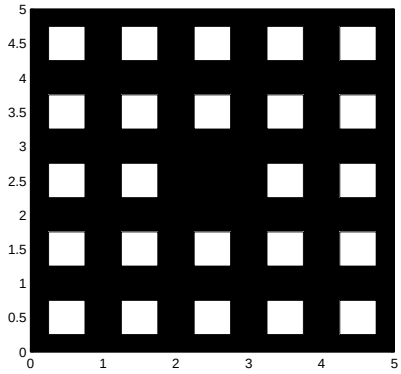

Fig. 3 The structure of the supercell used for the computations.

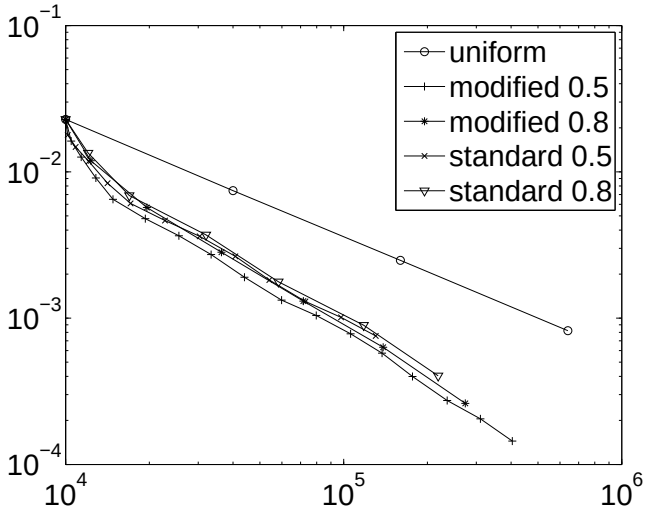

Fig. 4 Loglog plot of convergence of adaptive and uniform refinements for the TE problem on a supercell with quasimomentum $\kappa = (0, 0)$ and with $j = 28$.

the solution is inside the defect, which is a very small region. Moreover, the "modified" error estimator still performs a bit better than the standard one with no extra computational costs involved. Also in this case we computed the "exact" values of the eigenvalues λ_j using more than one million of DOFs.

In Figure 6 we depict the mesh coming from the fourth iteration of Algorithm 1 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 not to show important singularities, is because the trapped mode has a fast decay outside the defect and so the singularities at the corners of the inclusions are less important away from the defect. In Figure 7, we depict the eigenfunction corresponding to the mode "trapped" inside the defect. This eigenfunction is the one used to refine the mesh in Figure 6.

As explained above, it is important to use enough layers of periodic medium around the defect to have a narrow defect band. In Tables 12-14 we denote with λ^* the eigenvalue trapped in the defect and with λ_n^* the approximation of the trapped eigenvalue. We decided to change the notation because increasing the number

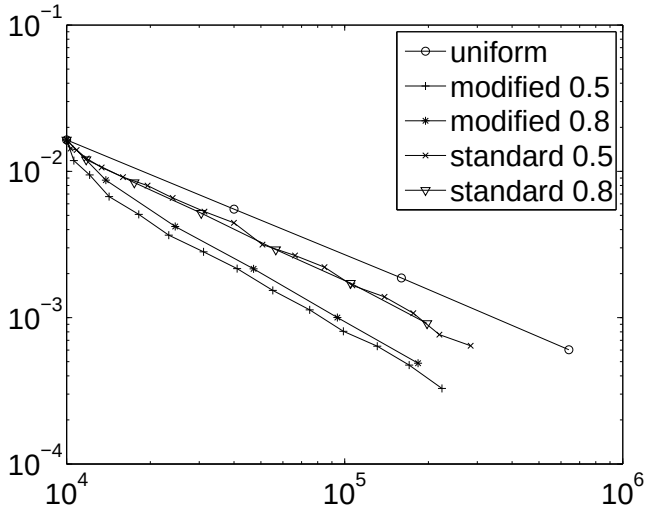

Fig. 5 Loglog plot of convergence of adaptive and uniform refinements for the TE problem on a supercell with quasimomentum $\kappa = (\pi/5, \pi/5)$ and with $j = 28$.

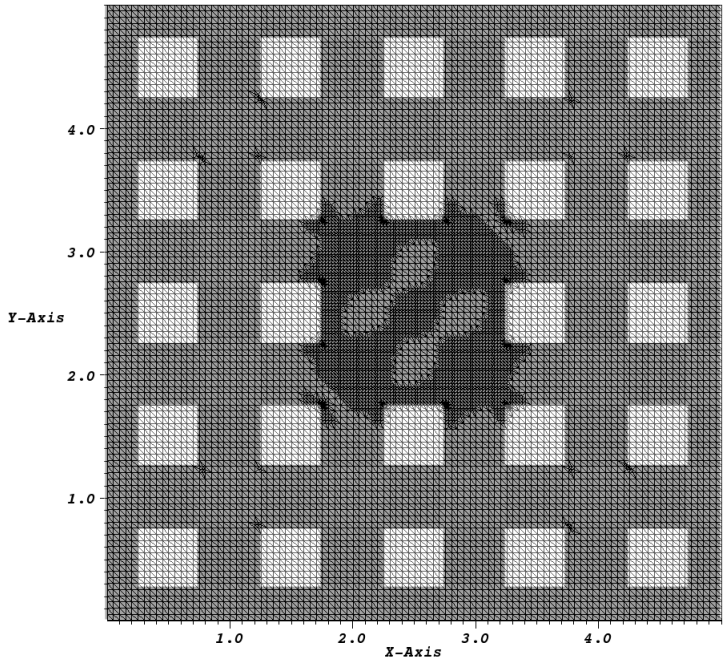

Fig. 6 An adapted mesh for a trapped eigenvalue for the TE case on a supercell with quasimomentum $\kappa = (0, 0)$ and with $j = 28$. The structure of the supercell is superimposed on the mesh

of periodic layers in the cell the index j of the trapped mode changes. In Tables 12 and 13 it is possible to see how the uniform and the adaptive methods behave when increasing the size of the supercell. In particular the superiority of the adaptive method is clearly visible. Finally in Table 14 we show the DOFs needed by the uniform and the adaptive methods to reach an accuracy higher than the order of the diameter of the defect band for different sizes of the supercell.

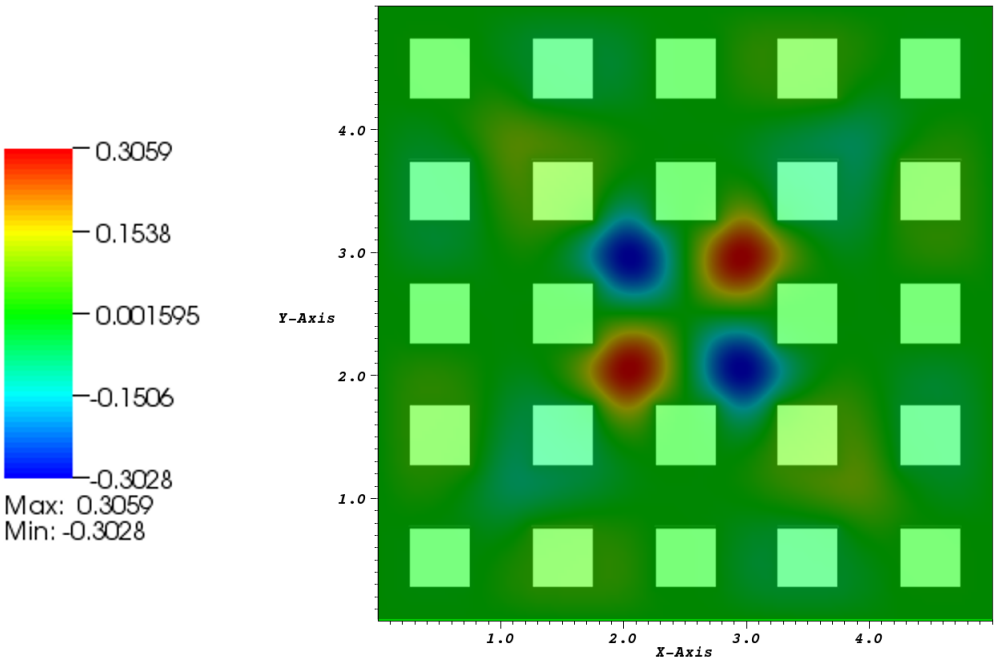

Fig. 7 A picture of the eigenfunction trapped in the defect for the TE case on a supercell with quasimomentum $\kappa = (0, 0)$ and with $j = 28$. The structure of the supercell is superimposed on the picture of the eigenfunction

Uniform			$\eta_{j,n}$							
			$\theta = 0.5$				$\theta = 0.8$			
n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	β	n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	β
1	0.0228	10000	1	0.0228	10000	-	1	0.0228	10000	-
2	0.0074	40000	6	0.0061	17128	2.4583	3	0.0069	16958	2.2677
3	0.0025	160000	9	0.0026	40791	0.9589	5	0.0018	58290	1.1002
4	0.0008	640000	13	0.0008	130455	1.0775	6	0.0009	118082	0.9687

Table 8 Comparison for $\kappa = (0, 0)$ and with $j = 28$ between the uniform refinement and the adaptive method with the "standard" error estimator on a supercell.

Uniform			$\tilde{\eta}_{j,n}$							
			$\theta = 0.5$				$\theta = 0.8$			
n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	β	n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	β
1	0.0228	10000	1	0.0228	10000	-	1	0.0228	10000	-
2	0.0074	40000	5	0.0065	14808	3.2038	3	0.0057	19598	2.0628
3	0.0025	160000	8	0.0027	33366	1.0704	4	0.0028	36356	1.1363
4	0.0008	640000	12	0.0008	105876	1.0794	6	0.0006	138720	1.1169

Table 9 Comparison for $\kappa = (0, 0)$ and with $j = 28$ between the uniform refinement and the adaptive method with the "modified" error estimator on a supercell.

References

1. M. Ainsworth and J.T. Oden, *A Posterior Error Estimation in Finite Element Analysis*, Wiley, 2000.
2. H. Ammari and F. Santosa, Guided waves in a photonic bandgap strucutre with a line defect, *SIAM J. Appl. Math.* 64(6):2018-2033, 2004.
3. N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, Holt, Rinehart and Winston, Philadelphia, 1976.
4. W. Axmann and P. Kuchment, An efficient finite element method for computing spectra of photonic and acoustic band-gap materials, *J. Comput. Physics* 150:468-481, 1999.
5. I. Babuška and J. Osborn. *Estimates for the errors in eigenvalue and eigenvector approximation by Galerkin methods, with particular attention to the case of multiple eigenvalues*, *SIAM J. Numer. Anal.*, 24:1249-1276, 1987.

Uniform			$\eta_{j,n}$							
			$\theta = 0.5$				$\theta = 0.8$			
n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	β	n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	β
1	0.0164	10000	1	0.0164	10000	-	1	0.0164	10000	-
2	0.0055	40000	9	0.0053	31329	0.9873	4	0.0052	30489	1.0114
3	0.0019	160000	14	0.0017	106654	0.9492	6	0.0017	105023	0.9401
4	0.0006	640000	18	0.0006	283900	0.9692	7	0.0009	197817	1.4987

Table 10 Comparison for $\kappa = (\pi/5, \pi/5)$ and with $j = 28$ between the uniform refinement and the adaptive method with the "standard" error estimator on a supercell.

Uniform			$\tilde{\eta}_{j,n}$							
			$\theta = 0.5$				$\theta = 0.8$			
n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	β	n	$\lvert\lambda_j - \lambda_{j,n}\rvert$	#DOFs	β
1	0.0164	10000	1	0.0164	10000	-	1	0.0164	10000	-
2	0.0055	40000	5	0.0051	18184	1.9584	3	0.0042	24580	1.3021
3	0.0019	160000	9	0.0015	55164	1.0793	4	0.0022	47044	1.8452
4	0.0006	640000	12	0.0006	131051	1.0132	4	0.0005	183654	0.6437

Table 11 Comparison for $\kappa = (\pi/5, \pi/5)$ and with $j = 28$ between the uniform refinement and the adaptive method with the "modified" error estimator on a supercell.

3 Layers				4 Layers			
Uniform		$\tilde{\eta}_{j,n}$, $\theta = 0.5$		Uniform		$\tilde{\eta}_{j,n}$, $\theta = 0.5$	
$\lvert\lambda^* - \lambda_n^*\rvert$	#DOFs	$\lvert\lambda^* - \lambda_n^*\rvert$	#DOFs	$\lvert\lambda^* - \lambda_n^*\rvert$	#DOFs	$\lvert\lambda^* - \lambda_n^*\rvert$	#DOFs
0.0324	12544	0.0324	12544	0.0356	20736	0.0356	20736
0.0092	50176	0.0080	19874	0.0100	82944	0.0101	47824
0.0013	200704	0.0008	60041	0.0015	331776	0.0010	156979

Table 12 Comparison for $\kappa = (0, 0)$ between the uniform refinement and the adaptive method with the "modified" error estimator on a supercells of different sizes.

3 Layers				4 Layers			
Uniform		$\tilde{\eta}_{j,n}$, $\theta = 0.5$		Uniform		$\tilde{\eta}_{j,n}$, $\theta = 0.5$	
$\lvert\lambda^* - \lambda_n^*\rvert$	#DOFs	$\lvert\lambda^* - \lambda_n^*\rvert$	#DOFs	$\lvert\lambda^* - \lambda_n^*\rvert$	#DOFs	$\lvert\lambda^* - \lambda_n^*\rvert$	#DOFs
0.0705	12544	0.0705	12544	0.0184	20736	0.0184	20736
0.0441	50176	0.0467	17675	0.0581	82944	0.0521	37406
0.0353	200704	0.0347	27281	0.0334	331776	0.0332	55463

Table 13 Comparison for $\kappa = (\pi/7, \pi/7)$ for the 3 layers case and for $\kappa = (\pi/9, \pi/9)$ for the 4 layers case between the uniform refinement and the adaptive method with the "modified" error estimator on a supercells of different sizes.

6. I. Babuška and J. Osborn. *Finite element-Galerkin approximation of the eigenvalues and eigenvectors of selfadjoint problems*, *Math. Comput.* 186:275-297, 1989.
7. I. Babuška and J. Osborn. *Eigenvalue Problems*, in Handbook of Numerical Analysis Vol II, eds P.G. Cairlet and J.L. Lions, North Holland, 641-787, 1991.
8. C. Bernardi and R.Verfürth, Adaptive finite element methods for elliptic equations with non-smooth coefficients, *Numer. Math.* 85:579-608, 2000.
9. M.S. Birman and T.A. Suslina, Periodic magnetic Hamiltonian with a variable metric. The problem of absolute continuity, St Petersburg Mth. J. 11: 203-232, 2000.
10. D. Boffi, M. Conforti and L. Gastaldi, Modified edge finite elements for photonic crystals, *Numer. Math.* 105:249-266, 2006
11. Y. Cao, Z. Hou and Y. Liu, Convergence problem of plane-wave expansion method for phononic crystals, *Physics Letters A* 327:247-253, 2004.
12. C. Carstensen and J. Gedicke, An oscillation-free adaptive FEM for symmetric eigenvalue problems Preprint 489, MATHEON, DFG Research Center on Mathematics for key technologies in Berlin, 2008.
13. C. Carstensen and J. Hu, Hanging nodes in the unifying theory of a posteriori finite element error control *J. Comput. Math.*, 27: 215-236, 2009.
14. D. C. Dobson, An Efficient Method for Band Structure Calculations in 2D Photonic Crystals, *J. Comp. Phys.* 149:363-376, 1999.
15. D. C. Dobson, J. Gopalakrishnan and J. E. Pasciak, An efficient method for band structure calculations in 3D photonic crystals, *J. Comput. Phys.* 161(2):668-679, 2000.
16. C. Engström and M. Wang, Complex dispersion relation calculations with the symmetric interior penalty method Int. J. Num. Meth. Engng, 84:849863, 2010.
17. A. Figotin and V. Goren, Resolvent method for computations of localized defect modes of H-polarization in two-dimensional photonic crystals, *Phys. Rev. E* 64:1-16, 2001.

		Uniform		$\tilde{\eta}_{j,n}$, $\theta = 0.5$	
N. Layers	Diameter	$\lvert\lambda^* - \lambda_n^*\rvert$	#DOFs	$\lvert\lambda^* - \lambda_n^*\rvert$	#DOFs
2	0.3008	0.0025	160000	0.0027	33366
3	0.0295	0.0092	50176	0.0080	19874
4	0.0154	0.0015	331776	0.0010	156979

Table 14 Comparison for $\kappa = (0, 0)$ between the uniform refinement and the adaptive method with the "modified" error estimator on a supercells of different sizes.

18. A. Figotin and V. Gorentsveig, Localized electromagnetic waves in a layered periodic dielectric medium with a defect, *Phys. Rev. B* 58(1):180-188, 1998.
19. A. Figotin and A. Klein, Localized classical waves created by defects, *J. Stat. Phys.* 86:165-177, 1997.
20. A. Figotin and A. Klein, Midgap defect modes in dielectric and acoustic media, *SIAM J. Appl. Math.* 58(6):1748-1773, 1998.
21. S. Giani. *Convergence of Adaptive Finite Element Methods for Elliptic Eigenvalue Problems with Application to Photonic Crystals*, PhD Thesis, University of Bath, 2008.
22. S. Giani and Ivan G. Graham, A Convergent Adaptive Method for Elliptic Eigenvalue Problems, *SIAM J. Numer. Anal.* 47(2):1067-1091, 2009.
23. 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.
24. B. Hiett. *Photonic Crystal modelling using finite element analysis*, PhD Thesis, University of Southampton, 2000.
25. HSL archive, http://hsl.rl.ac.uk/archive/hslarchive.html
26. J. D. Joannopoulos and S. G. Johnson, Block-iterative frequency-domain methods for Maxwell's equations in a planewave basis, *Optics Express* 8:173-190, 2001.
27. J. D. Joannopoulos, R. D. Meade, and J. N. Winn, Photonic Crystals. Molding the Flow of Light Princeton Univ. Press, Princeton, NJ, 1995.
28. A. Klöckner, *On the computation of maximally localized Wannier functions*, PhD Thesis, Karlsruhe University, 2004.
29. P. Kuchment, *Floquet Theory for Partial Differential Equations*, Birkhauser Verlag, 1993
30. P. Kuchment, *The mathematics of photonic crystals*, SIAM, Frontiers Appl. Math. 22:207-272, 2001.
31. 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
32. M. Luo, Q. H. Liu, and Z. Li, Spectral element method for band structures of two-dimensional anisotropic photonic crystals, Phys. Rev. E 79: 026705, 2009.
33. V. Mehrmann and A. Miedlar Adaptive solution of PDE-eigenvalue problems. Part I: eigenvalues, Preprint 565 MATHEON, DFG Research Center on Mathematics for key technologies in Berlin, 2009.
34. R. Norton and R. Scheichl, Convergence Analysis of Planewave Expansion Methods for Schroedinger Operators with Discontinuous Periodic Potentials, *SIAM Journal on Numerical Analysis* 47(6):4356-4380 (2010).
35. R. Norton and R. Scheichl, Analysis of planewave methods for photonic crystal fibres, BICS Preprint number 10/10, University of Bath, 2010.
36. R.A. Norton, Numerical Computation of Band Gaps in Photonic Crystal Fibres, Ph.D. thesis, University of Bath, 2008.
37. 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(19):195108, 2005
38. K. Sakoda, *Optical Properties of Photonic Crystals*, Springer-Verlag, 2001.
39. K. Schmidt and P. Kauf, Computation of band structure of two-dimensional photonic crystals with *hp* finite elements, Comput. Meth. Appl. Mech. Eng. 198: 1249 - 1259, 2009.
40. K. Schmidt and R. Kappeler, Efficient computation of photonic crystal waveguide modes with dispersive material, Optics Express,18: 7307-7322, 2010.
41. J. A. Scott, Sparse Direct Methods: An Introduction. *Lecture Notes in Physics*, 535, 401, 2000
42. R. L. Scott and S. Zhang, Finite element interpolation of nonsmooth functions satisfying boundary conditions, *Math Comp* 54:483-493, 1990.
43. S. Soussi, Convergence of the supercell method for defect modes calculations in photonic crystals, *SIAM J. Numer. Anal.* 43(3):1175-1201, 2005.
44. G. Strang and G. J. Fix. *An Analysis of the Finite Element Method.* Prentice-Hall, 1973.
45. R. Verfürth, *A Review of a Posteriori Error Estimation and Adaptive Mesh Refinement Techniques*, Wiley-Teubner ,1996.