Efficient and reliable hp-FEM estimates for quadratic eigenvalue problems and photonic crystal applications

Christian Engström[a], Stefano Giani[b], Luka Grubišić[c]

[a] *Department of Mathematics and Mathematical Statistics, Umeå University, Umeå, Sweden*
[b] *Durham University, School of Engineering and Computing Sciences, South Road, Durham DH1 3LE, United Kingdom.*
[c] *University of Zagreb, Department of Mathematics, Bijenička 30, 10000 Zagreb, Croatia*

Abstract

We present a-posteriori analysis of higher order finite element approximations (hp-FEM) for quadratic Fredholm-valued operator functions. Residual estimates for approximations of the algebraic eigenspaces are derived and we reduce the analysis of the estimator to the analysis of an associated boundary value problem. For the reasons of robustness we also consider approximations of the associated invariant pairs. We show that our estimator inherits the efficiency and reliability properties of the underlying boundary value estimator. As a model problem we consider spectral problems arising in analysis of photonic crystals. In particular, we present an example where a targeted family of eigenvalues cannot be guaranteed to be semisimple. Numerical experiments with hp-FEM show the predicted convergence rates. The measured effectivities of the estimator compare favorably with the performance of the same estimator on the associated boundary value problem. We also present a benchmark estimator, based on the dual weighted residual (DWR) approach, which is more expensive to compute but whose measured effectivities are close to one.

Keywords: Nonlinear eigenvalue problems, numerical methods, invariant pairs
2000 MSC: Primary: 47J10, Secondary: 65F15, 65N30

1. Introduction

A large number of processes in science are described by operator functions with a nonlinear dependence of a spectral parameter. In particular, dispersion and damping are commonly present in nature. Dispersion implies that the operator function is nonlinear and damping implies that the operator function is non-selfadjoint. For an overview of applications leading to nonlinear eigenvalue problems from a computational linear algebra point of view we refer to [10, 50]. Approximations of operator functions were also extensively studied; and the general operator theory provides basic a-priori convergence results for Fredholm-valued functions [35, 36, 53].

However, only few papers consider a-posteriori error estimations for Galerkin approximations of operator functions [2, 47]; see also [8] for a general approach to a-posteriori error estimation and [20] for an overview of techniques for linear eigenvalue problems. We point out that neither of the references considers higher order finite element approximations.

Email addresses: christian.engstrom@math.umu.se (Christian Engström), stefano.giani@durham.ac.uk (Stefano Giani), luka.grubisic@math.hr (Luka Grubišić)

This paper is concerned with a-posteriori analysis and quadratic eigenvalue problems which can be analyzed on an abstract Hilbert space by the Fredholm analytic theorem [44, Theorem 1.3.1]; see also [28]. In particular, for given sesquilinear forms $\mathfrak{a}_i[\cdot,\cdot]$, $i = 0, 1, 2$, we seek a vector $u \neq 0$ and a scalar λ such that

$$\mathfrak{a}(\lambda)[u, v] := \mathfrak{a}_0[u, v] + \lambda\mathfrak{a}_1[u, v] + \lambda^2\mathfrak{a}_2[u, v] = 0, \tag{1}$$

for all v. To this variational formulation we construct a quadratic operator valued function Q on an appropriate Hilbert space $\mathcal{H}$.

An operator T on $\mathcal{H}$ is here called a Fredholm operator if it is a bounded operator such that the dimensions of its null space $\mathrm{Ker}(T)$ and of the orthogonal complement of its range $\mathrm{Ran}(T)^\perp$ are finite. In the case when the dimensions agree $\dim \mathrm{Ker}(T) = \dim \mathrm{Ran}(T)^\perp$ we say that T is the Fredholm operator of index zero. In our setting, a quadratic polynomial Q is Fredholm-valued if there exist compact operators A_1 and A_2 and a Fredholm operator A_0 such that

$$Q(z) = A_0 + zA_1 + z^2A_2$$

for $z \in \mathbb{C}$. For the analytic Fredholm theorem to apply we have to additionally ascertain that there exists a complex number z_0 such that $Q(z_0)$ has a bounded inverse. In this case we may conclude [44, Theorem 1.3.1] that Q^{-1} is a finitely meromorphic function. This in turn implies that the spectrum consists of eigenvalues and the point spectrum $\sigma(Q) = \{\lambda \in \mathbb{C} \ : \ \dim(\mathrm{Ker}\, Q(\lambda)) > 0\}$ is countable. Moreover, for each $\lambda \in \sigma(Q)$ the dimension $\dim(\mathrm{Ker}\, Q(\lambda))$, which is called the geometric multiplicity of λ, is finite and the associated Jordan chains of generalized eigenvectors have finite length. The length of a chain of generalized eigenvectors is bounded by the algebraic multiplicity [44].

With the help of the phase space representation of the spectral problem for Q, we will define the notion of the algebraic space associated to λ and will argue that it makes sense to compute approximations of this space when numerically analyzing the spectral problem associated to $T(\cdot)$. Furthermore, we show an example where the space $\mathrm{Ker}(T(\lambda))$ may be a true subspace of the associated algebraic subspace. In this case we prove that by reducing our subspace error indicators we approximate a subspace of this algebraic space, which is asymptotically tending to the desired geometric eigenspace.

In the case when we can not prove semisimplicity of the targeted eigenspace a priori, the choice of approximating the nearest subspace of an algebraic space by reducing the residual subspace indicator is a reasonable choice. We propose it as an alternative to proving convergence to some averaged measure of the target eigenspace (eg. the arithmetic mean of targeted eigenvalues). We tackle the general case of approximating the whole algebraic eigensubspace by considering simple invariant pairs for Q, see [11, 12].

An invariant pair is a generalization of the notion of the eigenvector and eigenvalue. This analogy is based on the fact that invariant pairs provide for a coordinate representation (basis) of the action of the quadratic polynomial on its invariant subspace. In particular, this allows us to incorporate the effect of numerical linear algebra in the construction of approximate solutions through the use of local condition numbers in the overall approximation estimate. We argue that the size of standard local condition numbers (condition number for the returned basis of an eigensubspace) indicates whether there are generalized eigenvectors in an algebraic eigenspace, or the assumption that the targeted eigenvalues are semisimple is robust under perturbations.

Based on the this we propose two possible strategies. First, we can base our refinement on the reduction of the residuals (approximation defects) just for the constructed approximate eigenvectors. The constructed approximate eigenspace will be close to a subspace of the larger algebraic eigensubspace. Asymptotically this "nearest" subspace of the algebraic eigenspace converges to the subspace generated by eigenvectors, see [46]. Second, we optimize the local condition numbers by constructing a different basis of the algebraic eigenspace, eg. the Schur basis. Such subspace representation is much better conditioned and standard small scale algorithms can be used to reconstruct the desired spectral information in a postprocessing step.

We point out that in our photonic crystal applications we did not observe extremely ill-conditioned eigenvector basis (under the assumption of semisimplicity) and so we pursued the first strategy. We note that we were not able to exclude the possibility of the occurrence of associated generalized eigenvectors by an analytic argument. In this context we feel that issues of the computation of Schur vectors and related practical algorithm is better left for a subsequent paper where we will consider an appropriate application where generalized eigenvectors occur naturally. We point out that to our knowledge the only nontrivial application where generalized eigenvectors occur in numerical analysis literature is in reference [30, Section 7.2]. There the authors consider an application in hydrodynamics and aim to compute the Jordan basis of an algebraic subspace.

We now give a brief overview of the paper. In Section 2 we present residual estimates, based on resolvent calculus, for Galerkin approximations of general quadratic Fredholm-valued operator functions. In Section 3 we discuss one possible usage of quadratic operator functions in photonic crystal applications. Subsequently, in Section 4 we introduce two computable residual estimators for higher order finite element approximations of quadratic eigenvalue problems and in Section 5 we present numerical experiments which corroborate our theory.

2. The quadratic eigenvalue problem

In this section we provide details of the construction of both the operator representation of a variationally posed quadratic eigenvalue problem, as well as of the construction of its phase space linearization.

Let $\mathcal{V}$ be a complex Hilbert space with inner product $(\cdot,\cdot)_{\mathcal{V}}$ and let $||\cdot||_{\mathcal{V}}$ denote the norm on $\mathcal{V}$. Let $\mathfrak{a}_n : \mathcal{V}\times\mathcal{V}\to\mathbb{C}$, $n=1,2,3$ denote sesquilinear forms bounded on $\mathcal{V}\times\mathcal{V}$.

The variationally posed eigenvalue problem is to find u in $\mathcal{V}\backslash\{0\}$ and complex numbers λ such that

$$\mathfrak{a}(\lambda)[u,v] := \mathfrak{a}_0[u,v] + \lambda\mathfrak{a}_1[u,v] + \lambda^2\mathfrak{a}_2[u,v] = 0, \tag{2}$$

for all $v\in\mathcal{V}$. Riesz representation theorem implies that there exist bounded linear operators $A_n : \mathcal{V}\to\mathcal{V}$ such that

$$(A_n u, v)_{\mathcal{V}} := \mathfrak{a}_n[u,v], \quad n=0,1,2. \tag{3}$$

Hence, the quadratic eigenvalue problem (2) can be stated as follows: Find $u\in\mathcal{V}\backslash\{0\}$ and $\lambda\in\mathbb{C}$ such that

$$Q(\lambda)u = 0, \quad Q(\lambda) := A_0 + \lambda A_1 + \lambda^2 A_2. \tag{4}$$

Assume that A_1 and A_2 are compact and that A_0 is a Fredholm operator of index zero. The operator $Q(\lambda)$ is then for all $\lambda\in\mathbb{C}$ a Fredholm operator of index zero, see [44].

Let $\mathbf{M}$ denote the the linear pencil $\mathbf{M}(\lambda) := \mathbf{M}_0 - \lambda\mathbf{M}_1$, where

$$\mathbf{M}_0 = \begin{bmatrix} A_0 & A_1 \\ 0 & I \end{bmatrix}, \quad \mathbf{M}_1 = \begin{bmatrix} 0 & -A_2 \\ I & 0 \end{bmatrix}. \tag{5}$$

The eigenpairs of $\mathbf{M}$ are then eigenvectors $\mathbf{u} \in \mathcal{V} \oplus \mathcal{V}$ and complex numbers λ such that

$$\mathbf{M}(\lambda)\mathbf{u} = 0. \tag{6}$$

Since the two outer factors in the factorization

$$\mathbf{M}(\lambda) = \begin{bmatrix} I & A_1 + \lambda A_2 \\ 0 & I \end{bmatrix} \begin{bmatrix} Q(\lambda) & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} I & -A_2 \\ \lambda I & I \end{bmatrix} \tag{7}$$

are bounded and boundedly invertible, it follows that the Q and $\mathbf{M}$ have the same eigenvalues and it can be shown that these eigenvalues have the same multiplicities [42, Lemma 12.5].

When A_0 is invertible the eigenvalues of $\mathbf{L} = \mathbf{M}_0^{-1}\mathbf{M}_1$ are solutions of

$$\mathbf{L}\mathbf{u} = \mu\mathbf{u}, \quad \mathbf{L} = \begin{bmatrix} -A_0^{-1}A_1 & -A_0^{-1}A_2 \\ I & 0 \end{bmatrix}. \tag{8}$$

Consequently, the spectrum of the quadratic operator function Q coincide with the set of numbers

$$\sigma(Q) = \{\lambda \in \mathbb{C} \,:\, \lambda = 1/\mu,\, \mu \in \sigma(\mathbf{L})\}. \tag{9}$$

In this article, Fredholm-valued operator polynomials (4) with invertible leading coefficient A_0 are studied. Hence, from the analytic Fredholm theorem follows that $\sigma(\mathbf{L})$ consist only of isolated eigenvalues of finite multiplicity [44, Theorem 1.3.1]. For $\mu \in \sigma(\mathbf{L})\backslash\{0\}$ there exists a smallest positive integer α called the *ascent* of $\mu - \mathbf{L}$ such that Ker $(\mu - \mathbf{L})^\alpha$ = Ker $(\mu - \mathbf{L})^{\alpha+1}$. The elements of the finite dimensional subspace Ker $(\mu - \mathbf{L})^\alpha$ are called *generalized eigenvectors*. The *order* of a generalized eigenvector is the smallest positive integer n such that $u \in$ Ker $(\mu - \mathbf{L})^n$. The dimension of Ker $(\mu - \mathbf{L})^\alpha$ is called the algebraic multiplicity of μ and the dimension of Ker $(\mu - \mathbf{L})$ is called the geometric multiplicity. The eigenvalues of $\mathbf{M}$ and Q coincide and have the same multiplicity [42, Lemma 12.2].

2.1. Galerkin discretization

The approximation of Fredholm operator functions has been studied intensively [35, 36] and the general theory provide basic convergence results. However, we will study the block operator matrix formulation (6). This reformulation of the problem provides us with additional tools from the general spectral approximation theory of linear non-selfadjoint operators. Our approach to analyze the block operator matrix is partly similar to Kolata [38] and more recently [21].

Let $\mathcal{V}^\nu \subset \mathcal{V}$ denote a sequence of conforming finite element spaces with the approximation property

$$\lim_{\nu \to 0} \inf_{u^\nu \in \mathcal{V}^\nu} ||u - u^\nu||_{\mathcal{V}} = 0, \quad \text{for all}\, u \in \mathcal{V}. \tag{10}$$

The parameter ν can be thought of as the mesh size h for h-FEM and one divided by the number of degrees of freedom N for p-FEM.

Let $P^\nu : \mathcal{V} \to \mathcal{V}^\nu$ denote the projection of $\mathcal{V}$ into $\mathcal{V}^\nu$ defined by the inner product $(P^\nu u, v^\nu)_{\mathcal{V}} = (u, v^\nu)_{\mathcal{V}}$ for all $v^\nu \in \mathcal{V}^\nu$. For the operator polynomial

$$Q(\lambda) = A_0 + \lambda A_1 + \lambda^2 A_2, \tag{11}$$

define the projected operator function $Q^\nu : \mathcal{V}^\nu \to \mathcal{V}^\nu$ by $Q^\nu = P^\nu Q$. The Galerkin eigenvalue problem is to find vectors $u^\nu \in \mathcal{V}^\nu \setminus \{0\}$ and values $\lambda^\nu \in \mathbb{C}$ such that

$$Q^\nu(\lambda^\nu)u^\nu = 0. \tag{12}$$

The corresponding Galerkin eigenvalue problem for the pencil is to find vectors $\mathbf{u}^\nu \in \mathcal{V}^\nu \oplus \mathcal{V}^\nu \setminus \{0\}$ and values $\lambda^\nu \in \mathbb{C}$ such that

$$\mathbf{M}^\nu \mathbf{u}^\nu := (\mathbf{M}_0^\nu - \lambda \mathbf{M}_1^\nu)\mathbf{u}^\nu = 0, \tag{13}$$

where

$$\mathbf{M}_0^\nu = \begin{bmatrix} P^\nu A_0 & P^\nu A_1 \\ 0 & P^\nu \end{bmatrix}, \quad \mathbf{M}_1^\nu = \begin{bmatrix} 0 & -P^\nu A_2 \\ P^\nu & 0 \end{bmatrix}. \tag{14}$$

The pencil $\mathbf{M}^\nu$ does not converge to $\mathbf{M}$ in norm and we will therefore use Kolata's idea [38] to construct a pencil on $\mathcal{V} \oplus \mathcal{V}$, which has the same eigenvalues and generalized eigenvectors as (14) on $\mathcal{V}^\nu \oplus \mathcal{V}^\nu$ but it possible to use standard perturbation theory to compare this pencil with $\mathbf{M}_0 - \lambda \mathbf{M}_1$. Note that the operator A_0 can be written on the form $A_0 = I + \mathcal{K}$, where $\mathcal{K}$ is compact and define the operators

$$\widetilde{\mathbf{M}}_0^\nu = \begin{bmatrix} I + P^\nu \mathcal{K} & P^\nu A_1 \\ 0 & I \end{bmatrix}, \quad \widetilde{\mathbf{M}}_1^\nu = \begin{bmatrix} 0 & -P^\nu A_2 \\ I & 0 \end{bmatrix}. \tag{15}$$

In [21] it has been shown that $\widetilde{\mathbf{M}}_0 - \lambda \widetilde{\mathbf{M}}_1$ has the same spectrum as (14) and (12). Furthermore, Theorem 2.1 holds, which states that we have the norm convergence property of this auxiliary pencil. For further discussion and references see [38, 21].

Theorem 2.1. *[21] Assume that $A_0(\omega)$ for a given $\omega \in \mathbb{C}$ is boundedly invertible. Then the block operator matrix $\widetilde{\mathbf{L}}^\nu = (\widetilde{\mathbf{M}}_0^\nu)^{-1}\widetilde{\mathbf{M}}_1^\nu$ converge to $\mathbf{L} = \mathbf{M}_0^{-1}\mathbf{M}_1$ in norm.*

Assume that μ has algebraic multiplicity n. Given a circle $\mathcal{C}_\mu \in \rho(\mathbf{L})$ which encloses $\mu \in \sigma(\mathbf{L})$ and no other elements of $\sigma(\mathbf{L})$ the spectral projections $\mathbf{E}(\mu)$ and $\mathbf{E}^\nu(\mu)$ are defined by

$$\begin{aligned} \mathbf{E}(\mu) &= \frac{1}{2\pi \mathrm{i}} \int_{\mathcal{C}_\mu} (z - \mathbf{L})^{-1} dz, \\ \mathbf{E}^\nu(\mu) &= \frac{1}{2\pi \mathrm{i}} \int_{\mathcal{C}_\mu} (z - \mathbf{L}^\nu)^{-1} dz. \end{aligned} \tag{16}$$

The range of the operator $\mathbf{E}(\mu) : \mathcal{V} \oplus \mathcal{V} \to \mathcal{V} \oplus \mathcal{V}$, denoted by

$$\mathfrak{M}_\mu := \mathrm{Ran}(\mathbf{E}(\mu)) = \mathrm{Ker}(\mu - \mathbf{L})^\alpha, \tag{17}$$

is the corresponding generalized eigenspace. Moreover, let $\mathfrak{M}_\mu^\nu$ denote the range of the projection $\mathbf{E}^\nu(\mu)$. Note that this notation signifies the dependence of $\mathbf{E}^\nu(\mu)$ on $\mathcal{C}_\mu$ and Theorem 2.1 implies $\|\mathbf{E}(\mu) - \mathbf{E}^\nu(\mu)\| \to 0$.

2.2. Subspace containment gap

To study the properties of the convergence $\mathfrak{M}_\mu^\nu \to \mathfrak{M}_\mu$, where we allow the dimensions of $\mathfrak{M}_\mu$ and $\mathfrak{M}$ to differ, we use the notion of the subspace containment gap in addition to studying the norm difference of projections like $\|\mathbf{E}(\mu) - \mathbf{E}^\nu(\mu)\| \to 0$. Given two closed subspaces $\mathfrak{B}_1$ and $\mathfrak{B}_2$ of a Hilbert space $\mathcal{V}$ the proximity of the spaces are measured in terms of the containment gap

$$\delta(\mathfrak{B}_1, \mathfrak{B}_2) = \sup_{v_1 \in \mathfrak{B}_1} \inf_{v_2 \in \mathfrak{B}_2} \frac{||v_2 - v_1||_\mathcal{V}}{||v_1||_\mathcal{V}}. \tag{18}$$

Note that $\delta(\mathfrak{B}_1, \mathfrak{B}_2) = 0$ does not imply that $\mathfrak{B}_1 = \mathfrak{B}_2$, but rather that $\mathfrak{B}_1$ is equal to a subspace of $\mathfrak{B}_2$. We can conclude that $\mathfrak{B}_1 = \mathfrak{B}_2$ if and only if both $\delta(\mathfrak{B}_1, \mathfrak{B}_2) = 0$ and $\delta(\mathfrak{B}_2, \mathfrak{B}_1) = 0$. Let now $u_1, ..., u_r$, $r = \dim \mathfrak{B}_1$ be an orthonormal basis for $\mathfrak{B}_1$, then

$$\delta(\mathfrak{B}_1, \mathfrak{B}_2) \leq \sqrt{\sum_{k=1}^{\dim \mathfrak{B}_1} \inf_{v_2 \in \mathfrak{B}_2} ||v_2 - u_k||_{\mathcal{V}}^2} \ .$$

In the case when we do not have an orthonormal basis, we define the Gram matrix for vectors $u_1, ..., u_r$ by the formula $G = [(u_i, u_j)_{\mathcal{V}}]_{i,j=1,...,r}$. The condition κ number of a set of vectors $u_1, ..., u_r$ is defined as

$$\kappa = \sqrt{\sigma_1(G)\sigma_r^{-1}(G)}, \tag{19}$$

where $\sigma_1(G)$ is the largest and $\sigma_r(G)$ the smallest singular value of the matrix G. Then, the estimate of the containment gap reads

$$\delta(\mathfrak{B}_1, \mathfrak{B}_2) \leq \kappa \sqrt{\sum_{k=1}^{r} \inf_{v_2 \in \mathfrak{B}_2} \frac{||v_2 - u_k||_{\mathcal{V}}^2}{||u_k||_{\mathcal{V}}^2}} \ . \tag{20}$$

Note that this definition as well as statements hold for any pair of finite dimensional subspaces $\mathcal{A}$ and $\mathcal{B}$ of a Hilbert space $\mathcal{X}$. Since both the Gram matrix G, as well as the containment gap δ, depend on the scalar product of the Hilbert space $\mathcal{X}$ we will write $\delta_{\mathcal{X}}(\mathcal{A}, \mathcal{B})$ and $\kappa_{\mathcal{X}}(G)$ in situations where we have to notationally distinguish different Hilbert space constructions (e.g. the phase space condition number from the original space condition number).

2.3. Residual estimates based on resolvent calculus

Note that as the projections defined by (16) are not orthogonal projections in $\mathcal{V} \oplus \mathcal{V}$ we have for $\mathbf{f} \in \mathfrak{M}_\mu$, $\|\mathbf{f}\|_{\mathcal{V}\oplus\mathcal{V}} = 1$

$$\inf_{\mathbf{v} \in \mathfrak{M}_\mu^\nu} \|\mathbf{f} - \mathbf{v}\|_{\mathcal{V}\oplus\mathcal{V}} \leq \|\mathbf{f} - \mathbf{E}^\nu(\mu)\mathbf{f}\|_{\mathcal{V}\oplus\mathcal{V}} \ .$$

We will now use the resolvent calculus to compute an estimate of $\|\mathbf{f} - \mathbf{E}^\nu(\mu)\mathbf{f}\|_{\mathcal{V}\oplus\mathcal{V}}$. The following theorem is a specialization of [46, Theorem 1] to the block operator matrices which appear in the linearization (13) of the quadratic eigenvalue problem. Let us further point out that we recast the claim of the theorem slightly differently by concentrating on bounding the resolvent difference locally, that is on the vector $\mathbf{f}$ only, rather then emphasizing the "global" norm bound. This difference leads directly to residual estimates and we provide a proof to highlight the important difference to [46, Theorem 1].

Theorem 2.2. *Let $\mathbf{f}^\nu \in \mathfrak{M}_\mu^\nu$ be such that $\widetilde{\mathbf{L}}^\nu \mathbf{f}^\nu = \mu^\nu \mathbf{f}^\nu$ then*

$$\|\mathbf{f}^\nu - \mathbf{E}(\mu)\mathbf{f}^\nu\|_{\mathcal{V}\oplus\mathcal{V}} \leq \frac{\mathrm{len}(\mathcal{C}_\mu)}{2\pi} \sup_{z \in \mathcal{C}_\mu} \frac{\|(z - \mathbf{L})^{-1}\|}{|z - \mu^\nu|} \|(\mathbf{L} - \widetilde{\mathbf{L}}^\nu)\mathbf{f}^\nu\|_{\mathcal{V}\oplus\mathcal{V}} \ .$$

Proof. From (16) we compute

$$\|\mathbf{f}^\nu - \mathbf{E}(\mu)\mathbf{f}^\nu\|_{\mathcal{V}\oplus\mathcal{V}} = \frac{1}{2\pi} \| \int_{\mathcal{C}_\mu} (z - \mathbf{L})^{-1}(\mathbf{L} - \widetilde{\mathbf{L}}^\nu)(z - \widetilde{\mathbf{L}}^\nu)^{-1}\mathbf{f}^\nu \ dz \|_{\mathcal{V}\oplus\mathcal{V}}.$$

Since

$$(z-\widetilde{\mathbf{L}}^{\nu})^{-1}\mathbf{f}^{\nu}=\frac{1}{z-\mu^{\nu}}\mathbf{f}^{\nu},$$

the conclusion follows. □

Theorem 2.3. *Let A_0 be invertible and let $\lambda^{\nu}\neq 0$ and u^{ν} be such that $Q^{\nu}(\lambda^{\nu})u^{\nu}=0$. Then*

$$\mathbf{L}\mathbf{f}^{\nu}-\frac{1}{\lambda^{\nu}}\mathbf{f}^{\nu}=\frac{1}{\lambda^{\nu}}\begin{bmatrix}-A_0^{-1}(A_0u^{\nu}+\lambda^{\nu}A_1u^{\nu}+(\lambda^{\nu})^2A_2u^{\nu})\\ 0\end{bmatrix},\qquad \textit{for}\qquad \mathbf{f}^{\nu}=\begin{bmatrix}u^{\nu}\\ \lambda^{\nu}u^{\nu}\end{bmatrix},$$

and

$$\inf_{\mathbf{v}\in\mathfrak{M}_{\lambda^{-1}}}\|\mathbf{f}^{\nu}-\mathbf{v}\|_{\mathcal{V}\oplus\mathcal{V}}\leq C_{\mathcal{C}_\mu}\|Q(\lambda^{\nu})u^{\nu}\|_{\mathcal{V}}$$

where

$$C_{\mathcal{C}_\mu}=\frac{\mathrm{len}(\mathcal{C}_\mu)}{2\pi}\sup_{z\in\mathcal{C}_\mu}\frac{\|(z-\mathbf{L})^{-1}\|}{|z-\frac{1}{\lambda^{\nu}}|}\frac{\|A_0^{-1}\|}{\lambda^{\nu}}\ .$$

Proof. The proof is a direct consequence of Theorem 2.2 and a straightforward computation with the block matrix representation for $\mathbf{L}$ as given in (8) and the fact that $\widetilde{\mathbf{L}}^{\nu}\mathbf{f}^{\nu}=\mu^{\nu}\mathbf{f}^{\nu}$ and so $(\mathbf{L}-\widetilde{\mathbf{L}}^{\nu})\mathbf{f}^{\nu}=\mathbf{L}\mathbf{f}^{\nu}-\mu^{\nu}\mathbf{f}^{\nu}$. □

Theorem 2.4. *Assume that a circle $\mathcal{C}_\mu\in\rho(\mathbf{L})$ encloses $\mu=\lambda_0^{-1}\in\sigma(\mathbf{L})$ of geometric multiplicity r and no other elements of $\sigma(\mathbf{L})$ and that we are given a basis of Galerkin vectors u_i^{ν} such that all $\mu_i^{\nu}=(\lambda_i^{\nu})^{-1}$, $i=1,...,r$ are inside $\mathcal{C}_\mu\subset\rho(\mathbf{L})$. Then*

$$\delta_{\mathcal{V}\oplus\mathcal{V}}(\mathfrak{M}_\mu,\mathfrak{M}_\mu^{\nu})\leq\kappa_{\mathcal{V}\oplus\mathcal{V}}G_{C_\mu}\sqrt{\sum_{i=1}^{r}\|Q^{\nu}(\lambda_i^{\nu})u_i^{\nu}\|_{\mathcal{V}}^2}\ ,$$

in the case in which μ is semisimple it also holds

$$\delta_{\mathcal{V}\oplus\mathcal{V}}(\mathfrak{M}_\mu^{\nu},\mathfrak{M}_\mu)\leq\kappa_{\mathcal{V}\oplus\mathcal{V}}G_{C_\mu}\sqrt{\sum_{i=1}^{r}\|Q^{\nu}(\lambda_i^{\nu})u_i^{\nu}\|_{\mathcal{V}}^2}\ ,$$

Here $\kappa_{\mathcal{V}\oplus\mathcal{V}}$ is the condition number (19) *and $\mathfrak{M}_\mu^{\nu}$ is the span of vectors $\mathbf{f}^{\nu_i}$, $i=1,\cdots,r$.*

Proof. Since $\delta_{\mathcal{V}\oplus\mathcal{V}}(\mathfrak{M}_\mu^{\nu},\mathfrak{M}_\mu)\to 0$ for $\nu\to 0$ it follows from [46] that

$$\delta_{\mathcal{V}\oplus\mathcal{V}}(\mathfrak{M}_\mu,\mathfrak{M}_\mu^{\nu})\leq\frac{\delta_{\mathcal{V}\oplus\mathcal{V}}(\mathfrak{M}_\mu^{\nu},\mathfrak{M}_\mu)}{1-\delta_{\mathcal{V}\oplus\mathcal{V}}(\mathfrak{M}_\mu^{\nu},\mathfrak{M}_\mu)}\leq C\delta_{\mathcal{V}\oplus\mathcal{V}}(\mathfrak{M}_\mu^{\nu},\mathfrak{M}_\mu).\tag{21}$$

We define the constant

$$G_{C_\mu}=C\frac{\mathrm{len}(\mathcal{C}_\mu)}{2\pi}\sup_{z\in\mathcal{C}_\mu}\frac{\|(z-\mathbf{L})^{-1}\|}{|z-\frac{1}{\lambda^{\nu}}|}\frac{\|A_0^{-1}\|}{\lambda^{\nu}}\ ,$$

and use (20) and Theorem 2.3 to obtain the first estimate. □

Corollary 2.5. *Let $\mathfrak{B}_\mu^\nu$ be the linear span of the set of vectors u_i^ν, $i = 1, ..., r$ and let $\mathfrak{B}_\mu$ denote the linear span of eigenvectors and associated generalized eigenvectors of $\mu = \lambda^{-1}$. Moreover, let $\kappa_\mathcal{V}$ be the condition number of the basis u_i^ν, $i = 1, ..., r$. Then*

$$\delta_\mathcal{V}(\mathfrak{B}_\mu, \mathfrak{B}_\mu^\nu) \leq \kappa_\mathcal{V} G_{C_\mu} \sqrt{\sum_{i=1}^{r} \|Q^\nu(\lambda_i^\nu) u_i^\nu\|_\mathcal{V}^2} \ ,$$

and in the case in which μ is semisimple we have in addition

$$\delta_\mathcal{V}(\mathfrak{B}_\mu^\nu, \mathfrak{B}_\mu) \leq \kappa_\mathcal{V} G_{C_\mu} \sqrt{\sum_{i=1}^{r} \|Q^\nu(\lambda_i^\nu) u_i^\nu\|_\mathcal{V}^2} \ .$$

Proof. Note that for any $u \in \mathfrak{B}_\mu$ and $v \in \mathfrak{B}_\mu^\nu$ we have $\|u - v\|_\mathcal{V} \leq \sqrt{\|u - v\|_\mathcal{V}^2 + \|\mu\ u + \mu^\nu\ v\|_\mathcal{V}}$, and so the proof follows analogously as was done in the proof of Theorem 2.4. □

2.4. More detailed subspace estimates

This section is presented primarily for theoretical reasons. It purpose is to present more detailed estimates for invariant subspace approximations as well as to show the scope of the resolvent calculus approach to residual estimation. We will also give a geometrical interpretation of the approximation error measure $\delta(\mathfrak{M}_\mu, \mathfrak{M}_\mu^\nu)$ and propose it as an alternative to standard approximation error measures in the case when we are dealing with subspace approximations.

A simple invariant pair is an extension of the notion of the eigenvalue and an eigenvector. It is constructed to give a coordinate representation of the action of an operator on a particular invariant space.

Definition 2.6. Let $\mathcal{V}$ be a Hilbert space and $\mathbf{A}$ a bounded operator. A pair (X, M), where $X : \mathbb{C}^n \to \mathcal{V}$ is a bounded operator such that X^*X is invertible and $M \in \mathbb{C}^{n\times n}$, is a simple invariant pair of rank n for $\mathbf{A}$ if $\mathbf{A}\ X = X\ M$ and the algebraic multiplicities of the eigenvalues of M coincide with the algebraic multiplicities of the corresponding eigenvalues of $\mathbf{A}$. In the case in which M is in the Jordan form, we call (X, M) a Jordan pair.

Let e_i, $i = 1, \cdots, n$ denote the canonical basis vectors of $\mathbb{C}^n$. The operator $X : \mathbb{C}^n \to \mathcal{V}$ in Definition 2.6 is frequently called a quasi-matrix and

$$\|X\|_F := \sqrt{\mathrm{trace}(X^*X)} = \sqrt{\sum_{i=1}^{n} \|Xe_i\|_\mathcal{V}^2}, \tag{22}$$

denotes its Frobenius or Hilbert-Schmidt norm. For details and proofs see e.g. [52] and the references therein. We introduce the notion of a quasimatrix since it allows us to obtain a convenient coordinate parametrization of its image space

$$\mathrm{Ran}(X) = \{Xx \ : \ x \in \mathbb{C}^n\}.$$

Note that $(X^*X)_{ij} = (Xe_i, Xe_j)_\mathcal{V}$ is just a notation for a Gram matrix from (19). Recall that $\mathbf{L}$ and the quadratic polynomial Q, where $Q(z) = A_0 + A_1 z + A_2 z^2$, have the same spectra, providing A_0 is invertible.

Lemma 2.7. *Let* $Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} : \mathbb{C}^n \to \mathcal{V} \oplus \mathcal{V}$ *and let* (Y, M) *be a simple invariant pair of* $\mathbf{L}$ *with* M *invertible. Then* $Y_1 = Y_2 M$ *and*

$$A_0 Y_1 + A_1 Y_1 M^{-1} + A_2 Y_1 M^{-2} = 0.$$

Proof. First,

$$\begin{aligned} \mathbf{L}\, X - XM &= \begin{bmatrix} -A_0^{-1}A_1 & -A_0^{-1}A_2 \\ I & 0 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} - \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} M \\ &= \begin{bmatrix} -A_0^{-1}\big[A_1 Y_1 M^{-1} + A_2 Y_2 M^{-1} + A_0 Y_1\big] M \\ Y_1 - Y_2 M \end{bmatrix} = 0 \ . \end{aligned}$$

Hence $Y_1 = Y_2 M$ and so

$$A_0 Y_1 + A_1 Y_1 M^{-1} + A_2 Y_1 M^{-2} = 0 \ .$$

□

The pair (Y_1, M^{-1}), which we constructed in the lemma, is called a simple invariant pair for Q. To simplify the notation let $W := M^{-1}$. Then, note that (Y_1, W) is a simple invariant pair of Q if and only if $(Y_1 S^{-1}, SWS^{-1})$ is a simple invariant pair for any invertible matrix S. More to the point we have the following definition from [12].

Definition 2.8. A pair (Y, W), where $Y : \mathbb{C}^n \to \mathcal{V}$ and $W \in \mathbb{C}^{n\times n}$, is called an invariant pair of rank n, $n \in \mathbb{N}$ for the quadratic polynomial Q if the following holds

1. $Y^*Y + W^*Y^*YW$ is invertible
2. $A_0 Y + A_1 YW + A_2 YW^2 = 0$.

Let $l \in \mathbb{N}$ be the largest integer such that there exists an invariant pair $(\widetilde{Y}, \widetilde{W})$ of Q of rank l for which $\sigma(W) = \sigma(\widetilde{W})$. We define the multiplicity of (Y, W) as the number $l - n + 1$. An invariant pair of multiplicity one is called a simple invariant pair.

For a detailed discussion of simple and minimal invariant pairs in the matrix setting we refer to [11, 12]. Note that instead of using a measure which will be invariant for a coordinate transformation

$$(Y_1, W) \mapsto (Y_1 S^{-1}, SWS^{-1}) \ , \tag{23}$$

we have opted to compare the invariant subspaces of the linearization $\mathbf{L}$. There are several reasons for this. First, in the case in which $\sigma(W)$, consists solely of semisimple eigenvalues we can construct matrices W and W^ν as diagonal matrices. Then the containment gap $\delta(\mathfrak{M}_\mu, \mathfrak{M}_\mu^\nu)$ measures the distance between the graphs of the matrices W and W^ν in the geometry of the space $\mathcal{V} \oplus \mathcal{V}$. Recall, a graph of a matrix $W \in \mathbb{C}^{n\times n}$ is the space $\mathrm{Graph}(W) = \{x \oplus (Wx) \ : \ x \in \mathbb{C}^n\}$ and

$$\mathrm{Graph}_{\mathcal{V}\oplus\mathcal{V}}(W) = \{(Y_1 x) \oplus (Y_1 W x) \ : \ x \in \mathbb{C}^n\} \subset \mathcal{V} \oplus \mathcal{V}$$

is isomorphic to $\mathrm{Graph}(W)$.

Second, the examples that we present later in the article appear to have semisimple eigenvalues. The abstract spectral theory cannot guarantee this, but we were always able to construct a simple Jordan pair for $\mathbf{L}^\nu$ with W^ν being a diagonal matrix. The corresponding condition numbers κ were

always moderate and hence did not indicate the presence of associated generalized eigenvectors. However, should the condition numbers κ in Theorem 2.4 be too large, and a use of better conditioned basis is necessary or only desirable, then we might construct W in e.g. Schur form. Note that the inequality is still true, only the estimate might end up as being to coarse. The estimates for this case are given by Corollary 2.5.

Corollary 2.9. *Let* $F^\nu = \begin{bmatrix} Y^\nu \\ Y^\nu (M^\nu)^{-1} \end{bmatrix} : \mathbb{C}^n \to \mathcal{V} \oplus \mathcal{V}$ *and* $M^\nu \in \mathbb{C}^{n\times n}$ *be a simple invariant pair of* $\mathbf{L}^\nu$ *such that* $\sigma(M^\nu)$ *is enclosed by* $\mathcal{C}_\mu \subset \rho(\mathbf{L})$*. Assume that* $\mathcal{C}_\mu$ *encloses only the eigenvalue* μ *of algebraic multiplicity* n *and geometric multiplicity* r*. Then*

$$\|F^\nu - \mathbf{E}(\mu)F^\nu\|_F \leq \frac{\mathrm{len}(\mathcal{C}_\mu)}{2\pi} \sup_{z\in\mathcal{C}_\mu} \left\{ \|(z-\mathbf{L})^{-1}\| \|(z-M^\nu)^{-1}\| \right\} \|\mathbf{L}\ F^\nu - F^\nu M^\nu\|_F \ .$$

Set $\kappa = \sigma_1(G)/\sigma_n(G)$*, then for* $G_{ij} = (F^\nu e_i, F^\nu e_j)_{\mathcal{V}\oplus\mathcal{V}}$*, we have*

$$\delta(\mathrm{Ran}(F^\nu), \mathfrak{M}_\mu) \leq \kappa \frac{\mathrm{len}(\mathcal{C}_\mu)}{2\pi} \sup_{z\in\mathcal{C}_\mu} \left\{ \|(z-\mathbf{L})^{-1}\| \|(z-M^\nu)^{-1}\| \right\} \|\mathbf{L}\ F^\nu - F^\nu M^\nu\|_F \ . \tag{24}$$

Moreover, the following estimate holds

$$\|\mathbf{L}\ F^\nu - F^\nu M^\nu\|_F \leq \|A_0^{-1}\| \|M^\nu\| \|A_0 Y^\nu + A_1 Y^\nu W^\nu + A_2 Y^\nu (W^\nu)^2\|_F \ , \tag{25}$$

with $W^\nu = (M^\nu)^{-1}$*.*

Proof. For the proof we directly combine Theorem 2.4 and Corollary 2.5 together with the definition of the Frobenius norm of a quasimatrix (22). □

Note that the statement of Corollary 2.9 can equivalently be formulated for the quadratic polynomial Q. The gap $\delta(\mathrm{Ran}(F^\nu), \mathfrak{M}_\mu)$ is a containment gap between two graph spaces which represent local actions of operator $\mathbf{L}$ and $\mathbf{L}^\nu$ as defined by the contour $\mathcal{C}_\mu$. Any other approximation measure which measures the quality of a restriction of an operator to the subspace must take into account both eigenvalues as well as eigenvectors. A use of a graph distance is standard operator theoretic tool, but less frequently used as an approximation measure in finite element computations. We argue that it is appropriate for dealing with multiplicity since it is an appropriate measure for comparing two invariant pairs (X, W) and (X^μ, W^μ), where

$$\widetilde{\mathfrak{M}}_\mu = \mathrm{Ran}(\begin{bmatrix} X \\ XW \end{bmatrix}), \qquad \mathfrak{M}_\mu = \mathrm{Ran}(\begin{bmatrix} X^\mu \\ X^\mu W^\mu \end{bmatrix}).$$

Since given (X, W) we can compute the desired spectral information using standard robust procedures of numerical linear algebra on the matrix W. We emphasize that the equations (24) and (25) offer a natural opportunity to incorporate the effect of numerical linear algebra in the overall approximation process.

Computing an approximation to the simple invariant pair belonging to a collection of eigenvalues enclosed by a contour is an appropriate solution to the eigenvalue approximation problem associated to a cluster of eigenvalues (i.e. a group of eigenvalues enclosed by a contour) The local condition numbers κ, $\|M^\nu\| = \|(W^\nu)^{-1}\|$, together with the norm of the residual $\|A_0 Y^\nu + A_1 Y^\nu W^\nu + A_2 Y^\nu (W^\nu)^2\|_F$ are indicators which represent the mixture of the stability measures for the quality

of the computed Galerkin discretization. The quality of the discretization is primarily measured by the size of the norm of the residual, whereas $\kappa\|(W^\nu)^{-1}\|$ measure both the local condition number for the linearization as well as the influence of the choice of the basis for the approximate eigenspace. We argue that this measures the influence of numerical linear algebra in the overall process. Note that the residual norm (25) does not depend on the linearization, but is a property of the quadratic eigenvalue problem. This line of argument has been further elaborated in [12] and [11]. On the other hand κ depends directly on the linearization used. Large κ indicates that the basis for the eigenspace is close to being linearly dependent. This in turn indicates that we should construct a different basis for the approximate eigenspace. This is achieved by constructing an appropriate matrix S^ν in (23) which in turn changes all of the local stability numbers κ, $\|M^\nu\| = \|(W^\nu)^{-1}\|$ and $\|A_0Y^\nu + A_1Y^\nu W^\nu + A_2Y^\nu(W^\nu)^2\|_F$. Obviously the optimal choice, one giving the minimal estimate, is an open question. The possibilities which we consider are the choice of S^ν such that $S^\nu\ W^\nu(S^\nu)^{-1}$ is diagonal (e.g. semisimple Jordan form), or the choice of the orthogonal S such that $S^\nu\ W^\nu(S^\nu)^{-1}$ is upper triangular with eigenvalues appearing on the diagonal, e.g. $S^\nu\ W^\nu(S^\nu)^{-1}$ is in the Schur form. This will have the effect of minimizing κ. In our experiments in photonic crystal applications the κ for the semisimple Jordan form was always reasonable. We could not prove a priori that this must be so, and so we only rigorously conclude that by reducing the norm of the residual we have converged to the subspace of the algebraic eigenspace which is asymptotically tending to the subspace spanned by the eigenvectors (in the case in which there might be associated generalized eigenvectors).

2.5. Variational estimates of the residual

We will now present variational estimates of the subspace residual . First, we will present a variational formula to estimate the norm $\|Q(\lambda^\nu)u^\nu\|_{\mathcal{V}}$. Recall that

$$(Q(\lambda^\nu)u^\nu, v)_{\mathcal{V}} = \mathfrak{a}(\lambda^\nu;\omega)[u,v], \qquad v \in \mathcal{V}.$$

Then we have

$$\begin{aligned}\|Q(\lambda^\nu)u^\nu\|_{\mathcal{V}} &= \sup_{v\in\mathcal{V},v\neq 0} \frac{|(Q(\lambda^\nu)u^\nu, v)|}{\|v\|_{\mathcal{V}}} \\ &= \sup_{v\in\mathcal{V},v\neq 0} \frac{|\mathfrak{a}(\lambda^\nu)[u^\nu, v]|}{\|v\|_{\mathcal{V}}} \\ &=: \|\mathfrak{a}(\lambda^\nu)[u^\nu,\cdot]\|_{\mathcal{V}^*}\ .\end{aligned}$$

In the experiments section we will use direct estimates of the dual norm of the residual. In these examples the space $\mathcal{V}^*$ will be the dual of the first order Sobolev space, and so we can use direct residual estimators which are based on integration by parts formulae.

In the case in which we have a simple invariant pair $Y_1 : \mathbb{C}^n \to \mathcal{V}$, $W \in \mathbb{C}^{n\times n}$, we proceed using (22). First note that for $f \in \mathbb{C}^n$

$$\|A_0Y_1f + A_1Y_1Wf + A_2Y_1W^2f\|_{\mathcal{V}} = \|\mathfrak{a}_0[Y_1f,\cdot] + \mathfrak{a}_0[Y_1Wf,\cdot] + \mathfrak{a}_0[Y_1W^2f,\cdot]\|_{\mathcal{V}^*}, \tag{26}$$

and by a direct calculation we obtain the Frobenius norm

$$\|A_0Y_1 + A_1Y_1W + A_2Y_1W^2\|_F = \sqrt{\sum_{i=1}^n \|\mathfrak{a}_0[Y_1e_i,\cdot] + \mathfrak{a}_0[Y_1We_i,\cdot] + \mathfrak{a}_0[Y_1W^2e_i,\cdot]\|^2_{\mathcal{V}^*}}\ . \tag{27}$$

Formula (27) will not use the in practical computations. However, we will show in Section 4.2 how to construct an efficient and reliable estimator of (27).

3. Application to photonic crystal

In a photonic crystal, a complex wave vector k describes a wave attenuated along the direction of propagation for a given (real) frequency ω. These Bloch-waves with a complex wave vector are for example used to solve source problems in photonic crystals and photonic crystal wave-guides of finite size [16, 34, 14].

In this paper, we study electromagnetic wave propagation in a non-magnetic material with the relative permittivity $\epsilon(x_1, x_2)$ independent of the third coordinate x_3. The x_3- independent electromagnetic wave (E, H) is decomposed into transverse electric (TE) polarized waves and transverse magnetic (TM) polarized waves [18]. This decomposition reduces the spectral problem for the Maxwell operator to one scalar problem for H_3 and one scalar problem for E_3.

For the TM case with a real-valued permittivity (no losses), basic properties of the spectrum were derived in [24]. Recently, both polarizations were studied for a complex-valued (lossy) permittivity function [21].

3.1. TE and TM waves

A x_3-independent transverse magnetic (TM) wave is an electromagnetic wave (E, H), where the electric field is in the form $E = (0, 0, E_3)$ and the magnetic field is in the form $H = (H_1, H_2, 0)$. The TM-waves with frequency ω can then be determined from the scalar equation

$$-\Delta E_3 - \omega^2 \epsilon(x, \omega) E_3 = 0, \quad x \in \mathbb{R}^2, \tag{28}$$

and Maxwell's equations [18]. Similarly, the TE polarized waves $(E, H) = (E_1, E_2, 0, 0, 0, H_3)$ can be determined from

$$-\nabla \cdot \left(\frac{1}{\epsilon(x, \omega)} \nabla H_3 \right) - \omega^2 H_3 = 0 \tag{29}$$

and Maxwell's equations. Let Γ denote the lattice $\mathbb{Z}^2$ and denote by $\Omega = (0, 1]^2$ the unit cell of the lattice Γ. The dual lattice to Γ is

$$\Gamma^* = \{ q \in \mathbb{R}^2 \, : \, \gamma \cdot q \in 2\pi\mathbb{Z} \ , \ \forall \gamma \in \Gamma \} \tag{30}$$

and we define the Brillouin zone of the dual lattice Γ^* as the set

$$\Omega^* = (-\pi, \pi]^2. \tag{31}$$

A Bloch solution of (28) or (29) is a non-zero solution of the form

$$E_3(x) = \mathrm{e}^{\mathrm{i}k \cdot x} u(x), \text{ respectively } H_3(x) = \mathrm{e}^{\mathrm{i}k \cdot x} u(x), \tag{32}$$

where u is a Γ-periodic function and $k \in \mathbb{C}^2$ is the Floquet-Bloch wave vector [40, p. 152]. Since $\nabla(\mathrm{e}^{\mathrm{i}k \cdot x} u(x)) = \mathrm{e}^{\mathrm{i}k \cdot x}(\nabla + \mathrm{i}k) u(x)$ the Bloch solutions of (28),(29) are Γ-periodic solutions of

$$\text{TM:} \quad -(\nabla + \mathrm{i}k) \cdot (\nabla + \mathrm{i}k) u(x) - \omega^2 \epsilon(x, \omega) u = 0, \tag{33}$$

$$\text{TE:} \ -(\nabla + \mathrm{i}k) \cdot \left(\frac{1}{\epsilon(x, \omega)} (\nabla + \mathrm{i}k) u(x) \right) - \omega^2 u = 0. \tag{34}$$

The frequency ω as a multi-valued function of the wave vector k is called the dispersion relation and the graph of the dispersion relation defines the Bloch variety [40].

3.2. The variationally posed quadratic eigenvalue problems

We assume that electromagnetic energy may be transferred into the material, but electromagnetic energy is not transferred from the material into the electromagnetic field. Materials with this property, which called are called passive [18, 22], satisfy the condition

$$\omega\epsilon(\omega) \in \mathbb{C}_+ = \{z \in \mathbb{C} : \, 0 \leq \arg z < \pi, \, z \neq 0\} \quad \forall \omega \in \mathbb{C}_+. \tag{35}$$

An important consequence of (35) is $\Im\epsilon(\omega) \geq 0$ for $\omega > 0$. Assume that $\epsilon(\cdot, \omega) \in \mathrm{L}^\infty(\mathbb{T}^2)$ and that it exists positive constants c_0, c_1 such that

$$0 < c_0 \leq |\epsilon(x)| \leq c_1 \tag{36}$$

for almost all $x \in \mathbb{R}^2$. In our setting $\omega > 0$ is fixed and $k = \lambda\hat{k}$, where $\lambda \in \mathbb{C}$ and $\hat{k}$ is a fixed unit vector in $\mathbb{R}^2$. This means that, as in [32, 19, 23], the solutions $(k, \omega) \in \mathbb{C}^2 \times \mathbb{R}^+$ have collinear real and imaginary parts of k.

Let $\mathbb{T}^2 = \mathbb{R}^2/\Gamma$ denote the torus in two dimensions. The Sobolev space $H^1(\mathbb{T}^2)$ can be characterized by considering its Fourier series with coefficients $\hat{u}(n) \in \mathbb{C}$, $n \in \mathbb{Z}^2$. Then $u \in H^1(\mathbb{T}^2)$ if and only if

$$\sum_{n \in \mathbb{Z}^2} (1 + |2\pi n|^2)^2 |\hat{u}(n)|^2 < \infty. \tag{37}$$

Define the continuous sesquilinear forms

$$\mathfrak{s}_n : H^1(\mathbb{T}^2) \times H^1(\mathbb{T}^2) \to \mathbb{C}, \tag{38}$$

where $n = 0, 1, 2$ and

$$\begin{aligned} \mathfrak{s}_0(\omega)[u, v] &= \int_\Omega \nabla u \cdot \nabla \overline{v} - \omega^2 \epsilon u \overline{v} \, \mathrm{dx}, \quad \mathfrak{s}_1[u, v] = 2\mathrm{i} \int_\Omega u\hat{k} \cdot \nabla \overline{v} \, \mathrm{dx}, \\ \mathfrak{s}_2[u, v] &= \int_\Omega u \overline{v} \, \mathrm{dx}. \end{aligned} \tag{39}$$

The variationally posed spectral problem for TM-waves (33) is: Find vectors $u \in H^1(\mathbb{T}^2)\backslash\{0\}$ and complex numbers λ satisfying

$$\mathfrak{s}(\lambda; \omega)[u, v] = 0, \quad \mathfrak{s}(\lambda; \omega)[u, v] = \lambda^2 \mathfrak{s}_2[u, v] + \lambda \mathfrak{s}_1[u, v] + \mathfrak{s}_0(\omega)[u, v], \tag{40}$$

for all $v \in H^1(\mathbb{T}^2)$. Based on Riesz representation theorem we derive a quadratic operator polynomial formulation of (38). The variational problem (40) can therefore be written as the operator equation

$$Q_{\mathrm{TM}}(\lambda; \omega)u = 0, \quad Q_{\mathrm{TM}}(\lambda; \omega) = S_0(\omega) + \lambda S_1 + \lambda^2 S_2 \tag{41}$$

in $H^1(\mathbb{T}^2)$. The variationally posed quadratic eigenvalue problem for TE-waves (34) is derived in the same way. Define the continuous sesquilinear forms

$$\mathfrak{t}_n : H^1(\mathbb{T}^2) \times H^1(\mathbb{T}^2) \to \mathbb{C}, \tag{42}$$

where $n = 0, 1, 2$ and

$$\begin{aligned} \mathfrak{t}_0(\omega)[u, v] &= \int_\Omega \frac{1}{\epsilon} \nabla u \cdot \nabla \overline{v} - \omega^2 u \overline{v} \, \mathrm{dx}, \quad \mathfrak{t}_1(\omega)[u, v] = 2\mathrm{i} \int_\Omega \frac{1}{\epsilon} u\hat{k} \cdot \nabla \overline{v} \, \mathrm{dx}, \\ \mathfrak{t}_2(\omega)[u, v] &= \int_\Omega \frac{1}{\epsilon} u \overline{v} \, \mathrm{dx}. \end{aligned} \tag{43}$$

The variationally posed spectral problem for TE-waves is: Find vectors $u \in H^1(\mathbb{T}^2)\backslash\{0\}$ and complex numbers λ satisfying

$$\mathfrak{t}(\lambda;\omega)[u,v] = 0, \quad \mathfrak{t}(\lambda;\omega)[u,v] = \lambda^2\mathfrak{t}_2(\omega)[u,v] + \lambda\mathfrak{t}_1(\omega)[u,v] + \mathfrak{t}_0(\omega)[u,v], \tag{44}$$

for all $v \in H^1(\mathbb{T}^2)$.

The quadratic eigenvalue problem (44) can alternatively be represented by the operators $T_n : H^1(\mathbb{T}^2) \to H^1(\mathbb{T}^2)$, with

$$(T_n(\omega)u, v)_1 = \mathfrak{t}_n[u,v] \quad \text{for all } u, v \in H^1(\mathbb{T}^2) \tag{45}$$

and (44) can therefore alternatively be written in terms of the operator polynomial

$$Q_{\mathrm{TE}}(\lambda;\omega)u = 0, \quad Q_{\mathrm{TE}}(\lambda;\omega) = T_0(\omega) + \lambda T_1(\omega) + \lambda^2 T_2(\omega) \tag{46}$$

in $H^1(\mathbb{T}^2)$.

The frequency ω is in (41) and (46) a given real number. However, the spectral properties of Q_{TM} and of Q_{TE} depend on the choice of ω.

3.3. Spectral properties

The permittivity ϵ is in applications usually piecewise constant and we therefore restrict ϵ to the finite domain partitioning $\Omega = \cup_{n=1}^{N}\Omega_n$ and assume that ϵ can be written in the form

$$\epsilon(x) = \sum_{n=1}^{N} \epsilon_n \chi_{\Omega_n}, \tag{47}$$

where χ_{Ω_n} is the indicator function for subdomain Ω_n, which is of positive measure.

In section 2, Fredholm-valued operator polynomials with invertible leading coefficient are studied. The polynomials Q_{TM} and Q_{TE} do not in general have an invertible leading coefficient but we will consider two important cases where this requirement hold.

Lemma 3.1. *Assume that $\Im\epsilon = 0$ and that ω is band-gap frequency. That is there exist no Bloch wave with a real $k = \lambda\hat{k}$ [24]. The operators function S_0 and T_0 are then invertible and the functions Q_{TM} and Q_{TE} are Fredholm-valued.*

Proof. The function $\omega \mapsto S_0(\omega)$ in (41) is Fredholm. Assume $\Im\lambda \neq 0$. If $S_0(\omega_0)u_0 = 0$, then follows $Q_{\mathrm{TM}}(\lambda)u_0 = \lambda(S_1 + \lambda S_2)u_0$, where $S_1 + \lambda S_2$ is a compact operator and $\lambda = 0$ is a real eigenvalue. Hence, Q_{TM} is a Fredholm function of index zero [24]. The proof for T_0 is identical and Q_{TE} is then a Fredholm function of index zero [21, Lemma 3]. □

Note that if $S_0(\omega_0)u_0 = 0$ the vector u_0 can in general be an eigenvector for two different eigenvalues $\lambda = 0$ and the solution of $(S_1 + \lambda S_2)u_0 = 0$.

Lemma 3.2. *Assume that for a given $\omega > 0$ the imaginary part of the permittivity ϵ is non-zero in one of the subdomains Ω_n. The operators function S_0 and T_0 are then invertible and the functions Q_{TM} and Q_{TE} are Fredholm-valued.*

Proof. We can without loss of generality assume that ϵ_1 is a positive constant and $\epsilon_n \in \mathbb{C}$, $\Im\epsilon_n > 0$ for $n = 2, 3, \ldots, N$. The operator T_0 can then be written in the form

$$(T_0 u, v) = \frac{1}{\epsilon_1}\int_{\Omega_1} \nabla u \cdot \nabla \overline{v}\,\mathrm{dx} + \sum_{n=2}^{N} \frac{1}{\epsilon_n}\int_{\Omega_n} \nabla u \cdot \nabla \overline{v}\,\mathrm{dx} - \int_{\Omega} \omega^2 u\overline{v}\,\mathrm{dx}. \tag{48}$$

The operator T_0 is a Fredholm operator of index zero [21, Lemma 3]. Assume that ω is real and $(T_0 u, v) = 0$ for all $v \in H^1(\mathbb{T}^2)$. Then follows

$$0 = \Im(T_0 u, u) = \sum_{n=2}^{N} \Im\frac{1}{\epsilon_n}\int_{\Omega_n} |\nabla u|^2\,\mathrm{dx}. \tag{49}$$

Hence, $\nabla u = 0$ for almost all $x \in \cup_{n=2}^{N}\Omega_n$. This condition on the gradient of u implies that $\omega^2 \in \sigma(T_0) \cap \mathbb{R}$ only if it exist a $u \in H^1(\mathbb{T}^2)$ such that

$$\int_{\Omega} \frac{1}{\epsilon_1}\nabla u \cdot \nabla \overline{v} - \omega^2 u\overline{v}\,\mathrm{dx} = 0 \tag{50}$$

for all $v \in H^1(\mathbb{T}^2)$. However, the solutions of this problem have the form $e^{im\cdot x}$, $m \in \mathbb{Z}^2$, which do not satisfy the condition on the gradient of u. It was proved in [21] that S_0 is invertible. □

4. Finite element approximations

Let us now discretize our model problems (40) and (44) using *hp*-finite element spaces. These two problems can for fixed $\omega > 0$ be written in the form (2), where $\mathcal{V} = H^1(\mathbb{T}^2)$. In the following sections, we assume that the leading coefficient A_0 in the corresponding operator polynomial is invertible. Two cases where we can guarantee that A_0 is invertible were proved in Lemma 3.1 and in Lemma 3.2.

Let $\mathcal{T}$ be a triangulation of $\Omega = (0, 1]^2$ with the piecewise constant mesh function $h : \mathcal{T} \to (0, 1)$, $h(T) = \mathrm{diam}(T)$ for $T \in \mathcal{T}$. Further let there be a partition $\overline{\Omega} = \cup_{k=1}^{p}\overline{\Omega}_k$ of Ω into subdomains Ω_k with disjoint interiors and whose boundaries are piecewise smooth, such that $\epsilon(\cdot, \omega)|_{\Omega_k} \in W^{1,\infty}(\Omega_k)$ for each k.

We implicitly assume that $\mathcal{T}$ is subordinate to the polygonal partition of Ω in other words, each $T \in \mathcal{T}$ is contained in precisely one of the polygons Ω_k. Given a piecewise constant distribution of polynomial degrees, $p : \mathcal{T} \to \mathbb{N}$, we define the space

$$V_h^p = \{v \in H^1(\mathbb{T}^2) \ : \ v\big|_T \in \mathbb{P}_{p(T)} \text{ for each } T \in \mathcal{T}\}\ ,$$

where $\mathbb{P}_{p(T)}$ is the collection of polynomials of total degree not greater than p on a given element $T \in \mathcal{T}_h$ and the topology of the torus is imposed by mapping the parallelogram edges situated on one boundary on the corresponding parallelogram on opposite side.

Let $\mathcal{E}$ denote the set of edges in $\mathcal{T}$. Additionally, we let $\mathcal{T}(\mathbf{e})$ denote the two triangles having $\mathbf{e} \in \mathcal{E}$ as an edge, and we extend p to $\mathcal{E}$ by $p(\mathbf{e}) = \max_{T\in\mathcal{T}(\mathbf{e})} p(T)$. Without loss of generality, we assume that the family of spaces satisfy the following standard *regularity properties* on $\mathcal{T}$ and p: *There exists a constant* $\gamma > 0$ *for which*

($\mathfrak{C}\mathfrak{1}$) $\gamma^{-1}h(T) \le h(T') \le \gamma h(T)$ for adjacent $T, T' \in \mathcal{T}$, $\overline{T} \cap \overline{T'} \neq \emptyset$. In other words, the diameters of adjacent elements are comparable.

($\mathfrak{C}2$) $\gamma^{-1}(p(T)+1) \le p(T')+1 \le \gamma(p(T)+1)$ for adjacent $T, T' \in \mathcal{T}$, $\overline{T} \cap \overline{T'} \neq \emptyset$. In other words, the polynomial degrees associated with adjacent elements are comparable.

Let us now define the indicator function which will be used to compute an estimate of the residual. Given $u_{h,p} \in V_h^p \setminus \{0\}$ and $\lambda_{h,p} \in \mathbb{C}$ we define the element residual, $R_T(u_{h,p}, \lambda_{h,p})$, and edge residual, $R_{\mathbf{e}}(u_{h,p}, \lambda_{h,p})$, by

$$\begin{aligned} R_T(u_{h,p}, \lambda_{h,p}) &:= (-\nabla \cdot \alpha \nabla u_{h,p} - 2i\lambda_{h,p}\alpha \hat{k} \cdot \nabla u_{h,p} + (\alpha \lambda_{h,p}^2 - \omega^2 \beta) u_{h,p}) \ , \\ R_{\mathbf{e}}(u_{h,p}, \lambda_{h,p}) &:= -(\alpha \nabla u_{h,p} + 2i\lambda_{h,p}\alpha \hat{k} u_{h,p})_{|_T} \cdot \mathbf{n}_T - (\alpha \nabla u_{h,p} + 2i\lambda_{h,p}\alpha \hat{k} u_{h,p})_{|_{T'}} \cdot \mathbf{n}_{T'} \ , \end{aligned} \tag{51}$$

where T and T' are the two adjacent elements of $\mathbf{e} \in \mathcal{E}$, having outward unit normal vectors $\mathbf{n}_T$ and $\mathbf{n}_{T'}$, respectively. Here we have allowed for the function parameters α and β which take on appropriate values whether we tackle the TE or TM case in Section 3. For the TM case we have $\alpha := 1$, $\beta := \epsilon$ and for the TE case the parameters are $\alpha := \epsilon^{-1}$ and $\beta := 1$.

We also define the dual local indicators as

$$\begin{aligned} R_T^d(u_{h,p}^d, \lambda_{h,p}) &:= (-\nabla \cdot \overline{\alpha} \nabla u_{h,p}^d - 2i\overline{\lambda_{h,p}}\overline{\alpha} \hat{k} \cdot \nabla u_{h,p}^d + (\overline{\alpha}\overline{\lambda_{h,p}}^2 - \omega^2 \overline{\beta}) u_{h,p}^d) \ , \\ R_{\mathbf{e}}^d(u_{h,p}^d, \lambda_{h,p}) &:= -(\overline{\alpha} \nabla u_{h,p}^d)_{|_T} \cdot \mathbf{n}_T - (\overline{\alpha} \nabla u_{h,p}^d)_{|_{T'}} \cdot \mathbf{n}_{T'} \ . \end{aligned} \tag{52}$$

Let $\|\cdot\|_{0,S}$ and $|\cdot|_{1,S}$ denote the L^2-norm and the H^1-seminorm over $S \subset \Omega$, respectively.

As our global indicator we take

$$\eta(u_{h,p}, \lambda_{h,p})^2 := \sum_{T \in \mathcal{T}} \left(\frac{h(T)}{p(T)} \right)^2 \|R_T(u_{h,p}, \lambda_{h,p})\|_{0,T}^2 + \sum_{\mathbf{e} \in \mathcal{E}} \frac{h(\mathbf{e})}{p(\mathbf{e})} \|R_{\mathbf{e}}(u_{h,p}, \lambda_{h,p})\|_{0,\mathbf{e}}^2 \ .$$

and for the dual residual we use

$$\eta_d(u_{h,p}^d, \lambda_{h,p})^2 := \sum_{T \in \mathcal{T}} \left(\frac{h(T)}{p(T)} \right)^2 \|R_T^d(u_{h,p}^d, \lambda_{h,p})\|_{0,T}^2 + \sum_{\mathbf{e} \in \mathcal{E}} \frac{h(\mathbf{e})}{p(\mathbf{e})} \|R_{\mathbf{e}}^d(u_{h,p}^d, \lambda_{h,p})\|_{0,\mathbf{e}}^2 \ .$$

Below, we present a reliability estimate of the dual norm of the residual. Because we are imposing periodic boundary conditions, the edges along the boundary are considered interior edges. Then, for a given vertex z we define its neighborhood in $\mathcal{T}$ as the set of triangles ω_z which have z as a vertex. Theorem 4.1 states a result for the interpolation operator which will be used to obtain computable estimates of the residual.

Theorem 4.1 ([43]). *There is a linear operator $\mathcal{I} : \mathcal{V} \to V_h^p$ and a constant C depending only on the shape-regularity parameter γ, such that: For any vertex z and any edge $\mathbf{e}$ having z as a vertex,*

$$\|v - \mathcal{I}v\|_{0,\omega_z} + \frac{h_z}{p_z} |\mathcal{I}v|_{1,\omega_z} + \sqrt{\frac{h_z}{p_z}} \|v - \mathcal{I}v\|_{0,\mathbf{e}} \le C \frac{h_z}{p_z} |v|_{1,\Omega_z} \ .$$

Here, ω_z is the patch of triangles having z as a vertex, h_z is the largest of the diameters of these triangles, $p_z - 1$ is the largest of the polynomial degrees associated with these triangles, and $\Omega_z \supset \omega_z$ is a larger, but still localized, patch of triangles.

Remark 4.2. The precise choice of Ω_z is not essential here. It only matters that, if m_z is the number of triangles in Ω_z and $\#\mathcal{T}$ is the total number of triangles in $\mathcal{T}$, then $\sum_z m_z \le \delta(\#\mathcal{T})$ for some δ which depends only on the shape-regularity parameter γ. This is a consequence of the shape-regularity assumption ($\mathfrak{C}1$) (cf. [43]).

Theorem 4.3. *There is a constant C depending only on shape-regularity parameter γ for which*

$$\|\mathfrak{a}(\lambda_{h,p};\omega)[u_{h,p},\cdot]\|_{-1} \le C\eta(u_{h,p},\lambda_{h,p}),$$
$$\|\mathfrak{a}(\lambda_{h,p};\omega)[\cdot,u^d_{h,p}]\|_{-1} \le C\eta_d(u^d_{h,p},\lambda_{h,p}),$$

where $\|\cdot\|_{-1}$ denotes the norm of $\mathcal{V}^* = H^{-1}(\mathbb{T}^2)$.

Proof. It holds, because of Galerkin orthogonality, that

$$\begin{aligned}
|\mathfrak{a}(\lambda_{h,p};\omega)[u_{h,p},v]| &= |\mathfrak{a}(\lambda_{h,p};\omega)[u_{h,p},v-\mathcal{I}v]| \\
&\le \sum_{T\in\mathcal{T}} \|R_T(u_{h,p},\lambda_{h,p})\|_{0,T}\|v-\mathcal{I}v\|_{0,T} + \sum_{\mathbf{e}\in\mathcal{E}} \|R_{\mathbf{e}}(u_{h,p},\lambda_{h,p})\|_{0,\mathbf{e}}\|v-\mathcal{I}v\|_{0,\mathbf{e}} \\
&\lesssim \sum_{T\in\mathcal{T}} \|R_T(u_{h,p},\lambda_{h,p})\|_{0,T}\frac{h_{z(T)}}{p_{z(T)}}|v|_{1,\Omega_{z(T)}} + \sum_{\mathbf{e}\in\mathcal{E}} \|R_{\mathbf{e}}(u_{h,p},\lambda_{h,p})\|_{0,\mathbf{e}}\sqrt{\frac{h_{z(\mathbf{e})}}{p_{z(\mathbf{e})}}}|v|_{1,\Omega_{z(\mathbf{e})}} ,
\end{aligned}$$

where $z(T)$ is a vertex of T and $z(\mathbf{e})$ is a vertex of $\mathbf{e}$. The controlled overlap of patches (Remark 4.2) guarantees that

$$\sum_{T\in\mathcal{T}} |v|^2_{1,\Omega_{z(T)}} \lesssim |v|^2_1 \quad , \quad \sum_{\mathbf{e}\in\mathcal{E}} |v|^2_{1,\Omega_{z(\mathbf{e})}} \lesssim |v|^2_1 .$$

Now using the discrete Cauchy-Schwarz inequality and the fact that triangle diameters and polynomial degrees are comparable for nearby elements and edges, we see that

$$\mathfrak{a}(\lambda_{h,p};\omega)[u_{h,p},v] \lesssim \left(\sum_{T\in\mathcal{T}} \left(\frac{h(T)}{p(T)}\right)^2 \|R_T(u_{h,p},\lambda_{h,p})\|^2_{0,T} + \sum_{\mathbf{e}\in\mathcal{E}} \frac{h(\mathbf{e})}{p(\mathbf{e})}\|R_{\mathbf{e}}(u_{h,p},\lambda_{h,p})\|^2_{0,\mathbf{e}}\right)^{1/2} |v|_1 .$$

This completes the proof of the first statement. The proof of the second statement is analogous. □

Note that equation (27) can be used to formulate an analogous error estimators for the residual with the assumption that we have computed a Schur basis for a cluster of eigenvalues. We leave out the details for a subsequent paper where we will analyze a continuous problem for which we can a priori prove the existence of associated generalized eigenvectors.

4.1. Measures of the approximation error

In this paper we have opted to use the proximity measure of two eigenspaces of the linearization of the quadratic eigenvalue problem as measures of the quality of approximation. This differs somewhat from the usual approach taken in the literature. In this section we show how to derive standard results based on our estimates. Also, we point out our reasons for choosing this particular form of an estimator.

In Theorem 2.4 we have used the error measure from (18) to measure the quality of the approximation of a semisimple eigenvalue λ_0. The associated eigenspace $\mathfrak{M}_{\lambda^{-1}}$ of the linearized problem $\mathbf{L}$ consists of the functions of the form

$$\mathfrak{M}_{\lambda^{-1}} = \mathrm{Span}\left\{ \begin{bmatrix} u \\ \lambda u \end{bmatrix} \ : \ u \in H^1(\mathbb{T}^2) \text{ such that } Q(\lambda)u = 0 \right\}.$$

Let $\mathbf{v}_1 \in \mathfrak{M}_{\lambda^{-1}}$ and denote by $\mathbf{v}_2$ the vector $\mathbf{v}_2 = \begin{bmatrix} u_{h,p} \\ \lambda_{h,p} u_{h,p} \end{bmatrix}$, where $Q_{h,p}(\lambda_{h,p})u_{h,p} = 0$. Then we have

$$\frac{||\mathbf{v}_2 - \mathbf{v}_1||_{\mathcal{V}\oplus\mathcal{V}}}{||\mathbf{v}_1||_{\mathcal{V}\oplus\mathcal{V}}} = \frac{\sqrt{\|u - u_{h,p}\|_1^2 + \|\lambda u - \lambda_{h,p} u_{h,p}\|_1^2}}{\sqrt{\|u\|_1^2 + \lambda^2 \|u\|_1^2}},$$

where $\|\cdot\|_1$ denotes the norm of $\mathcal{V} = H^1(\mathbb{T}^2)$. In other words, the proximity measure $\delta(\mathfrak{B}_1, \mathfrak{B}_2)$ is an "averaged" measure of the approximation quality for a given eigenpair.

In comparison one could take the standard approach as in Pester [47] which follows form the work of Karma [35, 36] and obtain separate estimates for the approximation error in the eigenvalues as well as in eigenfunctions. We will first show these estimates in the case of a simple eigenvalue. Let as above u denote the right eigenfunction and let u^d denote the left eigenfunction. Then the distances between these functions and the approximating subspace V_h^p are given by

$$d_{h,p} = \inf_{v \in V_h^p} \|u - v\|_1, \qquad d_{h,p}^d = \inf_{v \in V_h^p} \|u^d - v\|_1 \ .$$

From [36, Theorem 3] follows

$$\frac{|\lambda - \lambda_{h,p}|}{|\lambda|} \leq C d_{h,p} d_{h,p}^d$$

and using the estimates from Theorem 2.3 we obtain

$$d_{h,p} = \inf_{v \in V_h^p} \|u - v\|_1 \leq C \|Q(\lambda_{h,p}) u_{h,p}\|_1 \leq C\eta(u_{h,p}, \lambda_{h,p}) \ .$$

Equivalently, $d_{h,p}^d$ is estimated by the left residual, which results in the practical estimate

$$\frac{|\lambda - \lambda_{h,p}|}{|\lambda|} \leq C\eta(u_{h,p}, \lambda_{h,p}) \eta_d(u_{h,p}^d, \lambda_{h,p}) \ . \tag{53}$$

Using the theory of [35, 36] it is possible to extend this estimate to semisimple and even defective eigenvalues. However, this involves intricate averaging of either estimators η or the error measures for the group of Galerkin values $\lambda^n u$.

Alternatively, based on the approach of Verfürth and using a newly dependent Clement type interpolation operator, Pester [47] has shown that for simple eigenvalues and the space of P1 Lagrange elements the following estimate holds

$$C_1 \eta^2(u_{h,p}, \lambda_{h,p}) \leq |\lambda - \lambda_{h,p}| + \|u - u_{h,p}\|_1^2 \leq C_2 \eta^2(u_{h,p}, \lambda_{h,p}) \tag{54}$$

It is straightforward to apply Verfürth's analysis from [54], as has been described in [29], in our setting. We further use the Clement type interpolation operator from Theorem 4.1 to obtain approximation error estimates. Although the constant C_2 depends on the shape regularity bound γ, and the constant C_1 depends both on γ as well as the maximal polynomial degree in V_h^p, these dependencies are not unexpected since the same holds for the boundary value estimator from [43].

Rather than to discuss these technical results, we propose that the separation distance as given by Theorem 2.4 is a good compromise to capture the approximating property for the whole algebraic eigenspace as well as the associated eigenvalue while at the same time giving an estimator for the whole group of approximating Galerkin values $\lambda_{h,p}^{(i)}$ in a simple form.

For numerical experiments we will also report the convergence history for the standard measures of the approximation error, since it will be easier to compare the performance of our estimators with what is reported in the literature elsewhere. Naturally, this is meaningful only for simple eigenvalues since both of the estimates (53) and (54) require this assumption.

4.2. A construction of a reliable end efficient residual estimator

In order to benchmark the performance of our estimator, we will develop a more reliable and efficient residual estimator. However, using this estimator to estimate the negative norm of the residuals is computationally much more expensive. The construction is based on the theory from Sections 5 and 4.1. This construction also shows one advantage of the flexibility which our operator theoretic approach to error analysis offers. The developed estimator is reliable and efficient and experiments will show effectivities close to one. We will use this method to compute benchmark results for the purposes of testing of our marking strategy.

Based on the results from [43] we have in Theorem 4.3 established a reliability estimate for the negative norm of the residual. Using technique from [43] it is easy to establish an efficiency estimate, eg. for the right residual

$$c\eta(u_{h,p},\lambda_{h,p}) \leq \|\mathfrak{a}(\lambda_{h,p};\omega)[u_{h,p},\cdot]\|_{-1} \ .$$

However, as in [43] not only does the constant c depend on the shape regularity parameter γ, but also on the maximal polynomial degree p. We note, that although in theory the efficiency bound deteriorates with the increase of the polynomial degree, we do not see this deterioration in experiments. The approach of goal oriented adaptivity is used to develop the estimator for computing the benchmark results. A similar estimator has been considered in eg. [26] for self adjoint eigenvalue problems. This will also illustrate a versatility of our approach to residual error estimation.

We define the following alternative representation of the residual. Let $\widetilde{u}_r \in \mathcal{V} := H^1(\mathbb{T}^2)$ be a vector such that

$$(\widetilde{u}_r, v)_1 = \mathfrak{a}(\lambda_{h,p};\omega)[u_{h,p},v], \qquad v \in \mathcal{V} \ . \tag{55}$$

Then, using the continuity $\mathfrak{a}(\lambda_{h,p};\omega)[\cdot,\cdot]$ with respect to $\|\cdot\|_1 = (\cdot,\cdot)_1^{1/2}$, we obtain

$$\frac{(\widetilde{u}_r,\widetilde{u}_r)_1}{\|\widetilde{u}_r\|_1} \leq \sup_{v\in\mathcal{V}\setminus\{0\}} \frac{|\mathfrak{a}(\lambda_{h,p};\omega)[u_{h,p},v]|}{\|v\|_1} \leq C \sup_{v\in\mathcal{V}\setminus\{0\}} \frac{\|\widetilde{u}_r\|_1\|v\|_1}{\|v\|_1} \tag{56}$$

and the conclusion

$$\|\widetilde{u}_r\|_1 \leq \sup_{v\in\mathcal{V}\setminus\{0\}} \frac{|\mathfrak{a}(\lambda_{h,p};\omega)[u_{h,p},v]|}{\|v\|_1} \leq C\|\widetilde{u}_r\|_1$$

follows. Obviously, computing $\widetilde{u}_r$ is as hard as solving the original problem. Therefore we compute $\widetilde{u}_{r,h'p'}$ using a higher order finite element space $V_{h'}^{p'}$ which contains as a subspace the space V_h^p which was used to compute $u_{h,p}$. We construct the space $V_{h'}^{p'}$, $V_h^p \subset V_{h'}^{p'}$, by using the standard refinement technique [43] for the source problem (55). As described in [26] we note that $\|\widetilde{u}_{r,h',p'}\|_1$ is a computable estimator for $\|\mathfrak{a}(\lambda_{h,p};\omega)[u_{h,p},\cdot]\|_{-1}$ which is both reliable and efficient. Following, [26] we will call the estimator defined by (55) the goal oriented dual weighted residual estimator and will abbreviate it by DWR. We treat the left residual $\|\mathfrak{a}(\lambda_{h,p};\omega)[\cdot,u_{h,p}^d]\|_{-1}$ equivalently.

Note that using the general nonlinear analysis of error estimators from [54] we have the estimate

$$C_1\|\mathfrak{a}(\lambda_{h,p};\omega)[\cdot,u^d_{h,p}]\|_{-1} \leq |\lambda-\lambda_{h,p}| + \|u-u_{h,p}\|_1^2 \leq C_2\|\mathfrak{a}(\lambda_{h,p};\omega)[\cdot,u^d_{h,p}]\|_{-1}\ , \tag{57}$$

where the constants C_1 and C_2 depend only on the sesquilinear form, but not on the approximation subspace [47]. Subsequently, we see that (55) provides an reliable and efficient estimator of the eigenvalue error with constants depending only on the shape regularity parameters of the space V_h^p.

Remark 4.4. The reliability constant C from (56) depends on $\lambda_{h,p}$. However, due to the continuity of the function $\mathfrak{a}(\cdot;\,\omega)[\cdot,\cdot]$ in all of its variables, it is obviously possible to obtain a uniform constant C for a given compact region of the complex plane. This is typical setting in applications since with finite element approximations one can only hope to directly approximate a finite component of the point spectrum.

Observe that (57) yields directly the efficiency and reliability for the containment gap measure of the error. Namely we have

$$\begin{aligned}\tilde{C}_1\|\mathfrak{a}(\lambda_{h,p};\omega)[\cdot,u^d_{h,p}]\|_{-1} &\leq \frac{\|u-u_{h,p}\|_1}{\|u\|_1}\\ &\leq \max\{1,|\lambda|\}\frac{\sqrt{\|u-u_{h,p}\|_1^2+\|\lambda u-\lambda_{h,p}u_{h,p}\|_1^2}}{\sqrt{\|u\|_1^2+\lambda^2\|u\|_1^2}}\\ &\leq \tilde{C}_2\|\mathfrak{a}(\lambda_{h,p};\omega)[\cdot,u^d_{h,p}]\|_{-1},\end{aligned}$$

where the last estimate follows directly from Theorem 2.4.

Remark 4.5. Finally, we return to the invariant subspace estimator for the approximate invariant pair (Y_1,W), $\mathrm{Ran}(Y_1)\subset V_h^p$, from Section 2.5. Starting from formulas (26)–(27) and using the technique from (55), we define the functions

$$(\tilde{f}_i,b) = \mathfrak{a}_0[Y_1e_i,v]+\mathfrak{a}_0[Y_1We_i,v]+\mathfrak{a}_0[Y_1W^2e_i,v],\qquad v\in\ \mathcal{V} \tag{58}$$

and construct the continuous representation

$$\|A_0Y_1+A_1Y_1W+A_2Y_1W^2\|_F = \sqrt{\sum_{i=1}^n\|\tilde{f}_i\|_{\mathcal{V}}^2}\ .$$

of the Frobenius norm of the residual. We now directly proceed to the construction of a practical estimator. Let $V_{h'}^{p'}$, $V_h^p\subset V_{h'}^{p'}$ be a given refinement of the current finite element space. Then from (26) we can construct functions $\tilde{f}_i'\in V_{h'}^{p'}$, $i=1,\cdots,n$ such that

$$(\tilde{f}_i',b) = \mathfrak{a}_0[Y_1e_i,v]+\mathfrak{a}_0[Y_1We_i,v]+\mathfrak{a}_0[Y_1W^2e_i,v],\qquad v\in\ V_{h'}^{p'}\ . \tag{59}$$

Hence, from (27) we obtain a reliable and efficient estimator for the norm of the invariant subspace residual $\|R\|_F$ since

$$\sqrt{\sum_{i=1}^n\|\tilde{f}_i'\|_{\mathcal{V}}^2} \leq \|A_0Y_1+A_1Y_1W+A_2Y_1W^2\|_F \leq C\sqrt{\sum_{i=1}^n\|\tilde{f}_i'\|_{\mathcal{V}}^2}\ . \tag{60}$$

We leave out the technical details, since the reasoning is equivalent to the semisimple case.

Remark 4.6. The use of estimates like (57) depends on the ability to reliably and efficiently compute the negative order Sobolev norms of the residuals, eg. $\|\mathfrak{a}(\lambda_{h,p};\omega)[\cdot,u^d_{h,p}]\|_{-1}$. This problem is equivalent to solving an auxiliary linear source type problem. Solving this problem, as has already been mentioned, is not feasible in a general situation. Therefore to obtain a practical bound or an estimator we solve the auxiliary problem numerically and estimate the error in the solution a-posteriori. To be specific, we have restricted the variational problem from (58) to the space $V_h^{p'}$ to obtain (59). If we have an a-posteriori error estimator for this source problem which is reliable and efficient, then the practical error estimator like (60) will inherit such properties. For the solution of problems like (59) there are simple reliable error estimators like [43]. Unfortunately, the efficiency constant in [43] depends on the polynomial degree. This feature is an artifact of the proof and we were not able to observe it in practical experiments. In such situation it is typical, as we have done in the construction of DWR estimator, to assume the saturation assumption – it amounts to assuming that the richer finite element space contains a better approximation of the source problem than the original space – and to use the solution of the auxiliary problem as an error estimator rather then as a bound. Technically, the saturation assumption allows us to prove that the part of the error, which is not bounded by the estimator, decays at a higher asymptotic rate and can thus be ignored. This is what we see in experiments like presented in Figure 5(b) in the next section. We emphasize, for a reader's convenience, that a bound like (57) presents an idealized but provably efficient and reliable error bound. It is at the point where we start computing negative order Sobolev norms that we either obtain a somewhat pessimistic bound as in Theorem 4.3 or an efficient and reliable DWR estimator (at the price of a saturation assumption cf. [17] for a justification of making a saturation assumption in a similar context). This can best be observed by comparing Figures 1 and 5 in the next section. The estimators η and η_d are provably reliable, but somewhat pessimistic. However, the DWR estimator whose reliability and efficiency depends on the saturation assumption has measured effecitivities close to one on our model problems.

5. Experiments

In this section we provide several numerical results which illustrate the efficiency of our a-posteriori error estimators and the exponential convergence of the error on a sequence of *hp*-adapted meshes. Following [6], we assume an eigenvalue error model of the form

$$\lambda_{h,p} = \lambda + Ce^{-2\gamma\sqrt{\#\mathrm{DOFs}}}\,, \tag{61}$$

for problems with analytic eigenvectors, and

$$\lambda_{h,p} = \lambda + Ce^{-2\gamma\sqrt[3]{\#\mathrm{DOFs}}}\,, \tag{62}$$

for problems possessing discontinuous coefficients in the second order term, which are expected to have eigenvectors with isolated singularities. The constants C and γ are determined by least-squares fitting [45]. The value of γ is reported for each problem and in all convergence plots a straight line of slope γ is added for comparison. Although all convergence rates in the experiments are seen to be exponential, with one of the two error models above, we will abuse terminology slightly in the experiments by referring to γ as the *convergence rate*; the context will make it clear if γ is to be associated with the model (61) or (62).

For a given approximate eigentriple $(\lambda_{h,p}, u_{h,p}, u^d_{h,p})$, we are interested in the relative eigenvalue error, its *a-posteriori* estimate, and the associated effectivity index. These quantities are given by

$$\frac{|\lambda_{h,p} - \lambda|}{|\lambda_{h,p}|}, \qquad \frac{\eta(u_{h,p}, \lambda_{h,p})\eta_d(u^d_{h,p}, \lambda_{h,p})}{|\lambda_{h,p}|}, \qquad \frac{\eta(u_{h,p}, \lambda_{h,p})\eta_d(u^d_{h,p}, \lambda_{h,p})}{|\lambda_{h,p} - \lambda|}.$$

Similarly for the eigenvectors $u_{h,p}$, $u^d_{h,p}$, we analyze the eigenvector errors, their *a-posteriori* estimates, and the associated effectivity indices:

$$\|u - u_{h,p}\|_1, \qquad \eta(u_{h,p}, \lambda_{h,p}), \qquad \eta(u_{h,p}, \lambda_{h,p})/\|u - u_{h,p}\|_1,$$

$$\|u^d - u^d_{h,p}\|_1, \qquad \eta_d(u^d_{h,p}, \lambda_{h,p}), \qquad \eta_d(u^d_{h,p}, \lambda_{h,p})/\|u^d - u^d_{h,p}\|_1.$$

Since the exact eigenvalues are not known for the problem under examination, we use highly accurate computations on very fine grids and adapted finite element spaces generated by the method described below to produce "exact eigenvalues" for our comparisons. In order to compute accurately the errors for eigenvectors on a sequence of refined spaces, it is necessary that all the eigenvectors in the sequence are approximations of the same continuous eigenvector. In general this is not true even for simple eigenvalues. It is common that the computed eigenvectors for the same eigenvalue on two consecutive refined spaces do not approximate the same continuous eigenvector, but two continuous eigenvectors for the same eigenvalue. For eigenvalues with multiplicity more than one, the situation is even more complicate because the same continuous eigenspace can split differently in discrete eigenspaces on two consecutive refined spaces. To recover the errors for eigenvectors, we used the reconstruction technique in [51], which has been extended to complex valued problems. Such technique guarantees that the computed eigentriples $(\lambda_{h,p}, u_{h,p}, u^d_{h,p})$ are always approximations of the same continuous eigentriple. On each finite element space the computed eigentriple is in general not an eigentriple of the discrete problem, but a linear combinations of one or more discrete eigentriples [51]. In case of a continuous eigenvalue of multiplicity more than one, a number of discrete eigentriples equal to the multiplicity of the continuous eigenvalue must be computed to ensure that the reconstruction technique works.

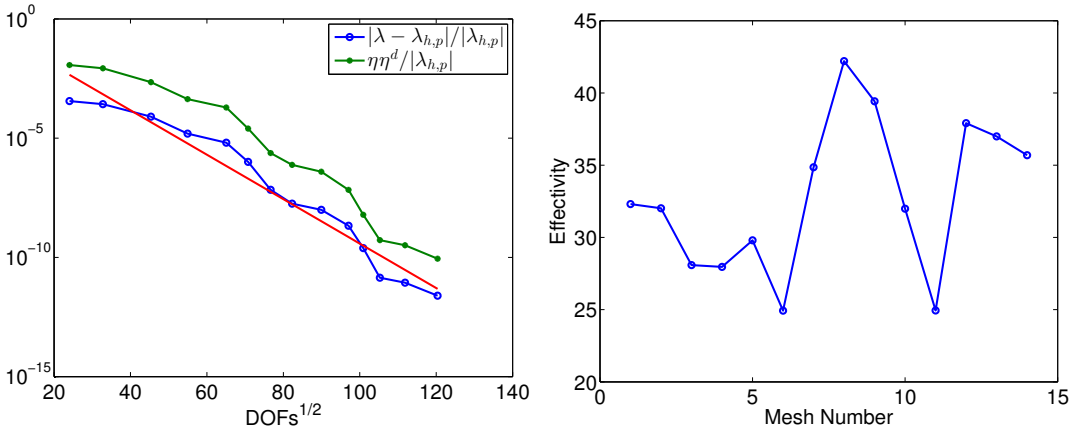
Let us shortly summarize the adaptive algorithm used in our simulations. At first we choose the indices i of the eigenvalues of interest. On the initial coarse mesh we compute the corresponding eigenpair $(\lambda_{h,p}, u_{h,p})$ and the *a-posteriori* error estimators. We determine the elements $T \in \mathcal{T}$ for refinement using a simple fixed-fraction strategy based on the values of the local error estimators

$$\eta(u_{h,p}, \lambda_{h,p})^2|_T := \left(\frac{h(T)}{p(T)}\right)^2 \|R_T(u_{h,p}, \lambda_{h,p})\|^2_{0,T} + \frac{1}{2}\sum_{\mathbf{e}\in\mathcal{E}(T)} \frac{h(\mathbf{e})}{p(\mathbf{e})}\|R_{\mathbf{e}}(u_{h,p}, \lambda_{h,p})\|^2_{0,\mathbf{e}} \ ,$$

where $\mathcal{E}(T)$ is the set of all edges of T.

The choice between refining the marked elements in h or p is based on an estimation of the local analyticity of the exact eigenvectors using the computed ones, see [31] for more details. On each successive refined space a desired number of eigenpairs are computed and the reconstruction technique in [51] is applied. The reconstructed eigentriple is then used in the a-posteriori error estimator and for refining the space.

All the experiments have been carried out using the APTOFEM package (`www.aptofem.com`) on a single processor desktop machine. In particular, we have used ARPACK [41] to solve the algebraic eigenvalue problems, employing MUMPS [1] to solve the necessary linear systems.



(a) Convergence of the target eigenvalue. (b) Effectivity index of the target eigenvalue.

Figure 1: Estimated convergence rate 0.1073.

Since ARPACK is based on the Arnoldi algorithm, we have to solve the projected eigenvalue twice. Once for the left eigenvectors and once for the right eigenvectors. In contrast, if we were to use a routine based on the nonsymetric Lanczos procedure such as implemented in ABLEpack, [7] then we would obtain both left as well as right approximate eigenvectors in one go. The issue of the choice of most efficient linear algebra routines is beyond the scope of this article.

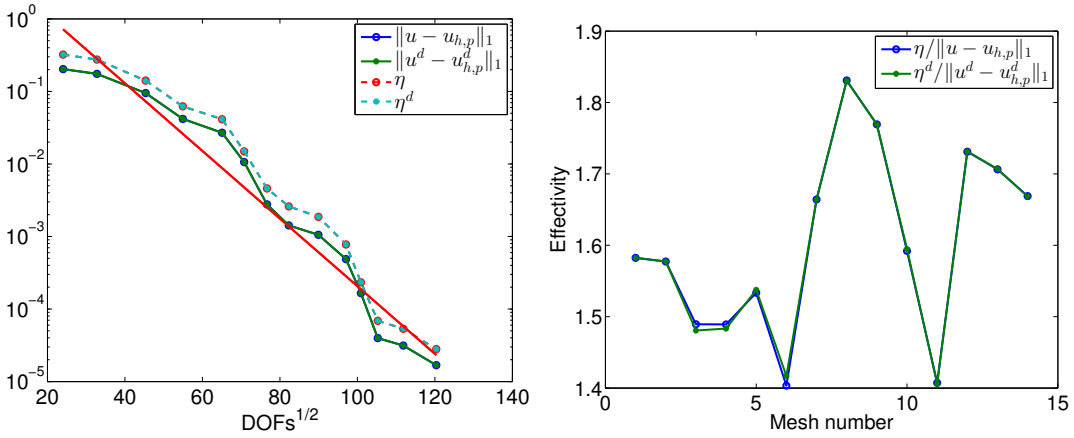
5.1. TM Waves

In the first experiment we consider the TM waves problem (40) on the unit square with periodic boundary conditions and with a square inclusion of size 0.5 in the center of the domain. For the numerical results below we set $\epsilon = 6.340528711808362 + 0.005341062818090i$ outside the inclusion and $\epsilon = 1$ inside. Also we set $\omega = \pi/50$ and $\hat{k} = (1, 0)$. The target eigenvalue has multiplicity two and its reference value is $6.2831865660155 + 6.2816125880788i$ with an accuracy of at least 12 digits.

In Figure 1(a) we present the relative eigenvalue errors and the error estimates for the target eigenvalue using our *hp*-adaptive scheme with 15% for refinement in the fixed-fraction marking strategy. In this case we have that the convergence rate for the eigenvalue estimated with least-squares fitting is $\gamma = 0.1957$. The corresponding effectivity indices are shown in Figure 1(b).

Similarly, the right and left eigenvector errors corresponding to the target eigenvalue with the associated error estimates are depicted in Figure 2(a). Here the convergence rate for the right and left eigenvectors estimated with least-squares fitting are $\gamma = 0.1168, 0.1120$, respectively. Figure 2(b) presents the effectivity indices for eigenvectors. The final *hp*-adapted mesh is displayed in Figure 3. The fact that the order of polynomials are quite high almost everywhere, suggests that the eigenvectors are smooth and so $p-$adaptivity is preferred to $h-$adaptivity. Finally in Figure 4(b) and Figure 4(a) the real and imaginary part of the left eigenfunction is presented.

We note that the measured effectivity of the estimators for eigenvalues is further away from one than is the effectivity of the eigenvectors. This is a typical behavior of the standard hp-residual eigenvalue estimator. However, we also point out that the decay rate for the error is well replicated by the decay rate of the estimator. This indicates, as the experiments corroborate that marking



(a) Eigenvalues. (b) Effectivity index.

Figure 2: Convergence of the left and right eigenvectors (eigenfunctions) corresponding to the target eigenvalue. Estimated convergence rates for the left and right eigenvectors (eigenfunctions) are respectively: 0.0537 and 0.0537.

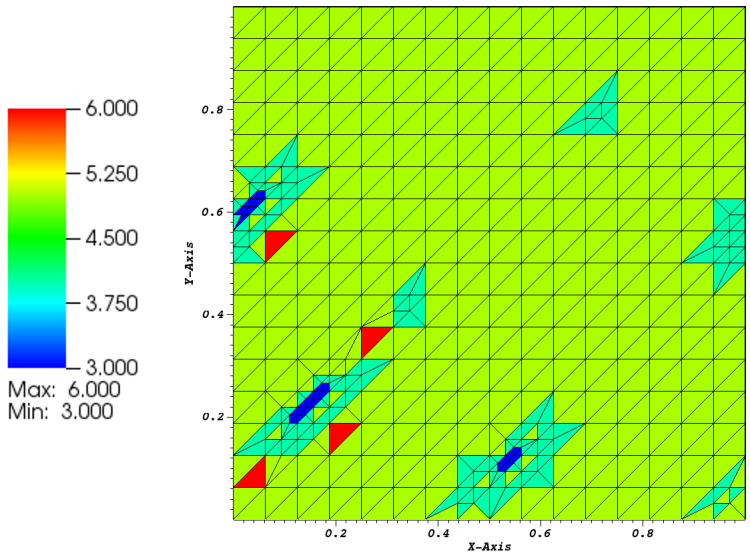
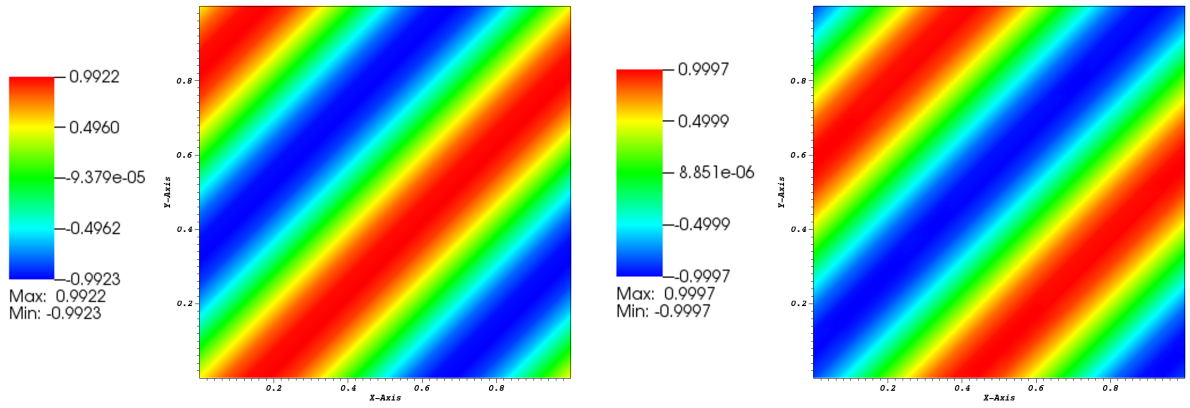


Figure 3: Final *hp*-adapted mesh for the TM waves problem with the order p of polynomials expressed on the color scale.



(a) Imaginary part of the eigenfunction for the target eigenvalue for the TM waves problem.

(b) Real part of the eigenfunction for the target eigenvalue for the TM waves problem.

Figure 4: TM eigenvector.

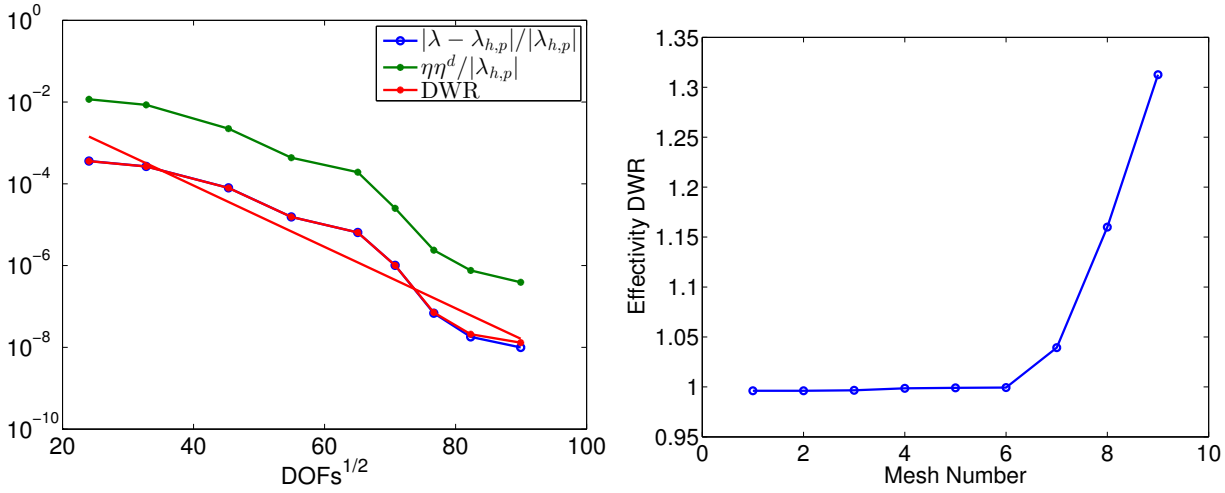
the triangles for refinement based on the hp-residual is a good strategy. We also make claims on the highly accurate benchmark eigenvalues. They were computed using a goal oriented approach on a much finer space. This estimator is defined in (55) and will be abbreviated as DWR (dual weighted residual estimator) on the plots in Figure 5. To this end we present a convergence history for the error estimator from Section 4.2. The observed effecitivities are close to one.

5.2. TE Waves

In the second experiment we consider the TE waves problem (42) on the unit square with periodic boundary conditions and with a square inclusion of size 0.5 in the center of the domain. For the numerical results below we set $\epsilon = 1/(6.340528711808362 + 0.005341062818090i)$ outside the inclusion and $\epsilon = 1$ inside. Also we set $\omega = \pi/50$ and $\hat{k} = (1, 0)$. The target eigenvalue has multiplicity one and its reference value is $5.432928082 - 3.125220283i$ with an accuracy of at least 8 digits.

In Figure 6(a) we present the relative eigenvalue errors and the error estimates for the target eigenvalue using our *hp*-adaptive scheme with 15% for refinement in the fixed-fraction marking strategy. In this case we have that the convergence rate for the eigenvalue estimated with least-squares fitting is $\gamma = 0.10723$. The convergence rate for this example is lower compared to the previous one, because the eigenfunction is less smooth around the corners of the inclusion. This is also supported by Figure 8 where clearly a lot of $h-$refinement has been done in those areas. The corresponding effectivity indices are shown in Figure 6(b).

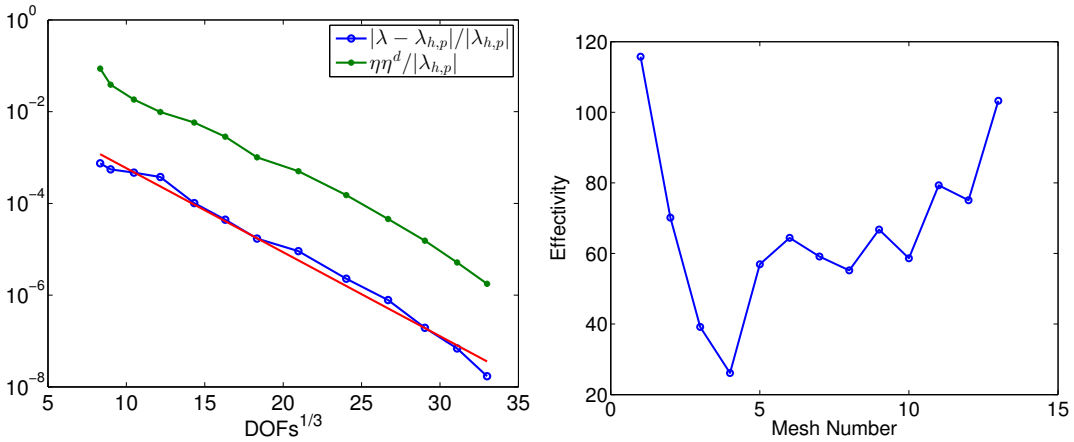
Similarly, the right and left eigenvector errors corresponding to the target eigenvalue with the associated error estimates are depicted in Figure 7(a). Here the convergence rate for the right and left eigenvectors estimated with least-squares fitting are $\gamma = 0.0537, 0.0537$, respectively. As can be seen the left end the right a-posteriori error estimators do not coincide as in Figure 2(a), this is due to the fact that along the faces between different values for ϵ the quantity $-(2i\lambda_{h,p}\alpha\hat{k}u_{h,p})_{|_T} \cdot \mathbf{n}_T - (2i\lambda_{h,p}\alpha\hat{k}u_{h,p})_{|_{T'}} \cdot \mathbf{n}_{T'}$ in (51), which is not present in (52), may not be zero. Figure 7(b) presents



(a) Comparison of the estimators

(b) Effectivity plot for the DWR error estimator.

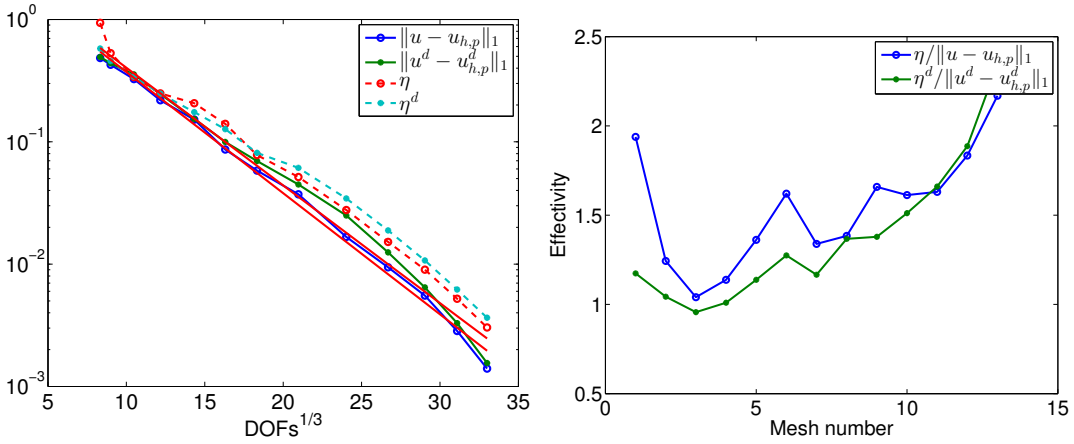
Figure 5: Comparison of the hp-residual and goal oriented residual for TM waves problem.



(a) Convergence of the target eigenvalue.

(b) Effectivity index of the target eigenvalue.

Figure 6: Estimated convergence rate 0.2109.



(a) Eigenvalues.

(b) Effectivity index.

Figure 7: Convergence of the left and right eigenvectors (eigenfunctions) corresponding to the target eigenvalue. Estimated convergence rates for the left and right eigenvectors (eigenfunctions) are respectively: 0.1141 and 0.1108.

the effectivity indices for eigenvectors. The final *hp*-adapted mesh is displayed in Figure 8. The fact that the order of polynomials are quite high almost everywhere, suggests that the eigenvectors are smooth and so $p-$adaptivity is preferred to $h-$adaptivity. Finally in Figure 9(a) and Figure 9(b) the real and imaginary part of the left eigenfunction is presented.

We now present in Figure 10 the performance of the DWR estimator, which was defined in (55), in the TE case. The measured effectivity is again close to one. With this we justify the use of our benchmark values for estimating convergence rates in the experiments.

Remark 5.1. Note that we use fixed fraction marking strategy for our h refinement. This marking strategy does not always result in a refined mesh that possesses any of the symmetries of the approximated functions. The established estimates are operator theoretic results that do not assume or need any symmetry properties of the approximating functions. Furthermore, the convergence rates we observed in our experiments were not influenced by symmetry related properties. However, after many iterations the sequence of refined meshes tended to replicate symmetries of the approximated functions and we also observed that p refinement is eventually preferred. In this paper, we concentrated on validating our convergence rate estimates and did not theoretically analyze those features of the algorithm further.

6. Conclusion

In this paper we presented an analysis of the hp-residual estimators for Fredholm valued polynomial eigenvalue problems. The experiments have shown that the estimator is correctly capturing the convergence rate of the eigenvalues and eigenvectors. However, the measured empirical effectivity of the estimator was frequently away from the ideal effectivity one. This is a known feature of the hp-residual estimator and when one compares the performance of our nonlinear estimator, then one observes that it is similar with the performance of the estimator in the linear case [3]. On the other hand, our operator analysis reduced the analysis of the approximation problem for the eigenvalue problem on the analysis of the reliability and the efficiency the estimates for the

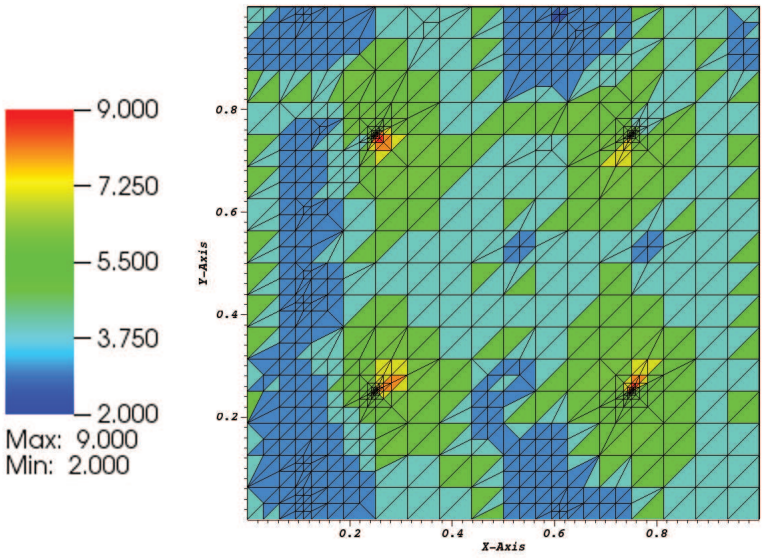
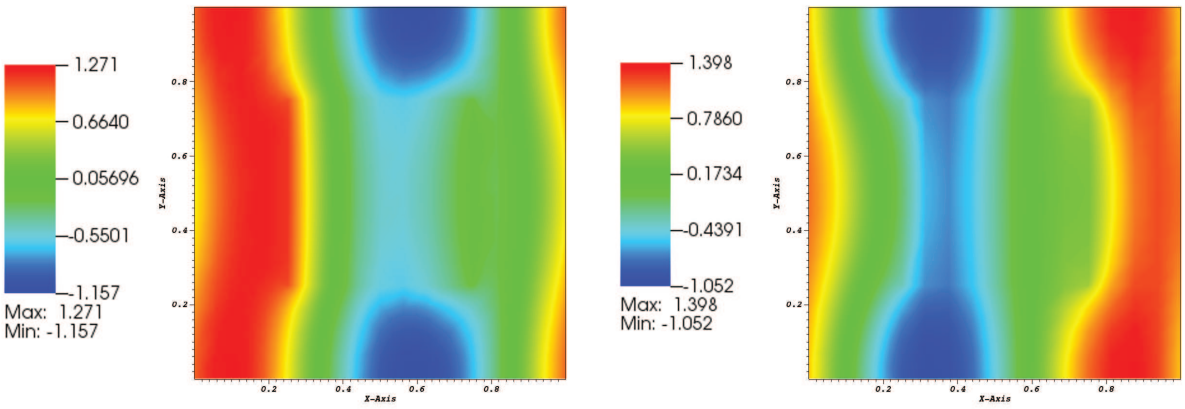


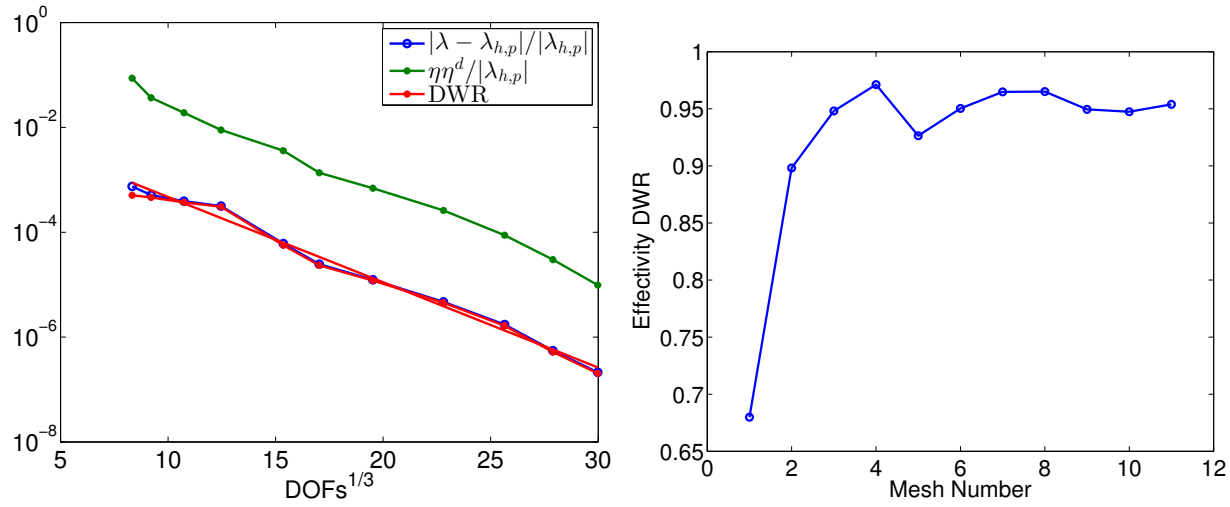
Figure 8: Final *hp*-adapted mesh for the TE waves problem with the order p of polynomials expressed on the color scale.



(a) Real part of the eigenfunction for the target eigenvalue for the TE waves problem.

(b) Imaginary part of the eigenfunction for the target eigenvalue for the TE waves problem.

Figure 9: Computed eigenvectors.



(a) Comparison of the estimators

(b) Effectivity plot for the DWR error estimator.

Figure 10: Comparison of the hp-residual and goal oriented residual for TE waves problem.

negative order Sobolev norm of the left and right residual. This residual norm has been shown as a reliable and efficient estimator of the eigenvalue and eigenfunction error. We have also shown how to construct a reliable a-posteriori error estimator for the invariant subspace approximation error. In this case we have theoretically shown how to use different bases for the invariant subspace than a basis of eigenvectors. Furthermore, any reliable and efficient estimator of the negative order Sobolev norm would equally be reliable and efficient estimator of an eigenvalue error. Based on this, we have constructed an auxiliary subspace error estimator. We call the estimator a goal oriented error estimator since its construction was motivated by the standard techniques in goal oriented adaptivity. The measured performance of this estimator showed not only that it is reliable and efficient, but also has a measured effectivity close to one.

Acknowledgement

The work of L.G. has been supported by the Croatian Science Foundation grant number 9345. C.E. gratefully acknowledge the support of the Swedish Research Council under the project grant *Spectral analysis and approximation theory for a class of operator functions*.

References

[1] P.R. Amestoy, I.S. Duff and J.-Y. L'Excellent. Multifrontal parallel distributed symmetric and unsymmetric solvers. Comput. Methods in Appl. Mech. Eng. 184, 501–520, 2000.

[2] Th. Apel, A.-M. Sändig and S.I. Solovev. Computation of 3D vertex singularities for linear elasticity: Error estimates for a finite element method on graded meshes. Math. Model. Numer. Anal., 36:1043–1070, 2002.

[3] M.G. Armentano, C. Padra, R. Rodríguez, and M. Scheble. An *hp* finite element adaptive scheme to solve the Laplace model for fluid-solid vibrations, Comput. Methods Appl. Mech. Engrg. 200:178–188, 2011.

[4] I. Babuška and J. Osborn. Eigenvalue problems, Handbook of numerical analysis, Vol. II. pp. 641–687. North-Holland, Amsterdam, 1991.

[5] I. Babuška and M. Suri. The *h-p* version of the finite element method with quasiuniform meshes. RAIRO Model. Math. Anal. Numar. 21, 119–238, 1987.

[6] I. Babuška and M. Suri. The *p* and *h-p* versions of the finite element method, basic principles and properties. SIAM Rev. 36(4), 578–632, 1994.

[7] Z. Bai et al. Templates for the Solution of Algebraic Eigenvalue Problems: a Practical Guide. SIAM, Philadelphia, 2000.

[8] R. Becker and R. Rannacher, An optimal control approach to a-posteriori error estimation in finite element methods. Acta Numerica 10, 1–102, 2001.

[9] A. Bermudez, R.G. Duran, R. Rodriguez and J. Solomin. Finite element analysis of a quadratic eigenvalue problem arising in dispersive acoustics. SIAM J. Numer. Anal. 38(1), 267–291, 2001.

[10] T. Betcke, N. Higham, V. Mehrmann, C. Schröder and F. Tisseur. NLEVP: A Collection of Nonlinear Eigenvalue Problems ACM Trans. Math. Softw. 39 (2), 7:1–7:28, 2013.

[11] T. Betcke and D. Kressner. Perturbation, extraction and refinement of invariant pairs for matrix polynomials. Linear Algebra Appl. 435 (3), 536–574, 2011.

[12] W. Beyn and V.Thümmler. Continuation of invariant subspaces for parameterized quadratic eigenvalue problems SIAM J. Matrix Anal. Appl. 31, 1361–1381, 2009.

[13] D. Bourne, H. Elman and J.E. Osborn. A non-self-adjoint quadratic eigenvalue problem describing a fluid-solid interaction Part II: analysis of convergence. Commun. Pure Appl. Anal. 8(1), 143–160, 2009.

[14] H. Brandsmeier, K. Schmidt and C. Schwab. A multiscale hp-fem for 2d photonic crystal bands. J. Comput. Phys. 230(2), 349 – 374, 2011.

[15] S.Brenner and R. Scott. The mathematical theory of finite element methods, volume 15 of Texts in Applied Mathematics, second edition. Springer-Verlag, New York, 2002.

[16] K. Busch, S.F Mingaleev, A. Garcia-Martin, M. Schillinger and D. Hermann. The wannier function approach to photonic crystal circuits. J. Phys.: Condens. Matter 15, 2003. R1233

[17] C. Carstensen, J. Gedicke, V. Mehrmann and A. Miedlar. An adaptive finite element method with asymptotic saturation for eigenvalue problems Numer. Math., 128(4), 615–634, 2014.

[18] M. Cessenat. Mathematical Methods in Electromagnetism. Series on Advances in Mathematics for Applied Sciences — Vol. 41. World Scientific Publisher, Singapore, 1996.

[19] M. Davanco, Y. Urzhumov and G. Shvets. The complex Bloch bands of a 2D plasmonic crystal displaying isotropic negative refraction. Opt. Express 15(15), 9681–9691, 2007.

[20] R.G. Durán, C. Padra and R. Rodríguez. A posteriori error estimates for the finite element approximation of eigenvalue problems. Math. Models Methods Appl. Sci. 13, 1219, 2003.

[21] C. Engström. Spectral approximation of quadratic operator polynomials arising in photonic band structure calculations. Num. Mat., Volume 126(3), 413–440, 2014.

[22] C. Engström. On the spectrum of a holomorphic operator-valued function with applications to absorptive photonic crystals. Math. Models Methods Appl. Sci. 20, 1319–1341, 2010.

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

[24] C. Engström and M. Richter. On the spectrum of an operator pencil with applications to wave propagation in periodic and frequency dependent materials. SIAM J. Appl. Math. 70(1), 231–247, 2009.

[25] S. Giani and I.G. Graham. Adaptive finite element methods for computing band gaps in photonic crystals. Numer. Math. 121(1), 31– 64, 2012.

[26] S. Giani, L. Grubisic and J. Ovall. Benchmark results for testing adaptive finite element eigenvalue procedures. *Applied Numerical Mathematics*, 62(2):121–140, 2012.

[27] I.C. Gohberg and M.G. Kreĭn. Introduction to the theory of linear nonselfadjoint operators, *Translations of mathematical monographs*, vol. 18. American Mathematical Society, Providence, Rhode Island, 1969.

[28] I. C. Gohberg and E. I. Sigal. An operator generalization of the logarithmic residue theorem and Rouché's theorem. *Mat. Sb. (N.S.)*, 84(126):607–629, 1971.

[29] V. Heuveline and R. Rannacher. A posteriori error control for finite element approximations of elliptic eigenvalue problems. Advances in Comput. Math. 15, 1–32, 2001.

[30] V. Heuveline and R. Rannacher. Adaptive FEM for eigenvalue problems with application in hydrodynamic stability analysis. Univ. Heilderberg Preprint, 2006.

[31] P. Houston and E. Süli. A note on the design of *hp*-adaptive finite element methods for elliptic partial differential equations. Comp. Methods in Appl. Mech. Eng. 194, 229–243, 2005.

[32] K.C. Huang and E. Lidorikis, X. Jiang, J.D. Joannopoulos, K.A. Nelson, P. Bienstman and S. Fan. Nature of lossy Bloch states in polaritonic photonic crystals. Phys. Rev. B 69, 195111, 10 Pages, 2004.

[33] E. Istrate, A.A. Green, and E.H. Sargent. Behavior of light at photonic crystal interfaces. Phys. Rev. B. 71(19), 2005. 195122

[34] E. Istrate and E.H. Sargent. Photonic crystal heterostructures and interfaces. Rev. Modern Phys. 78, 455–481, 2006.

[35] O. Karma. Approximation in eigenvalue problems for holomorphic Fredholm operator functions. I. Numer. Funct. Anal. Optim. 17(3–4), 365–387, 1996.

[36] O. Karma. Approximation in eigenvalue problems for holomorphic Fredholm operator functions. II. Numer. Funct. Anal. Optim. 17(3–4), 389–408, 1996.

[37] T. Kato. Perturbation Theory for Linear Operators. Springer-Verlag, Berlin, 1980.

[38] W.G. Kolata. Spectral approximation and spectral properties of variationally posed nonselfadjoint problems. Dissertation, University of Maryland, 1976.

[39] W.G. Kolata. Approximation in variationally posed eigenvalue problems. Numer. Math 29, 159–171, 1978.

[40] P. Kuchment. Floquet Theory for Partial Differential Equations. Birkhäuser, Basel, 1993.

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

[42] A.S. Markus. Introduction to the Spectral Theory of Polynomial Operator Pencils. Transl. Math. Monogr. 71, AMS, Providence, 1988.

[43] J. M. Melenk and B. I. Wohlmuth. On residual-based a-posteriori error estimation in hp-FEM. *Adv. Comput. Math.*, 15(1-4):311–331, 2001.

[44] R. Mennicken and M. Möller. *Non-self-adjoint boundary eigenvalue problems*, volume 192 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 2003.

[45] J.C. Miller and J.N. Miller. Statistics for Analytical Chemistry. Ellis Horwood Ltd, 1993.

[46] J.E. Osborn. Spectral approximation for compact operators. Math. Comput. 29, 712–725, 1975.

[47] C. Pester. A posteriori error estimation for non-linear eigenvalue problems for differential operators of second order with focus on 3-D vertex singularities. Dissertation, TU. Chemnitz, 2006.

[48] K. Schmidt and R. Kappeler. Efficient computation of photonic crystal waveguide modes with dispersive material. Opt. Express 18(7), 7307–7322, 2010.

[49] K. Schmidt and P. Kauf. Computation of the band structure of two-dimensional photonic crystals with *hp* finite elements. Comput. Methods Appl. Mech. Engrg. 198, 1249–1259, 2009.

[50] F.Tisseur and K. Mehrbergen. The quadratic eigenvalue problem SIAM review 43(2), 235–286, 2001.

[51] P. Solin and S.Giani. An iterative adaptive hp-FEM method for non-symmetric elliptic eigenvalue problems. Computing 95(1), 183-213, 2013.

[52] L. N. Trefethen. *Approximation theory and approximation practice.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2013.

[53] G.M. Vaĭnikko. On the convergence speed of convergence of approximative methods in eigenvalue problems. U.S.S.R. Comp. Math. and Math. Phys. 7, 18–32, 1967.

[54] R. Verfürth. A posteriori error estimates for nonlinear problems. Finite element discretizations of elliptic equations. *Math. Comp.*, 62(206):445–475, 1994.